Passenger Flow Estimation Method for Ridesharing Systems using Inverse Problem

Koki Oda
Department of Civil and Environmental Engineering
Institute of Science Tokyo
Tokyo, Japan
oda.k.ae@m.titech.ac.jp

Toru Seo
Department of Civil and Environmental Engineering
Institute of Science Tokyo
Tokyo, Japan
seo.t.aa@m.titech.ac.jp

***Abstract*—Ridesharing is attracting attention as a new means of transportation. However, in general, passenger flows in ridesharing systems are difficult to be recognized by third parties due to several reasons. First, unlike conventional private cars, flows of people and cars are disentangled. Second, although companies operating ridesharing systems know their passenger flow in detail, such information might not be fully shared to third parties due to privacy protection and corporate confidentiality. This will pose substantial difficulty for the management or regulation of ridesharing services by public authorities and the integration with other transportation modes provided by other parties. To account for this challenge, this study proposes a novel estimation method for passenger flow in ridesharing systems by using the inverse problem framework. The proposed method estimates passenger flow as well as operational strategy of ridesharing systems from observable vehicle flow data. Results of numerical validations show that the method accurately estimates these information even under erroneous observation.**

***Index Terms*—ridesharing, inverse problem, passenger flow estimation, operational strategy estimation**

I. Introduction

Ridesharing has emerged as a promising new mode of transportation. In this study, ridesharing is defined as a service where individuals or groups with temporally and spatially similar destinations share a vehicle and travel together, mutually bearing the incurred transportation costs. It should be noted that ride-sourcing services, such as the standard offerings of Transportation Network Companies (e.g., Uber and Lyft), are excluded from this definition. Ridesharing enables solo drivers to carpool, thereby reducing overall travel distances despite potential detours or schedule adjustments, and also provides mobility options for individuals unable to drive. With the advancement of internet connectivity and the widespread use of smartphones, the sharing economy is rapidly expanding, and ridesharing is expected to continue its growth.

Ridesharing offers several distinctive advantages: a higher number of passengers per vehicle compared to private cars, leading to reduced environmental impacts; greater route flexibility compared to trains and buses; and lower transportation costs compared to taxis. However, it has been suggested that appropriate planning and regulation are essential to fully realize these benefits (Beojone and Geroliminis [1]; Narayanan et al. [2]). Consequently, effective planning and regulatory measures are crucial for the widespread adoption and sustainable development of ridesharing.

For transportation and urban planning, public authorities must accurately understand traffic conditions, particularly passenger flow data such as flow rates and origin-destination (OD) demands. These data are necessary for forecasting demand, optimizing public transport networks, and designing road infrastructure. In the context of ridesharing, they are also vital for planning waiting areas, regulating services, and improving service efficiency.

However, in a society where ridesharing is prevalent, passenger flow is one of the most difficult types of data to observe. The reasons are twofold. First, unlike conventional private cars, flows of people and cars are disentangled in ridesharing systems. This is due to the unique features of ridesharing systems, such as ridesharing, empty vehicles running, detour for matching. Second, although operators can collect detailed data through matching systems and vehicle tracking, the sharing of such information with public authorities is often limited due to concerns about privacy protection and corporate confidentiality. D'Agostino et al. [3] have discussed the challenges associated with restricted mobility data sharing and proposed principles to address privacy and confidentiality issues, although many of these principles remain largely unimplemented. Unlike vehicle flow, passenger flow cannot be readily observed externally, posing significant challenges for public authorities in monitoring and managing ridesharing systems. To the authors' knowledge, no studies have investigated this issue.

Various methods have been developed to estimate passenger flow in conventional public transportation systems. Many developed countries conduct transportation surveys, such as household travel surveys, based on questionnaires or the analysis of IC card usage records. However, these surveys are conducted infrequently and often involve limited sample sizes, making them insufficient for comprehensive planning and regulation. Traditional estimation models include the gravity model (Low et al. [4]), the network equilibrium approach (Nguyen et al. [5]), and the entropy maximization method (Zuylen et al. [6]). More recently, advances in automated data collection systems have enabled the use of large-scale data for passenger flow estimation. Kumar et al. [7] proposed a method

based on trip chaining using smart card data from buses and railways, while Ge and Fukuda [8] developed an estimation method utilizing mobile device trajectory data with entropy maximization techniques. Mohammed et al. [9] have provided a comprehensive review of various other estimation methods. Additionally, short-term forecasting techniques incorporating Kalman filters and deep learning methods have been proposed (Jiao et al. [10]; Liu et al. [11]).

Research on ridesharing, meanwhile, has primarily focused on optimizing dynamic matching between passengers and vehicles (Herbawi and Weber [12]; Agatz et al. [13]). Seo and Asakura [14] formulated a multi-objective optimization model for the operation and planning of shared autonomous vehicles (SAVs), demonstrating that ridesharing can reduce costs for passengers, operators, and society as a whole. Nevertheless, studies investigating the operational state of ridesharing remain scarce, and methods for estimating operational strategies from observed traffic conditions have not yet developed.

Liu et al. [15] investigated a similar issue. They proposed a model for the equilibrium analysis of Mobility-on-Demand (MoD) services that accounts for the potential non-disclosure of operational strategies. Conventional MoD equilibrium models (Ban et al. [16]; Xu et al. [17]) assume access to operational strategies, yet such information is often treated as corporate confidential data and is unlikely to be shared with regulatory authorities. Similar challenges are anticipated in the context of ridesharing, underscoring the need for methods capable of simultaneously estimating passenger flow and operational strategies.

The objective of this study is to develop a method for estimating passenger flow using relatively easily obtainable data, assuming a future society where ridesharing has become widespread. In addition to passenger flow, the method aims to estimate parameters that quantitatively represent operational strategies. Since this kind of problems have not been investigated by the existing study, we have proposed a novel methodology based on the inverse problem framework, in which ridesharing systems' operation model is integrated and a tailored estimation algorithm is incorporated. Numerical simulations were conducted to validate the performance of the proposed method and to identify challenges for future theoretical development.

II. Method

A. Assumptions

This study considers the problem of estimating passenger flow and the operator's strategy based on information available to a third party (e.g., public organizations) in a setting where only ridesharing services are provided. The operator is assumed to manage the system based on a specific strategy. The behavior of ridesharing vehicles is represented using a Dynamic Traffic Assignment (DTA) model on a road network of nodes and links with respective capacities. The DTA model in this study does not account for accidents, congestion, or inflows and outflows from outside the network, and all vehicles are assumed to be of the same type.

