https://doi.org/10.1016/j.tre.2022.102931

Scheduling Zonal-based Flexible Bus Service under Dynamic Stochastic Demand and Time-dependent Travel Time

Enoch Lee[a], Xuekai Cen[b*], Hong K. Lo[a]

[a] Department of Civil and Environmental Engineering,
The Hong Kong University of Science and Technology, Hong Kong
[b] School of Traffic & Transportation Engineering, Central South University, China
[*]Corresponding author: cenxuekai@csu.edu.cn

Abstract
This paper schedules the zonal-based flexible bus service (ZBFBS) considering elastic stochastic demand, stochastic location, time-dependent travel time, and passenger time window constraints based on a scheduled-based formulation. Unlike a traditional time-space network that stipulates the precise arrival and departure times on specific nodes, a zonal-based time-space network is proposed to define the routes in terms of zonal visits of the flexible buses while allowing for flexibility in their arrival and departure times to cater for randomness. The ZBFBS scheduling problem is formulated as a two-stage decision-dependent stochastic problem with recourse. The first stage schedules the zonal visits of flexible buses and the second stage matches each passenger with either flexible bus or ad hoc service, with the latter incurring extra cost to carry the unmatched passengers. To effectively solve the problem, a state-augmented network, that integrates time and zone, is proposed to reduce the number of variables. Moreover, relaxation formulations based on vehicle types and routes are introduced, with an insertion heuristic implemented for vehicle scheduling. The problem is solved by a gradient-based solution approach. Numerical studies demonstrate the efficiency and quality of the solution methods under a variety of ride requests, as well as its advantage over the frequency-based approach in substantially reducing the ad hoc service cost. The applicability of the model is validated by solving an instance of Chengdu, China, with real data.

Keywords: Flexible bus, stochastic dynamic demand, demand responsive transit, reliability

1. Introduction

Public transport services aim to address the transportation needs of people. While most of these services are operated with fixed routes and fixed schedules, they can be improved by better responding to real-time demand. Recently, the development of information technology enables real-time demand-responsive mobility services, such as Didi, Lyft, and Uber, which have become an integral part of transportation services. Despite being convenient and providing direct door-to-door service, taxi and ride-sourcing, which are the most common types of demand-responsive service, aggravate both traffic congestion and pollution (Mourad et al., 2019). Flexible bus services were explored in the 1970s, which provide on-demand mobility services to passengers, without worsening congestion by vehicle sharing (Gong et al., 2021; Kim and Schonfeld, 2015; Liu and Ceder, 2015). As flexible bus service can address the traffic congestion problem while providing high-quality service to travelers efficiently, it is worthwhile to further investigate the service. In the literature, existing studies on flexible buses were mostly based on deterministic demand. For example, Zhang et al. (2017) optimized a flexible bus service for long-distance and single-direction commuter trips. Moreover, Ma et al. (2017) proposed a framework based on large-scale demand data processing and analysis for the flexible bus network design. Readers are referred to Errico et al. (2013) for a comprehensive review of flexible transit services.

As the travel demand typically exhibits some stochastic patterns, it is beneficial to schedule the service while considering stochastic demand, whose distribution can be inferred from historical data. In the proposed zonal-based flexible bus service, the service region is demarcated into separate zones. The service schedule entails a sequence of zonal visits, which is fixed before demand realization, with the vehicles' exact routes within each zone to be determined after the demands are realized. By doing so, passengers can make an informed decision on mode choice as they know ahead of time whether the service vehicles will be covering the zones of their interest. In the literature, demand-responsive transit (DRT) services are often planned based on deterministic demand information to keep the modeling effort simple. For example, Bertsimas et al. (2019) and Liu et al. (2021) planned DRT services after the realization of ride requests. However, considering demand stochasticity while planning DRT services could reduce the cost by as much as 65% (Ritzinger et al., 2015). Ho and Haugland (2011) scheduled a comprehensive DRT service considering demand stochasticity, wherein the service routes were adjusted immediately if the demand was not realized. Lee et al. (2021b) showed that by capturing demand and detour time stochasticity in service planning, a more profitable schedule could be achieved, especially when the demand distributions were not symmetric.