1) Available information: Considering the ease of acquisition, the information available to the estimator (e.g., public organizations) in the proposed method is as follows:

- Road network information
  - Network topology
  - Link length d_{ij}
  - Travel time on the link t_{ij}
  - Capacities of links and nodes μ_{ij}, κ_i
- Vehicle information
  - Vehicle capacity ρ
  - Total number of vehicles N
- Vehicle flow rate x_{ij}^t
- OD demand ratio ${r_M}_{s,i}^k$

The vehicle flow rate x_{ij}^t can be obtained through external observations and is considered easier to acquire than passenger flow, as it can also be estimated using existing traffic state estimation methods. The OD demand ratio, ${r_M}_{s,i}^k$, represents the proportion of each OD demand relative to the total OD demand, ensuring a sum of 1. If location information for some passengers is available, partial OD information can be derived. Assuming these passengers are evenly distributed, the OD demand ratio can be estimated from their OD data. Since location data can be shared via services other than ridesharing, such as smartphone GPS, it is believed to be more accessible than passenger flow.

2) Optimal vehicle assignment problem: It is assumed that all vehicles and passengers follow Optimal vehicle assignment problem (OVAP) (1) – (13) under a unified strategy. This problem is formulated based on the operational terms of the unified optimization problem for the operation and infrastructure design of the SAV system, as formulated by Seo and Asakura [14]. The mathematical notations are provided in the Appendix.

$$\underset{\substack{x_{ij}^t, y_{s,ij}^{k,t},\\ B_{s,i}^{k,t}, A_s^{k,t}}}{\text{minimize}} \quad \alpha \sum_{ij, i\neq j, t} d_{ij} x_{ij}^t + \sum_{ij,s,k,t\in T_k} t_{ij} y_{s,ij}^{k,t} \tag{1}$$

$$\text{subject to} \quad \sum_j x_{ji}^{t-t_{ji}} - \sum_j x_{ij}^t = 0, \quad \forall i, t \in (0, t_{\max}) \tag{2}$$

$$\sum_j y_{s,ji}^{k,t-t_{ji}} - \sum_j y_{s,ij}^{k,t} + B_{s,i}^{k,t} = \begin{cases} A_s^{k,t} & (\text{if } i = s) \\ 0 & (\text{if } i \neq s) \end{cases}, \quad \forall i, s, k, t \in T_k \tag{3}$$

$$\sum_{s,k} y_{s,ij}^{k,t} \leq \rho x_{ij}^t, \quad \forall ij, i \neq j, t \tag{4}$$

$$\tag{5}$$

$$x_{ij}^t \leq \mu_{ij}, \quad \forall ij, i \neq j, t \tag{6}$$

$$x_{ii}^t \leq \kappa_i, \quad \forall i, t \tag{7}$$

$$\sum_{t \in T_k} B_{s,i}^{k,t} = M_{s,i}^k, \quad \forall i, s, k \tag{8}$$

$$\sum_{t \in T_k} A_s^{k,t} = \sum_i M_{s,i}^k, \quad \forall s, k \tag{9}$$

$$\sum_{ij} x_{ij}^0 = N \tag{10}$$

$$x_{ij}^t \geq 0, \quad \forall ij, t \tag{11}$$

$$y_{s,ij}^{k,t} \geq 0, \quad \forall ij, s, k, t \in T_k \tag{12}$$

$$B_{s,i}^{k,t} \geq 0, \quad \forall i, s, k, t \in T_k \tag{13}$$

$$A_s^{k,t} \geq 0, \quad \forall s, k, t \in T_k \tag{14}$$

The objective function consists of two terms: the first term represents the total travel distance of the vehicles, and the second term represents the total travel time of the passengers. These are simultaneously minimized. The priority between total travel distance and total travel time can be freely set by the strategy parameter α. In this study, to explicitly address the trade-off relationship between total travel distance and total travel time, as proven by Seo and Asakura [14], it is assumed that $\alpha > 0$ and that phenomena where vehicles travel regardless of passenger demand do not occur.

Constraint (2) represents the vehicle conservation law at each node. Constraint (3) represents the passenger conservation law at each node. The time-expanded network of the incoming and outgoing passenger flow, focusing on node i, is shown in Fig. 1. Constraint (4) represents the vehicle capacity constraint. Constraint (5) represents the capacity constraint on each link. Constraint (6) represents the capacity constraint at each node. Constraint (7) ensures that the number of departures satisfies the OD demand. Constraint (8) ensures that the number of arrivals satisfies the OD demand. Constraint (9) represents the constraint on the total number of vehicles in the road network. Constraints (10) – (13) represent the non-negativity conditions for each decision variable.

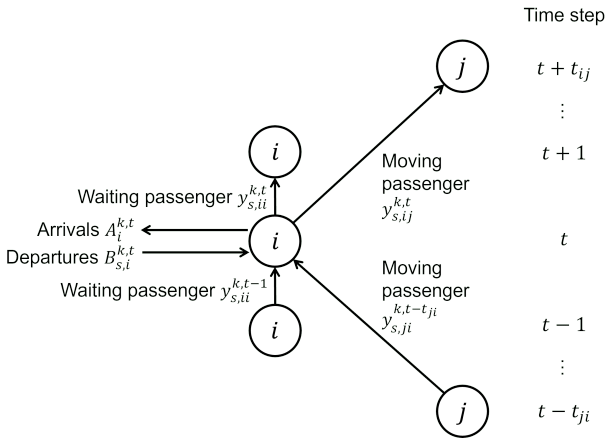

Fig. 1. Passenger flow in time-expanded network

In OVAP, as in Seo and Asakura [14], there is a trade-off relationship between total travel distance D and total travel time T. Multiple Pareto optimal solutions exist for OVAP, and they depend on the strategy parameter α. Moreover, the Pareto optimal solutions and the strategy parameter α do not have a one-to-one relationship. Generally, for a given Pareto optimal solution, there exists a continuous range of α values.

In this study, a method is constructed to estimate passenger flow based on the available information in a society following OVAP. Additionally, strategy parameter is estimated simultaneously with passenger flow. Specifically, the following variables are estimated:

- Passenger flow rate $y_{s,ij}^{k,t}$
- Strategy parameter α, representing how the rideshare operating company values the vehicle operation costs over passenger travel time when determining vehicle operations.
- Vehicle flow rate x_{ij}^t
- OD demand $M_{s,i}^k$
- Number of departing passengers $B_{s,i}^{k,t}$
- Number of arriving passengers $A_s^{k,t}$

B. Formulation of basic method

As described in the previous section, it is assumed in this study that all vehicles follow OVAP and behave rationally to satisfy the OD demand. In this case, passenger flow is inferred from vehicle flow, and a method for inferring the operational strategy from both vehicle flow and passenger flow is considered.

Furthermore, when OVAP that determines vehicle flow from the OD demand is regarded as the forward problem, the problem of determining OD demand from vehicle flow can be called the inverse problem. Therefore, it is theoretically possible to estimate the OD demand from vehicle flow following any model by using the inverse problem. Based on the above, the inverse problem for OVAP is formulated, in which OD demand is estimated from vehicle flow, while passenger flow and the strategy parameter are also estimated simultaneously.

Ahuja et al. [20] formulated the inverse problem for linear programming problems. The inverse problem minimizes the L_p norm concerning the observed and estimated values, subject to the following three constraints. Satisfying these three conditions means that the solution is the optimal solution of OVAP. Thus, this inverse problem can be considered as a problem that seeks the closest solution to the observed values among the optimal solutions of OVAP.

- Feasible for the primal problem
- Feasible for the dual problem
- Complementary slackness condition

Referring to the inverse problem formulated by Ahuja et al. [20], the inverse problem in this study is formulated as shown in equations (14) to (50). This is referred to as basic estimation method. The mathematical notations are provided in the Appendix.

$$\underset{\substack{\hat{M}^k_{i,s},\\ \hat{x}^t_{ij},\hat{y}^{k,t}_{s,ij},\\ B^{k,t}_{s,i},A^{k,t}_s}}{\text{minimize}} \quad \beta \sum_{\substack{ij,t\\,i\neq j}} (\hat{x}^t_{ij} - \tilde{x}^t_{ij})^2 + \sum_{\substack{i,s,k\\,i\neq s}} (\hat{M}^k_{i,s} - r\tilde{M}^k_{i,s} \sum_{\substack{i,s,k\\,i\neq s}} \hat{M}^k_{i,s})^2 \tag{15}$$

$$\text{subject to} \quad \sum_j \hat{x}^{t-t_{ji}}_{ji} - \sum_j \hat{x}^t_{ij} = 0, \quad \forall i, t \in (0, t_{\max}) \tag{16}$$

$$\sum_j \hat{y}^{k,t-t_{ji}}_{s,ji} - \sum_j \hat{y}^{k,t}_{s,ij} + \hat{B}^{k,t}_{s,i} = \begin{cases} \hat{A}^{k,t}_s & i = s \\ 0 & i \neq s \end{cases}, \quad \forall i, s, k, t \in T_k \tag{17}$$

$$\sum_{s,k} \hat{y}^{k,t}_{s,ij} \leq \rho \hat{x}^t_{ij}, \quad \forall ij, i \neq j, t \tag{18}$$

$$\hat{x}^t_{ij} \leq \mu_{ij}, \quad \forall ij, i \neq j, t \tag{19}$$

$$\hat{x}^t_{ii} \leq \kappa_i, \quad \forall i, t \tag{20}$$

$$\sum_{t \in T_k} \hat{B}^{k,t}_{s,i} = \hat{M}^k_{s,i}, \quad \forall i, s, k \tag{21}$$

$$\sum_{t \in T_k} \hat{A}^{k,t}_s = \sum_i \hat{M}^k_{s,i}, \quad \forall s, k \tag{22}$$

$$\sum_{ij} \hat{x}^0_{ij} = N \tag{23}$$

$$\alpha d_{ij} + \lambda_j^{t+t_{ij}} - \lambda^t_i - \rho \xi^t_{ij} + \gamma^t_{ij} \geq 0, \quad \forall ij, i \neq j, t \neq 0 \tag{24}$$

$$\alpha d_{ij} + \lambda_j^{t_{ij}} - \rho \xi^0_{ij} + \gamma^0_{ij} + \eta \geq 0, \quad \forall ij, i \neq j \tag{25}$$

$$\lambda_i^{t+\Delta t} - \lambda^t_i + \epsilon^t_i \geq 0, \quad \forall i, t \neq 0 \tag{26}$$

$$\lambda_i^{\Delta t} + \epsilon^0_i + \eta \geq 0, \quad \forall i \tag{27}$$

$$t_{ij} + \nu^{k,t+t_{ij}}_{s,j} - \nu^{k,t}_{s,i} + \xi^t_{ij} \geq 0, \quad \forall ij, i \neq j, s, k, t \in T_k \tag{28}$$

$$\Delta t + \nu^{k,t+\Delta t}_{s,i} - \nu^{k,t}_{s,i} \geq 0, \quad \forall i, s, k, t \in T_k \tag{29}$$

$$\nu^{k,t}_{s,i} + \phi^k_{s,i} \geq 0, \quad \forall i, s, k, t \in T_k \tag{30}$$

$$-\nu^{k,t}_{s,i} + \zeta^k_s \geq 0, \quad \forall s, k, t \in T_k \tag{31}$$

$$\left(\alpha d_{ij} + \lambda_j^{t+t_{ij}} - \lambda^t_i - \rho \xi^t_{ij} + \gamma^t_{ij}\right) \hat{x}^t_{ij} = 0, \quad \forall ij, i \neq j, t \neq 0 \tag{32}$$

$$\left(\alpha d_{ij} + \lambda_j^{t_{ij}} - \rho \xi^0_{ij} + \gamma^0_{ij} + \eta\right) \hat{x}^0_{ij} = 0, \quad \forall ij, i \neq j \tag{33}$$

$$\left(\lambda_i^{t+\Delta t} - \lambda^t_i + \epsilon^t_i\right) \hat{x}^t_{ii} = 0, \quad \forall i, t \neq 0 \tag{34}$$

$$\left(\lambda_i^{\Delta t} + \epsilon^0_i + \eta\right) \hat{x}^0_{ii} = 0, \quad \forall i \tag{35}$$

$$\left(t_{ij} + \nu^{k,t+t_{ij}}_{s,j} - \nu^{k,t}_{s,i} + \xi^t_{ij}\right) \hat{y}^{k,t}_{s,ij} = 0, \quad \forall ij, i \neq j, s, k, t \in T_k \tag{36}$$