Transit planning models can be classified as either frequency-based or schedule-based (Fu et al., 2012). Frequency-based models design services on a frequency basis without explicitly stating the exact timetable (Jiang and Szeto, 2016; Szeto and Jiang, 2014). In contrast, schedule-based models consider the timetables of the routes, which include the precise departure/arrival times of the services, time-dependent passenger demands, and variable travel times at different times of a day (Hamdouch et al., 2014, 2011; Tong and Wong, 1999; Zhang et al., 2010). In the context of high-speed rail services, Cascetta and Coppola (2016) compared a frequency-based model with a schedule-based model, and found that under uniform demand distribution, the frequency-based model tended to overestimate the modal share by up to 5%; under non-uniform demand distribution, the estimation error was even higher, ranging from -10% to 10%. In general, the schedule-based approach yields a better transit schedule by capturing time-dependent travel time and dynamic demand. For example, in Chengdu, the average driving speed is about 25.7km/h, but the driving speed in peak hours is only 14.75km/h, which is a substantial speed reduction (Sun et al., 2020). Moreover, the schedule-based approach is also proven to be able to improve the reliability of the schedule for some demand-responsive services (Schilde et al., 2014). It suffices to say that the schedule-based approach produces superior results as compared with the frequency-based approach, at the expense of higher model complexity and solution challenges. On the theoretical basis of Lee et al. (2021a, 2021b) for the zonal-based flexible bus service design with the frequency-based approach, this paper aims to formulate and solve a comprehensive schedule-based DRT model that captures dynamic demand and dynamic route travel time with detour to pick up or drop off passengers.

Zonal-based flexible bus service (ZBFBS) plans the fare and zonal visit sequences (routes) of vehicles based on stochastic elastic demand distribution. Then, after demand realization, passengers are assigned to vehicles, and the exact detour routes of vehicles within each zone to pick up and drop off passengers are determined, while satisfying maximum travel time and vehicle capacity constraints. In the event that certain passengers cannot be served due to vehicle capacity or detour time constraints, ad hoc services, such as taxis or ride-hailing services, are provided for them, which incur additional costs. Extending from Lee et al. (2021a), we formulate a schedule-based two-stage stochastic program to optimize the routes (as a sequence of zonal visits), schedules, and pricing of flexible bus service, as schematically illustrated in Figure 1 using a scenario with three zones, two time intervals, and one route. The route, which is determined in the first stage of the stochastic program, is illustrated by bold lines. Then, in the second stage, based on the realized demand locations, illustrated by small circles, the exact detour routes of vehicles within each zone are determined to pick up or drop off passengers. The detour time to pick up or drop off passengers, if within the same time window, is illustrated by a line with double arrows, and if it spans across two consecutive time intervals, is illustrated by a thin dotted line. For simplicity, following Lee et al. (2021a), we assume that the detour time between service requests is independent. The proposed flexible bus optimization model considers elastic stochastic demand, stochastic demand locations, and time-dependent travel times, while obeying the vehicle capacity constraint, maximum detour time constraint, and passengers' time window constraints. The objective of the model is to maximize the profit, and by setting the fare to zero to minimize the total operating cost. The advantage of the flexible bus is its certainty in its zonal route or visits with a specified schedule, which is important for passenger trip planning, while providing door-to-door demand responsive service, striking a good balance between traditional bus service and demand-responsive services such as taxi. A here-and-now decision of vehicle deployment in terms of the sequence of zonal visits is available to the operators and passengers, which allows them to make better operational and travel decisions, such as driver allocations. Then, the demand is assigned to vehicles only after its realization.