$$\left(\Delta t + \nu^{k,t+\Delta t}_{s,i} - \nu^{k,t}_{s,i}\right) \hat{y}^{k,t}_{s,ii} = 0, \quad \forall i, s, k, t \in T_k \tag{37}$$

$$\left(\nu^{k,t}_{s,i} + \phi^k_{s,i}\right) \hat{B}^{k,t}_{s,i} = 0, \quad \forall i, s, k, t \in T_k \tag{38}$$

$$\left(-\nu^{k,t}_{s,s} + \zeta^k_s\right) \hat{A}^{k,t}_s = 0, \quad \forall s, k, t \in T_k \tag{39}$$

$$\xi^t_{ij} \left(\sum_{s,k} \hat{y}^{k,t}_{s,ij} - \rho \hat{x}^t_{ij} \right) = 0, \quad \forall ij, i \neq j, t \tag{40}$$

$$\gamma^t_{ij} \left(\hat{x}^t_{ij} - \mu_{i,j}\right) = 0, \quad \forall ij, i \neq j, t \tag{41}$$

$$\epsilon^t_i \left(\hat{x}^t_{ii} - \kappa_i\right) = 0, \quad \forall i, t \tag{42}$$

$$\hat{M}^k_{s,s} = 0, \quad \forall s, k \tag{43}$$

$$\hat{M}^k_{s,i} \geq 0, \quad \forall i, s, i \neq s, k \tag{44}$$

$$\hat{x}^t_{i,j} \geq 0, \quad \forall ij, t \tag{46}$$

$$\hat{y}^{k,t}_{s,ij} \geq 0, \quad \forall ij, s, k, t \in T_k \tag{47}$$

$$\hat{B}^{k,t}_{s,i} \geq 0, \quad \forall i, s, k, t \in T_k \tag{48}$$

$$\hat{A}^{k,t}_s \geq 0, \quad \forall s, k, t \in T_k \tag{49}$$

$$\xi^t_{ij} \geq 0, \quad \forall ij, i \neq j, t \tag{50}$$

$$\gamma^t_{ij} \geq 0, \quad \forall ij, i \neq j, t \tag{51}$$

$$\epsilon^t_i \geq 0, \quad \forall i, t \tag{52}$$

The constraints (15) to (22) represent the constraints of OVAP. The constraints (23) to (30) represent the constraints of the dual problem. The constraints (31) to (41) represent the complementary slackness conditions. Constraint (42) indicates that no movement occurs within the same node. Constraints (43) to (50) represent the non-negativity conditions.

The non-uniqueness of α is addressed by treating it as a constant rather than a variable. First, the minimum and maximum values of α and its discretization width are set. Starting from the minimum, α is incremented by the discretization width, solving the inverse problem and computing the objective function value at each step. This continues until the maximum value is reached, and the α that minimizes the objective function, along with the corresponding decision variables, is selected as the estimate. If α is given outside the true range, the observed values $\tilde{x}^t_{i,j}$ and $r\tilde{M}^k_{s,i}$ will not be within the feasible region. Consequently, the objective function value will be significantly larger compared to the cases where α falls within the true range.

C. Limitation of basic method

The complementary slackness conditions among the constraints of the basic estimation method take a quadratic form. In optimization problems, quadratic constraints complicate the feasible region compared to linear constraints, making the problem more challenging to solve due to their non-convex nature. The number of complementary slackness conditions increases with the number of nodes n_{node}, the number of links n_{link}, and the total number of time steps n_{time}. In other words, as the target road network expands or is subdivided, the estimation period lengthens, or the temporal discretization becomes finer, the number of these constraints increases.

D. Formulation of improved method

To address the limitation described in the previous section, the big-M method [21] is introduced, and observed values are substituted into the complementary slackness conditions. The big-M method replaces quadratic constraints with two linear constraints by introducing a binary variable b and a sufficiently large constant S. Additionally, observed values $\tilde{x}^t_{ij}$ are substituted into the complementary slackness conditions for x^t_{ij}, replacing them with linear constraints. Consequently, constraints (31) to (41) are reformulated as equations (68) to (89). Equations (51) to (98) are referred to as the improved method.

$$\underset{\substack{\hat{M}^k_{i,s},\\ \hat{x}^t_{ij},\hat{y}^{k,t}_{s,ij},\\ B^{k,t}_{s,i},A^{k,t}_s}}{\text{minimize}} \quad \beta \sum_{\substack{ij,t\\ ,i\neq j}} (\hat{x}^t_{ij} - \tilde{x}^t_{ij})^2 + \sum_{\substack{i,s,k\\ ,i\neq s}} (\hat{M}^k_{i,s} - r\tilde{\mathrm{M}}^k_{i,s} \sum_{\substack{i,s,k\\ ,i\neq s}} \hat{M}^k_{i,s})^2 \tag{53}$$

$$\text{subject to} \quad \sum_j \hat{x}^{t-t_{ji}}_{ji} - \sum_j \hat{x}^t_{ij} = 0, \quad \forall i, t \in (0, t_{\max}) \tag{54}$$

$$\sum_j \hat{y}^{k,t-t_{ji}}_{s,ji} - \sum_j \hat{y}^{k,t}_{s,ij} + \hat{B}^{k,t}_{s,i} = \begin{cases} \hat{A}^{k,t}_s & i = s \\ 0 & i \neq s \end{cases}, \quad \forall i, s, k, t \in T_k \tag{55}$$

$$\sum_{s,k} \hat{y}^{k,t}_{s,ij} \leq \rho \hat{x}^t_{ij}, \quad \forall ij, i \neq j, t \tag{56}$$

$$\hat{x}^t_{ij} \leq \mu_{ij}, \quad \forall ij, i \neq j, t \tag{57}$$

$$\hat{x}^t_{ii} \leq \kappa_i, \quad \forall i, t \tag{58}$$

$$\sum_{t \in T_k} \hat{B}^{k,t}_{s,i} = \hat{M}^k_{s,i}, \quad \forall i, s, k \tag{59}$$

$$\sum_{t \in T_k} \hat{A}^{k,t}_s = \sum_i \hat{M}^k_{s,i}, \quad \forall s, k \tag{60}$$

$$\sum_{ij} \hat{x}^0_{ij} = N \tag{61}$$

$$\alpha d_{ij} + \lambda_j^{t+t_{ij}} - \lambda^t_i - \rho \xi^t_{ij} + \gamma^t_{ij} \geq 0, \quad \forall ij, i \neq j, t \neq 0 \tag{62}$$

$$\alpha d_{ij} + \lambda_j^{t_{ij}} - \rho \xi^0_{ij} + \gamma^0_{ij} + \eta \geq 0, \quad \forall ij, i \neq j \tag{63}$$

$$\lambda_i^{t+\Delta t} - \lambda^t_i + \epsilon^t_i \geq 0, \quad \forall i, t \neq 0 \tag{64}$$

$$\lambda_i^{\Delta t} + \epsilon^0_i + \eta \geq 0, \quad \forall i \tag{65}$$

$$t_{ij} + \nu^{k,t+t_{ij}}_{s,j} - \nu^{k,t}_{s,i} + \xi^t_{ij} \geq 0, \quad \forall ij, i \neq j, s, k, t \in T_k \tag{66}$$

$$\Delta t + \nu^{k,t+\Delta t}_{s,i} - \nu^{k,t}_{s,i} \geq 0, \quad \forall i, s, k, t \in T_k \tag{67}$$

$$\nu^{k,t}_{s,i} + \phi^k_{s,i} \geq 0, \quad \forall i, s, k, t \in T_k \tag{68}$$

$$-\nu^{k,t}_{s,i} + \zeta^k_s \geq 0, \quad \forall s, k, t \in T_k \tag{69}$$

$$\left(\alpha d_{ij} + \lambda_j^{t+t_{ij}} - \lambda^t_i - \rho \xi^t_{ij} + \gamma^t_{ij}\right) \leq S b_{x}{}^t_{ij}, \quad \forall ij, i \neq j, t \neq 0 \tag{70}$$

$$\tilde{x}^t_{ij} \leq S(1 - b_{x}{}^t_{ij}), \quad \forall ij, i \neq j, t \neq 0 \tag{71}$$

$$\left(\alpha d_{ij} + \lambda_j^{t_{ij}} - \rho \xi^0_{ij} + \gamma^0_{ij} + \eta\right) \leq S b_{x}{}^0_{ij}, \quad \forall ij, i \neq j \tag{72}$$

$$\tilde{x}^0_{ij} \leq S(1 - b_{x}{}^0_{ij}), \quad \forall ij, i \neq j \tag{73}$$

$$\left(\lambda_i^{t+\Delta t} - \lambda^t_i + \epsilon^t_i\right) \leq S b_{x}{}^t_{ii}, \quad \forall i, t \neq 0 \tag{74}$$

$$\tilde{x}^t_{ii} \leq S(1 - b_{x}{}^t_{ii}), \quad \forall i, t \neq 0 \tag{75}$$

$$\left(\lambda_i^{\Delta t} + \epsilon^0_i + \eta\right) \leq S b_{x}{}^0_{ii}, \quad \forall i \tag{76}$$

$$\tilde{x}^0_{ii} \leq S(1 - b_{x}{}^0_{ii}), \quad \forall i \tag{77}$$

$$\left(t_{ij} + \nu^{k,t+t_{ij}}_{s,j} - \nu^{k,t}_{s,i} + \xi^t_{ij}\right) \leq S b_{y}{}^{k,t}_{s,ij}, \quad \forall ij, i \neq j, s, k, t \in T_k \tag{78}$$

$$\hat{y}^{k,t}_{s,ij} \leq S(1 - b_{y}{}^{k,t}_{s,ij}), \quad \forall ij, i \neq j, s, k, t \in T_k \tag{79}$$

$$\left(\Delta t + \nu^{k,t+\Delta t}_{s,i} - \nu^{k,t}_{s,i}\right) \leq S b_{y}{}^{k,t}_{s,ii}, \quad \forall i, s, k, t \in T_k \tag{80}$$

$$\hat{y}^{k,t}_{s,ii} \leq S(1 - b_{y}{}^{k,t}_{s,ii}), \quad \forall i, s, k, t \in T_k \tag{81}$$

$$\left(\nu^{k,t}_{s,i} + \phi^k_{s,i}\right) \leq S b_{B}{}^{k,t}_{s,i}, \quad \forall i, s, k, t \in T_k \tag{82}$$

$$\hat{B}^{k,t}_{s,i} \leq S(1 - b_{B}{}^{k,t}_{s,i}), \quad \forall i, s, k, t \in T_k \tag{83}$$

$$\left(-\nu^{k,t}_{s,s} + \zeta^k_s\right) \leq S b_{A}{}^{k,t}_{s}, \quad \forall s, k, t \in T_k \tag{84}$$

$$\hat{A}^{k,t}_s \leq S(1 - b_{A}{}^{k,t}_{s}), \quad \forall s, k, t \in T_k \tag{85}$$

$$\xi^t_{ij} \leq S b_{\xi}{}^t_{ij}, \quad \forall ij, i \neq j, t \tag{86}$$

$$\left(\sum_{s,k} \hat{y}^{k,t}_{s,ij} - \rho \hat{x}^t_{ij}\right) \leq S(1 - b_{\xi}{}^t_{ij}), \quad \forall ij, i \neq j, t \tag{87}$$

$$\gamma^t_{ij} \leq S b_{\gamma}{}^t_{ij}, \quad \forall ij, i \neq j, t \tag{88}$$

$$\left(\hat{x}^t_{ij} - \mu_{i,j}\right) \leq S(1 - b_{\gamma}{}^t_{ij}), \quad \forall ij, i \neq j, t \tag{89}$$

$$\epsilon^t_i \leq S b_{\epsilon}{}^t_i, \quad \forall i, t \tag{90}$$

$$\left(\hat{x}^t_{ii} - \kappa_i\right) \leq S(1 - b_{\epsilon}{}^t_i), \quad \forall i, t \tag{91}$$

$$\hat{M}^k_{s,s} = 0, \quad \forall s, k \tag{92}$$

$$\hat{M}^k_{s,i} \geq 0, \quad \forall i, s, i \neq s, k \tag{93}$$

$$\hat{x}^t_{i,j} \geq 0, \quad \forall ij, t \tag{94}$$

$$\hat{y}^{k,t}_{s,ij} \geq 0, \quad \forall ij, s, k, t \in T_k \tag{95}$$

$$\hat{B}^{k,t}_{s,i} \geq 0, \quad \forall i, s, k, t \in T_k \tag{96}$$

$$\hat{A}^{k,t}_s \geq 0, \quad \forall s, k, t \in T_k \tag{97}$$

$$\xi^t_{ij} \geq 0, \quad \forall ij, i \neq j, t \tag{98}$$

$$\gamma^t_{ij} \geq 0, \quad \forall ij, i \neq j, t \tag{99}$$

$$\epsilon^t_i \geq 0, \quad \forall i, t \tag{100}$$

The improved method transforms all quadratic constraints in the example into linear constraints involving binary variables. In addition, some quadratic constraints are assigned observables, resulting in very simple linear constraints. Thus, the improved method converts the problem into a form that can be more easily solved using general-purpose optimization solvers.