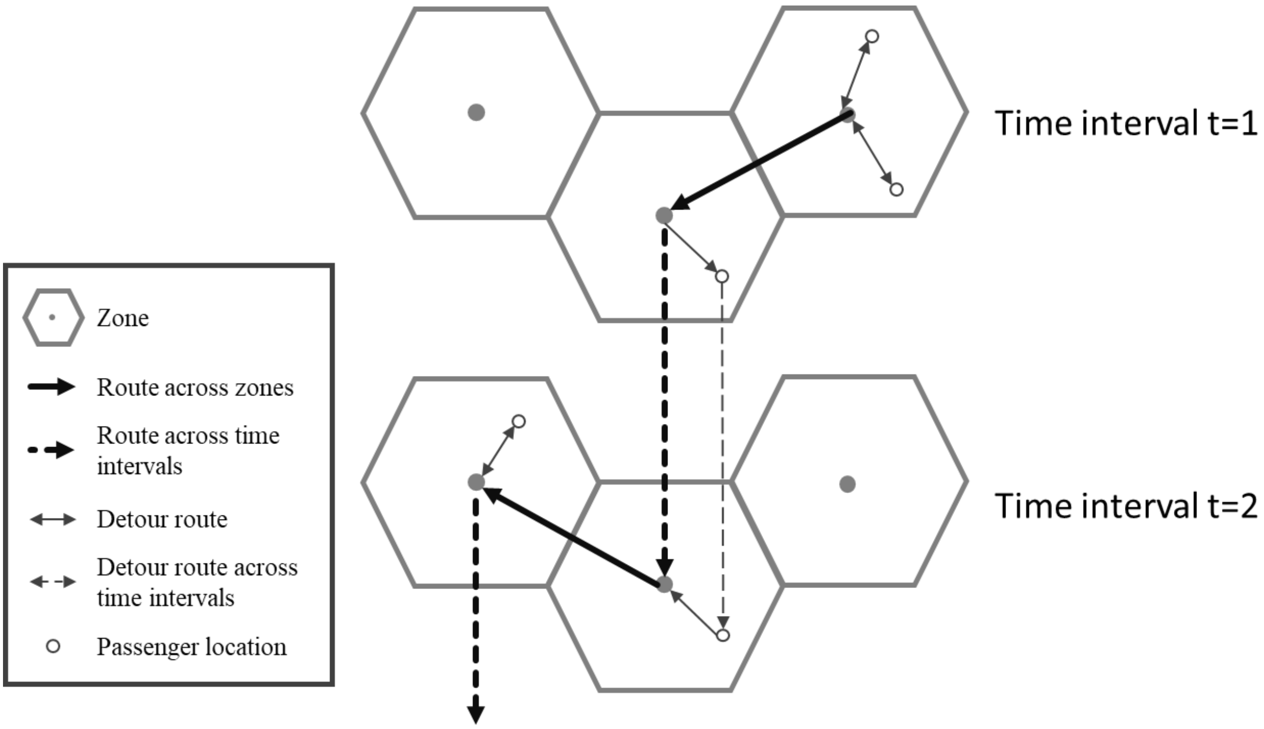

Figure 1. Illustration of the zonal-based flexible bus service under zonal-based time-space network

To address time-dependent travel time and dynamic demand, a novel zonal-based time-space network is developed, while considering vehicle capacity and detour time constraints. Time-space networks have been applied to logistic operations such as convoy movement scheduling (Mokhtar et al., 2020), or transit service scheduling (An and Lo, 2014; Mahmoudi and Zhou, 2016; Steinzen et al., 2010; Tong et al., 2017). Zhao et al. (2018) applied a time-space network to model electric vehicle and staff balancing by discretizing the time of traveling or waiting. However, as time is continuous in nature, time discretization, which is required by models formulated with traditional time-space networks, may affect model performance (e.g. Steinzen et al., 2010). Moreover, traditional time-space networks are unable to represent the stochastic detour times for pick-up or drop-off, as a prespecified deterministic time is needed to construct the time-space network. To address these shortcomings, we propose a novel zonal-based time-space network, which considers travel time across zones and detour time within zone simultaneously in every zone-time node. The zonal-bsed time-space network considers time as continuous. Specifically, in this model, a visit to a zone-time node means that the vehicle must be in the zone and stay there for some time within the time interval (TI); other than this, it has the flexibility to be traveling within the zone to pick up or drop off passengers within the TI, or on the way traveling to the next zone. For example, in the first TI in Figure 1, the vehicle can detour within the first zone to pick up or drop off passengers or travel from the right zone to the middle zone. By considering the time window constraints and precise vehicle schedule, this model is capable of solving also dial-a-ride problems (DARP), as will be demonstrated in the numerical example.