Finally, the computational procedure of the Improved method is presented as follows:

Step 1: Preliminary Preparation
- **Step 1.1**: Set the network and time discretization width.
- **Step 1.2**: Collect available information.

Step 2: Solve the Inverse Problem
- **Step 2.1**: Determine the minimum value, maximum value, and discretization width of α.
- **Step 2.2**: Set α to its minimum value.
- **Step 2.3**: Solve the inverse problem using α and the available information, then record the results.
- **Step 2.4**: Increment α by the discretization width.
- **Step 2.5**: If α is less than the maximum value, return to Step 2.3.

Step 3: Determine the Estimated Values
- **Step 3.1**: Compare the objective function values for each α and select the minimum as the estimated value of α.
- **Step 3.2**: The computational results corresponding to the estimated α are taken as the estimated values for other variables.

III. Validation

A. Settings

To numerically validate the proposed method, simulation experiments were conducted. The approach involves first setting the true traffic states (vehicle flow rate $x^t_{i,j}$, passenger flow rate $y^{k,t}_{s,ij}$, departure counts $B^{k,t}_{s,i}$, arrival counts $A^{k,t}_s$) and operational strategy (strategy parameter α). Then, the available information was provided to the inverse problem, and the obtained solution was checked for consistency with the true traffic state and strategy.

In this experiment, a network consisting of 6 nodes and 14 links, as illustrated in Fig. 2, is used. The travel times of the links connecting Node 1 to Node 2 and Node 2 to Node 3 are relatively short, assuming these links represent arterial roads that enable high-speed travel.

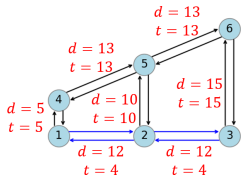

Fig. 2. Network

The true value of α was set to 5, and the true traffic state was generated by solving the optimal vehicle allocation problem with the pre-set true OD demand. The true passenger flow rate $y_{s,ij}^{k,t}$ is aggregated for each departure time k and visualized as shown in Fig. 3. The inverse problem was solved with $\alpha = 5$ to validate whether the true traffic state can be estimated. In other words, this experiment examines whether Fig. 3 can be reproduced. Additionally, the true range of α in this case is $4 \leq \alpha \leq 6$, and it is examined whether the objective function value becomes smaller when α is set within this range.

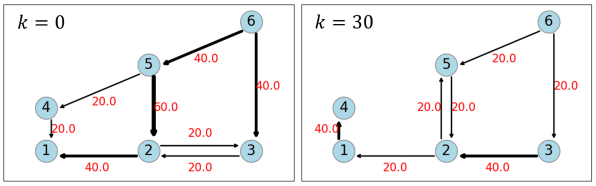

Fig. 3. True passenger flow rate

According to the three estimation scenarios shown in TABLE I, the estimation was conducted. First, calculations were performed using the basic method. Through this experiment, the computational difficulty of the basic method was confirmed. Next, calculations were performed using the improved method. This experiment aimed to validate the effectiveness of the improved method. Finally, calculations were conducted considering the presence of observation errors in $\tilde{x}_{ij}^{t}$ and $\tilde{r}_{\mathrm{M}i,s}^{k}$. This experiment examined the impact of observation errors on estimation accuracy. For solving the inverse problem, Gurobi Optimizer [22], one of the mathematical optimization solvers, was employed.

TABLE I
ESTIMATION SCENARIOS

Scenario	Method	Error	Given α
Basic	Basic	None	3, 5, 7
Improved	Improved	None	0.5, 1, 1.5, . . . , 12
Improved w/ error	Improved	5%	0.5, 1, 1.5, . . . , 12

B. Results and discussion

In Scenario "Improved," the relationship between α and the objective function value is shown in Fig. 4. Since similar results were obtained in other scenarios, only the results for "Improved" are presented as a representative case. The shape of each point indicates whether the optimal solution was obtained within the time limit of six hours; a circle represents cases where the optimal solution was found, while a triangle indicates cases where it was not. Additionally, the color of each point represents whether the given α is included in the range of the true α; red points indicate inclusion, whereas blue points indicate exclusion.

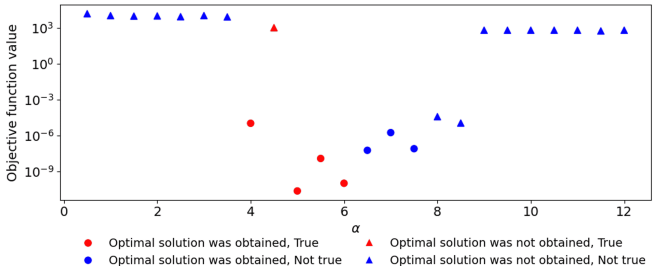

Fig. 4. Strategy parameter α - Objective function value relationship

From Fig. 4, it is suggested that the objective function value is minimized when α is equal to its true value, and that the approximate estimation of α is possible based on the objective function value from the results obtained within the time limit. In the improved method, the optimal solution was obtained when α was close to the range of the true α. Although the objective function value became large when $\alpha = 4.5$, where the optimal solution was not obtained, for other points included in the range of the true α (red points), the objective function value decreased, achieving the intended result. However, the objective function value also decreased for $6.5 \leq \alpha \leq 8.5$, leading to an estimated range of $5 \leq \alpha \leq 7.5$, which indicates that the exact range of α could not be determined.