We further propose a solution method to decompose and solve the hard elastic stochastic problem. By introducing reliability measures for demand volume and detour times within each zone, the problem is decomposed into two mixed-integer linear programs (MILPs). The flexible bus schedule is subsequently optimized to meet the designated demand volume and detour time by reliability measures. A gradient-based solution method is applied to optimize the reliability measures and price. The usage of the zonal-based time-space network enables the precise scheduling of departure and arrival times at each zone, which creates a more detailed service schedule than the previous static frequency-based model. Relaxations and heuristics are introduced to further improve the solution method for better scalability.

The contributions in this paper are as follows:

1. To the best of our knowledge, this paper presents the first model that formulates and solves DARP instances while considering elastic stochastic dynamic passenger volume and service locations, and time-dependent travel times.
2. In optimizing the schedule of flexible buses, a novel zonal-based time-space network is developed to consider the stochasticity of demand volume, stochastic zonal detour time, passenger time window constraints, and time-dependent travel time between zones.
3. A reliability-based gradient solution method with relaxations is developed to decompose and solve this inherently non-convex problem. Also, a heuristic solution method is developed, enabling the model to be applicable for real problems, as illustrated through its application to Chengdu, China, based on real data.

The structure of the paper is as follows. The formulation of the ZBFBS scheduling problem is introduced in Section 2. To solve the problem, in Section 3, we introduce reliability measures to decompose the flexible bus scheduling formulation into two separate problems: vehicle-to-route assignment and ride requests-to-vehicle assignment. A gradient-based solution method is then introduced to optimize the price and reliability measures. An insertion heuristic and relaxations are implemented to enhance the solution efficiency while maintaining the solution quality. Numerical studies are conducted in Section 4 to demonstrate its performance and applicability to a real-world scenario in Chengdu, China. Finally, Section 5 concludes this paper.

2. Formulation

2.1. *Notation Table*

This notation table contains every notation used throughout this paper except the appendix.

Notation	Description
Sets	
$\mathcal{N}$	Set of zone-time node
$\mathcal{Z}$ / $\mathcal{Z}^{\text{aug}}$ / $\mathcal{Z}_p$ / $\mathcal{Z}_v$	Set of zones / Set of augmented zone/ Set of augmented zones traversed by route $p \in \mathcal{P}$ / Set of augmented zones traversed by vehicle $v \in \mathcal{V}$
$\mathcal{T}$	Set of time intervals
$\mathcal{T}_r$	Set of feasible time interval pairs for ride request $r \in \mathcal{R}$
$\mathcal{P}$	Set of routes
$\mathcal{V}$	Set of vehicles
$\mathcal{U}$	Set of vehicle types
$\mathcal{K}$	Set of customer realization scenarios
$\mathcal{R}$ / $\mathcal{R}^{\kappa}$	Set of ride requests / Set of ride requests under scenario $\kappa \in \mathcal{K}$
$\mathcal{R}_{z+(-)}$	Set of ride requests with origin (destination) in $z \in \mathcal{Z}^{\text{aug}}$