Next, the objective function value, Symmetric Mean Absolute Percentage Error (SMAPE), and computation time for each scenario are presented in TABLE II. SMAPE is calculated as shown in Eq. (99) and represents the overall estimation accuracy. The time limit for Scenario "Basic" was set to 24 hours, while it was set to 6 hours for the other scenarios. If the optimal solution was not obtained within these time limits, the computation results at that point were reported. In the table, * indicates that the time limit was reached.

$$\text{SMAPE} = \begin{cases} 0 & X = \hat{X} = 0 \\ \frac{2}{n} \sum_{X \in \{x_{ij}^{t}, y_{s,ij}^{k,t}, B_{s,i}^{k,t}, A_{s}^{k,t}, M_{s,i}^{k}\}} \frac{|X - \hat{X}|}{|X| + |\hat{X}|} & \text{else} \end{cases} \tag{101}$$

In the basic method, the optimal solution was not obtained within 24 hours. As a result, the objective function value and SMAPE remained relatively large, indicating that the estimation accuracy was not very high. By using the improved method, the optimal solution was obtained in approximately 20 minutes, demonstrating a significant improvement in computational efficiency. The obtained optimal solution led to

TABLE II
ESTIMATION RESULTS

Scenario	Objective function value	SMAPE	Computation time (sec)
Basic	4814.23	4.72	86400*
Improved	0.00	0.07	1197
Improved w/ error	8.93	0.21	21600*

a substantial reduction in the objective function value and SMAPE, indicating a highly accurate estimation. When observation errors were present, the optimal solution was not obtained. However, the objective function value and SMAPE remained relatively small, suggesting that a solution close to the optimal one was obtained.

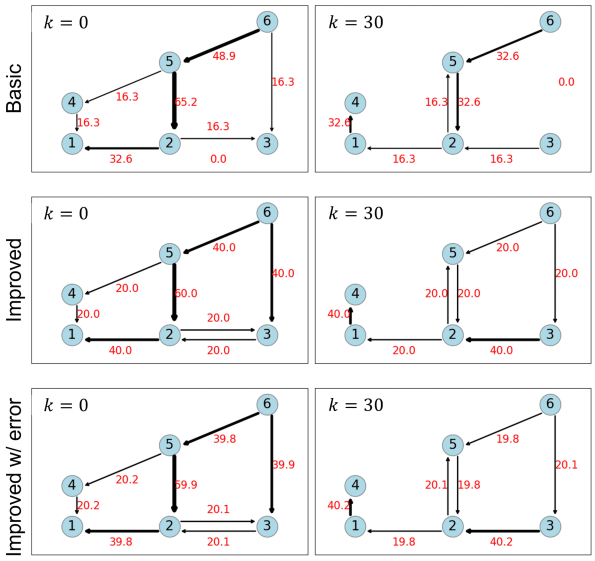

Fig. 5. Estimated passenger flow rate

In the basic method, although the estimation error is large, the overall trend is captured. Therefore, even if obtaining the optimal solution is computationally difficult, passenger flow can still be estimated with a certain level of accuracy by performing partial solution exploration. When the improved method is used, the estimated values are confirmed to match the true values exactly. This indicates that when there is no observation error and the optimal solution is obtained, the proposed method can accurately reproduce the true values. When observation errors are present, the estimated values are nearly identical to the true values, suggesting that passenger flow can be accurately estimated even in the presence of observation errors.

The estimated passenger flow rate $y_{s,ij}^{k,t}$ for each estimation scenario is plotted against the true values in a two-dimensional heatmap, as shown in Fig. 6. The horizontal axis represents the true values, while the vertical axis represents the estimated values. Points located on the red line indicate that the true values were successfully estimated. In the basic method, many points deviate from the red line, suggesting lower estimation accuracy in the detailed estimation results beyond aggregated values. In contrast, with the improved method, all points lie on the red line, indicating that the true values were accurately estimated. Furthermore, even in the presence of observation errors, all points remain on the red line, confirming the high accuracy of the detailed estimation results.

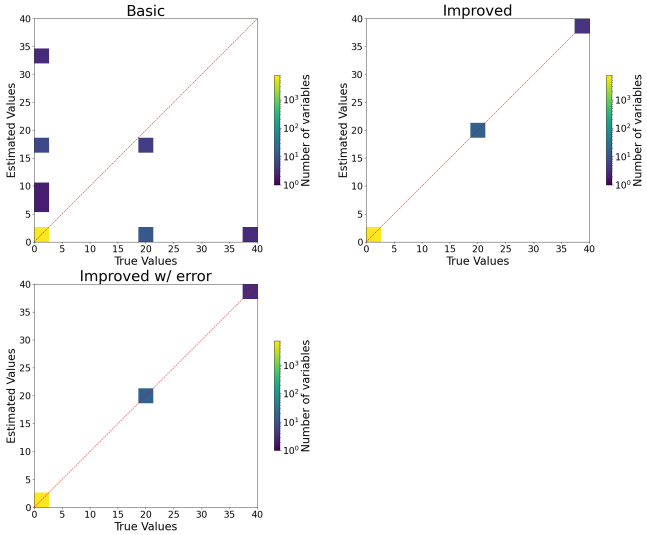

Fig. 6. Comparison of true and estimated passenger flow rate

IV. CONCLUSION

In this research, the method was proposed to simultaneously estimate passenger flow (e.g., passenger flow rate and OD demand) and operational strategy from vehicle flow rate and OD demand ratio by solving the inverse problem of OVAP. Through experiments using simulation data, the characteristics and challenges of the proposed method were validated.

Focusing on ridesharing as a new mode of transportation, and based on the premise that passenger flow estimation is difficult compared to traditional transportation modes, and that the model follows one optimized for the operator and society rather than the user equilibrium model, the completely new passenger flow estimation method was constructed.

In the simulation experiments, it was confirmed that when the actual traffic state follows OVAP, the passenger flow can be estimated accurately even in the presence of observation errors. On the other hand, it was also confirmed that the computational difficulty is a significant challenge of the proposed method. In particular, the calculations became difficult when observation errors were included or when a non-true value of α was given. Consequently, it was suggested that accurate estimation of the strategy parameter α is difficult.

These results contribute to estimating passenger flow using only the highly available data in a society where ridesharing is prevalent. Furthermore, it was suggested that the characteristics of the ridesharing system allow for the collection of necessary data for better operation and society, while protecting corporate confidentiality and passenger privacy, by

ensuring that vehicles and passengers behave in an optimal manner for the operator and society.

A major limitation of the current study is that the proposed algorithm cannot solve realistic, large-scale scenarios due to the computational difficulty of the estimation problem. Development of more sophisticated algorithms is necessary. Another possible direction is the incorporation of the other operation strategies for ridesharing operators by modifying the OVAP. The current strategy is assumed to be system cost minimization, and it is quantified by α, which represents the priority of vehicle operation costs over passenger travel time. However, there should be various strategies such as profit maximization. Other challenges include considerations other transportation modes and multiple ridesharing operators. These factors are essential to make the method more general and realistic.

ACKNOWLEDGEMENTS

This research was partly supported by Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan Science and Technology Agency (JST) / Japan International Cooperation Agency (JICA), Grant No. JPMJSA2405 and a JSPS KAKENHI Grant-in-Aid for Scientific Research 24K01002.

APPENDIX

The meanings of the symbols used in this study are as follows:

- ΔT : Departure time discretization width
- Δt : Time discretization width
- T_k : Set of times considered to have departed at departure time k
- α : Strategy parameter
- $M_{s,i}^k$: Demand from node i to node s at departure time k
- x_{ij}^t : Vehicle flow rate (number of vehicles starting to travel on link ij at time t)
- $y_{s,ij}^{k,t}$: Passenger flow rate (number of passengers who departed from node s and started traveling on link ij at time $t \in T_k$)
- $B_{s,i}^{k,t}$: Number of passengers departing from node i to node s at time $t \in T_k$
- $A_s^{k,t}$: Number of passengers arriving at node s at time $t \in T_k$
- d_{ij} : Length of link ij
- t_{ij} : Travel time of link ij
- μ_{ij} : Capacity of link ij
- κ_i : Capacity of node i
- ρ : Vehicle capacity
- N : Total number of vehicles
- $r_{M_{s,i}^k}$: Ratio of $M_{s,i}^k$
- β: Weight parameter for the norm of x_{ij}^t in relation to the norm of $r_{M_{s,i}^k}$
- λ_i^t, $\nu_{s,i}^{k,t}$, ξ_{ij}^t, γ_{ij}^t, ϵ_i^t, $\phi_{s,i}^k$, ζ_s^k, η: Dual variables

Additionally, $\tilde{}$ denotes observed values, and $\hat{}$ denotes estimated values.

REFERENCES

[1] Beojone, C. V. and Geroliminis, N., "On the inefficiency of ride-sourcing services towards urban congestion," Transportation Research Part C: Emerging Technologies, Vol. 124, p. 102890, 2021.
[2] Narayanan, S., Chaniotakis, E., and Antoniou, C., "Shared autonomous vehicle services: A comprehensive review," Transportation Research Part C: Emerging Technologies, Vol. 111, pp. 255-–293, 2020.
[3] Mollie D'Agostino, Pellaton, P. and Brown, A., "Mobility data sharing: Challenges and policy recommendations," UCDAVIS, pp. 1—29, 2019.
[4] D. Low, "A new approach to transportation systems modelling," Traffic Quarterly, Vol. 26, pp. 391–404, 1972.
[5] S. Nguyen, "Estimating an o-d matrix from network data: a network equilibrium approach," Centre de Recherche surles Transports, No. 87, 1977.
[6] V. Zuylen, "Some remarks on the information minimising method," Verkeersakademie Tilburg, 10 1973.
[7] Qing He, Pramesh Kumar, Alireza Khani, "A robust method for estimating transit passenger trajectories using automated data," Transportation Research Part C, Vol. 95, pp. 731–747, 2018.
[8] Ge. Qian, Daisuke Fukuda, "Updating origin–destination matrices with aggregated data of GPS traces," Transportation Research Part C: Emerging Technologies, Vol. 69, pp. 291–312, 2016.
[9] Mohammed, M. and J., Oke., "Origin-destination inference in public transportation systems: Origin-destination inference in public transportation systems: A comprehensive review," International Journal of Transportation Science and Technology, Vol. 12, pp. 315-–328, 2023.
[10] Tuo Sun, Zenghao Hou, Pengpeng Jiao, Ruimin Li and Amir Ibrahim, "Three revised kalman filtering models for short-term rail transit passenger flow prediction," Mathematical Problems in Engineering, 1 2016.
[11] Ruo Jia, Yang Liu, Zhiyuan Liu, "Deeppf: A deep learning based architecture for metro passenger flow prediction," Transportation Research Part C, Vol. 101, pp. 18–34, 2019.
[12] W. Herbawi and M. Weber, "The ridematching problem with time windows in dynamic ridesharing: A model and a genetic algorithm," IEEE congress on evolutionary computation, pp. 1–8, 2012.
[13] Erera A. L., Savelsbergh M. W., Agatz, N. and X Wang, "Dynamic ride-sharing: A simulation study in metro atlanta," Procedia-Social and Behavioral Sciences, Vol. 17, pp. 532–550, 2011.
[14] Seo, T. and Asakura, Y., "Multi-objective linear optimization problem for strategic planning of shared autonomous vehicle operation and infrastructure design," IEEE Transactions on Intelligent Transportation Systems, Vol. 23, No. 4, pp. 3816-–3828, 2022.
[15] Joseph Y. J. Chow, Bingqing Liu, David Watling, "An operation-agnostic stochastic user equilibrium model for mobility-on-demand networks with congestible capacities," European Journal of Operational Research, pp. 1–41, 12 2024.
[16] Ban. X. J., et al., "A general equilibrium model for transportation systems with e-hailing services and flow congestion," Transportation Research Part B: Methodological, Vol. 129, pp. 273–304, 2019.
[17] Xu. S. J., et al., "Network learning via multiagent inverse transportation problems," Transportation Science, Vol. 52. 6, pp. 1347–1364, 2018.
[18] D. Burton and P.L. Toint, "On an instance of the inverse shortest paths problem," Mathematical Programming, Vol. 53, pp. 45–61, 1992.
[19] Robert B. Dial, "Minimal-revenue congestion pricing part i: A fast algorithm for the singleorigin case," Transportation Research Part B, Vol. 33, pp. 189–202, 1999.
[20] Ahuja R. K. and Orlin J. B., "Inverse optimization," Operations research, Vol. 49, No. 5, pp. 771-–783, 2001.
[21] G.B. Dantzig, "Programming in a linear structure. Comptroller," United Air Force, Tech rep, 1948.
[22] Gurobi Optimization Inc, "Gurobi optimization," Jan. 2025, [Online]. Available: http://www.gurobi.com.