$\mathcal{C}$ / $\mathcal{C}_p$	Set of demand categories / Set of demand categories that can be served by route p, $p \in \mathcal{P}$
$\mathcal{O}_c, \mathcal{D}_c \subseteq \mathcal{Z}^{\text{aug}}$	Set of possible origins and destinations for demand category $c \in \mathcal{C}$
$\mathcal{O}_{rp}, \mathcal{D}_{rp} \subseteq \mathcal{Z}^{\text{aug}}$	Set of possible origins and destinations for ride request $r \in \mathcal{R}$ or demand category $r \in \mathcal{C}$ by route $p \in \mathcal{P}$
$\mathbb{N}_0$	Set of non-negative integers
Parameters	
$(z,s) \in \mathcal{N}$	Zone-time node of zone z in time interval s, $z \in \mathcal{Z}, s \in \mathcal{T}$
τ_{IJ}^{s}	Travel time between zone I and zone J, departures at time interval s $\text{I},\text{J} \in \mathcal{Z}, s \in \mathcal{T}$
$\tau_{\text{I}p}$	Travel time from zone I to next zone in route $p \in \mathcal{P}$, $\text{I} \in \mathcal{Z}$
$z_{ps}^{\text{first}}, z_{ps}^{\text{second}} \subseteq \mathcal{Z}$	The first and the second zone visited by route p at time interval s, $p \in \mathcal{P}, s \in \mathcal{T}$
$\overline{T}$	Duration of time intervals
c_{pu}^{fixed}	The fixed cost for vehicle v or vehicle type u serving route p, $v \in \mathcal{V}, u \in \mathcal{U}, p \in \mathcal{P}$
c_r^{adhoc}	The ad hoc service cost for ride request $r \in \mathcal{R}$
p_κ	The probability that scenario κ occurs, $\kappa \in \mathcal{K}$
$\tau_{r+(-)}$	The detour time for picking up (dropping off) ride request $r \in \mathcal{R}$ or for any ride requests in zone $r \in \mathcal{Z}$
$\alpha_{r+(-)}^{z}$	Indicating whether the origin (destination) of ride request r is zone z, $r \in \mathcal{R}, z \in \mathcal{Z}$
n_r	Number of passengers for ride request $r \in \mathcal{R}$ or demand category $r \in \mathcal{C}$
$e_r^{\text{dep}}, \ell_r^{\text{dep}}, e_r^{\text{arr}}, \ell_r^{\text{arr}} \in \mathcal{T}$	The earliest departure time, the latest departure time, the earliest arrival time, and the latest arrival time for ride request $r \in \mathcal{R}$ or demand category $r \in \mathcal{C}$
$\mathbf{B}_p$	A converting matrix from passenger flow to the number of in-vehicle passengers for vehicles of route $p \in \mathcal{P}$
$\mathbf{B}_{p+(-)}$	Converting matrix from passenger pick-up (drop-off) to number of passengers boarded (egressed) for vehicles of route $p \in \mathcal{P}$
cap_v	The vehicle capacity of vehicle $v \in \mathcal{V}$ or vehicle type $v \in \mathcal{U}$
M	A very large number
m_p	Sum of the number of time intervals and the number of zones visited by route $p \in \mathcal{P}$
$O_r, D_r \in \mathcal{Z}$	Origin and destination of ride request $r \in \mathcal{R}$
Δ_c	Demand volume distribution of demand category $c \in \mathcal{C}$
$\Lambda_{z+(-)}$	Detour time distribution in zone z for access (egress), $z \in \mathcal{Z}$
N_u	Number of vehicles of vehicle type $u \in \mathcal{U}$
$TW(z) \in \mathcal{T}$	Time interval of the augmented zone $z \in \mathcal{Z}^{\text{aug}}$
Decision variables	
$\overline{t}_{vz}^{s}$	Maximum allowable detour time in zone z for vehicle v at time interval s, $z \in \mathcal{Z}, v \in \mathcal{V}, s \in \mathcal{T}$
$\overline{t}_{uzp}$ / $\overline{t}_{pz}$	Maximum allowable detour time in augmented zone $z \in \mathcal{Z}^{aug}$ for vehicles of type $u \in \mathcal{U}$ and route $p \in \mathcal{P}$ / for route $p \in \mathcal{P}$
$\mathbf{X} = \{x_{pv}\}$	A matrix indicates if vehicle v is serving route p, $v \in \mathcal{V}, p \in \mathcal{P}$
$\mathbf{X}' = \{x_{pu}\}$	Number of vehicles of type u assigned to route p, $u \in \mathcal{U}, p \in \mathcal{P}$
$\mathbf{W} = \{w_{rv}^{st\kappa}\}$	The portion of detour time of ride request r served by vehicle v, from time interval s to time interval t for scenario κ, $v \in \mathcal{V}, s,t \in \mathcal{T}, r \in \mathcal{R}^{\kappa}, \kappa \in \mathrm{K}$
$\overline{\mathbf{W}} = \{\overline{w}_{rv}^{st\kappa}\}$	Indicate if space reserved for ride request r from time interval s to time interval t for vehicle v for scenario κ, $v \in \mathcal{V}, s,t \in \mathcal{T}, r \in \mathcal{R}^{\kappa}, \kappa \in \mathrm{K}$
$\mathbf{W}'' = \{w_{rp}^{zz'}\}$	The portion of ride request r picked up in zone z and dropped off in zone z', for vehicles of route p, $p \in \mathcal{P}, z, z' \in \mathcal{Z}^{aug}, r \in \mathcal{R}$
$\boldsymbol{\rho}^{\text{I}} = \{\rho_c^{\text{I}}\}$	Vector of volume reliability measure for demand category $c \in \mathcal{C}$
$\boldsymbol{\rho}^{\text{II}} = \{\rho_{z+}^{\text{II}}, \rho_{z-}^{\text{II}}\}$	Vector of detour time reliability measure, defined for each zone $z \in \mathcal{Z}$

$\mathbf{Y}_{+(-)} = \{y^{z}_{cup+(-)}\}$	The portion of ride requests of category c picked up (dropped off) by vehicles of type u and route p in augmented zone z , $u \in \mathcal{U}, c \in \mathcal{C}, p \in \mathcal{P}, z \in \mathcal{Z}^{\text{aug}}$
$\bar{\mathbf{Y}}_{+(-)} = \{\bar{y}^{z}_{cup+(-)}\}$	The number ride requests of category c that their space reserved (released) in augmented zone z in vehicles of type u and route p , $u \in \mathcal{U}, c \in \mathcal{C}, p \in \mathcal{P}, z \in \mathcal{Z}^{\text{aug}}$
$\boldsymbol{\pi}$	Price factor to be optimized.
δ_c	The number of ride requests of demand category $c \in \mathcal{C}$ required to be served in phase-1
$\zeta^{\kappa}_{v} = \{\zeta^{\text{IJ}st,\kappa}_{v}\}$	The number of passengers traveling from node (I, s) to node (J, t) , $\forall(\text{I}, s), (\text{J}, t) \in \mathcal{N}, v \in \mathcal{V}, \kappa \in \mathcal{K}$
$\zeta'_{p} = \{\zeta^{zz'}_{p}\}$	The number of passengers traveling from zone z to z' by vehicles of route $p \in \mathcal{P}$, $\forall z, z' \in \mathcal{Z}^{\text{aug}}$

2.2. *Definitions of Temporal Variables and Constraints*

To consider passenger time constraints and time-dependent travel times, a zonal-based time-space network is used to formulate the model. A small zonal-based time-space network with three zones and three TIs is illustrated in Figure 2. Each *time-space network node* $(z,s) \in \mathcal{N}$ corresponds to a zone $z \in \mathcal{Z}$ and a time interval $s \in \mathcal{T}$. The time interval is represented by natural numbers 1,2,3,…, $|\mathcal{T}|$. Visiting node (z,s) in a route requires the vehicle to be in arbitrary location in zone z at the beginning of TI s . If the next zone to be visited in $s+1$ is z' , the time in s can be utilized for either detour in z or z' , or travel from z to z' , providing flexibility. At the beginning of any TI, the vehicle can be on the way to pick up or drop off passengers or at the centroid. For simplicity, we assume that a vehicle may visit at most two zones in any TI. We define the route notation such that each alphabet refers to the zone where the vehicle is in within one TI, and the last digit corresponds to the starting TI of the route. For example, ABC1 is the route that starts from zone A at the beginning of the first TI (denoted by the last digit), or zone-time node (A,1). Then, it starts traveling to zone B before the beginning of the second TI, and arrives at zone B within the second TI, i.e., visiting zone-time nodes (B,1) and (B,2), and finally starts traveling to zone C before the beginning of the third TI, and arrives at zone C within the third TI, i.e., visiting zone-time nodes (C,2) and (C,3). If a vehicle stays in a zone for three consecutive TIs, the zone is repeated in the notation. For example, the notation ABB2 represents a route that starts in zone A at time interval 2 (denoted by the last digit), travels to zone B before time interval 3, and stays in zone B until time interval 4. The zone-time nodes visited are (A,2), (B,2), (B,3), and (B,4). The *time interval duration*, denoted by $\bar{T}$, can be arbitrarily defined, say 10 minutes. Note that the time interval cannot be too long as it would increase the time required in a TI, which might result in bus holding within the zone in case of low demand to comply with the scheduling. Also, it is assumed that the travel time between each pairs of adjacent zones z and z' is shorter than the time interval duration, i.e. $\tau^{s}_{zz'} < \bar{T}$ for all TI s . The assumption is justified for small zones relatively to the TI duration.

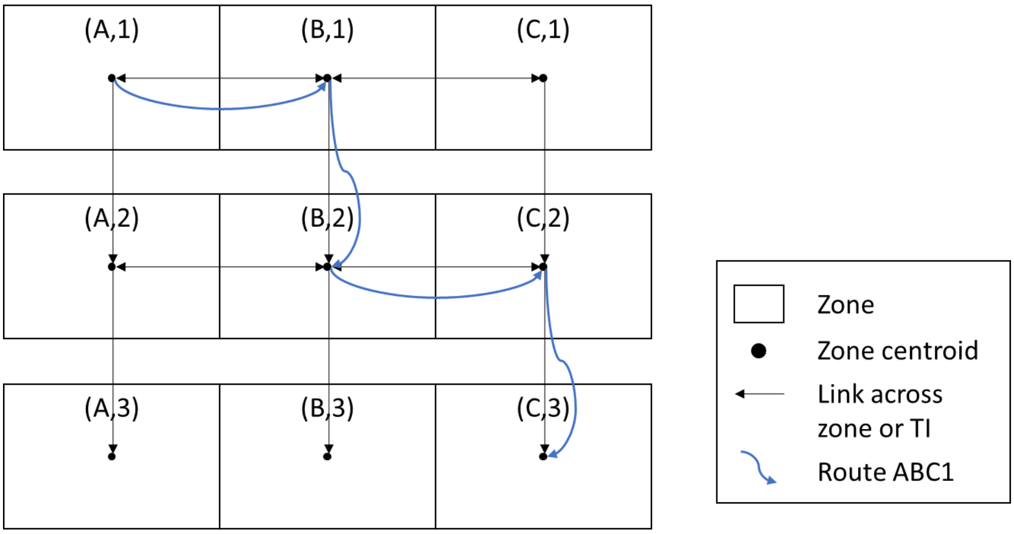

Figure 2. A small zonal-based time-space network and route ABC1

For a time interval s when a certain vehicle $v \in \mathcal{V}$ travels between zone A and zone B, the time interval s, with a duration of $\overline{T}$, is divided into three parts: the first part is the allowable detour time in the first zone (e.g. $\overline{t}_{v\text{A}}^{s}$); the second part is the travel time between adjacent zones (e.g. τ_{AB}^{s}); and the third part is the allowable detour time in the second zone (e.g. $\overline{t}_{v\text{B}}^{s}$), as illustrated in Figure 3 for route ABC1. The allowable detour time is not required to be used up; the slack time from not completely utilizing the time can be considered as waiting time of the vehicle to match the schedule. For each route p, z_{ps}^{first} and z_{ps}^{second} denote the first and second zone visited within the TI s, respectively. For example, $z_{\text{ABC1,1}}^{\text{first}} = \text{A}$ and $z_{\text{ABC1,1}}^{\text{second}} = \text{B}$. In general, if x_{pv} indicates the route of flexible bus, the split of time between allowable detour time and interzonal travel time can be expressed as,

$$\overline{t}_{vz_{ps}^{\text{first}}}^{s} + \overline{t}_{vz_{ps}^{\text{second}}}^{s} = \sum_{p \in \mathcal{P}} x_{pv} \left(\overline{T} - \tau_{z_{ps}^{\text{first}} z_{ps}^{\text{second}}}^{s} \right), \forall v \in \mathcal{V}, \forall s \in \mathcal{T} \quad (1)$$

The time-dependent travel time $\tau_{zz'}^{s}$ is exogenously given, but the detour times are variables to be optimized. If the vehicle is not required to travel to another zone in some intervals, the entire TI is allocated to the allowable detour time within the zone, as illustrated in the 3rd time interval ($t = 3$) in Figure 3. Note that the maximum allowable detour time is not required to be used up. In this case, speed reduction or bus holding control may be applied to maintain the bus schedule (Liang et al., 2016; Sánchez-Martínez et al., 2016).

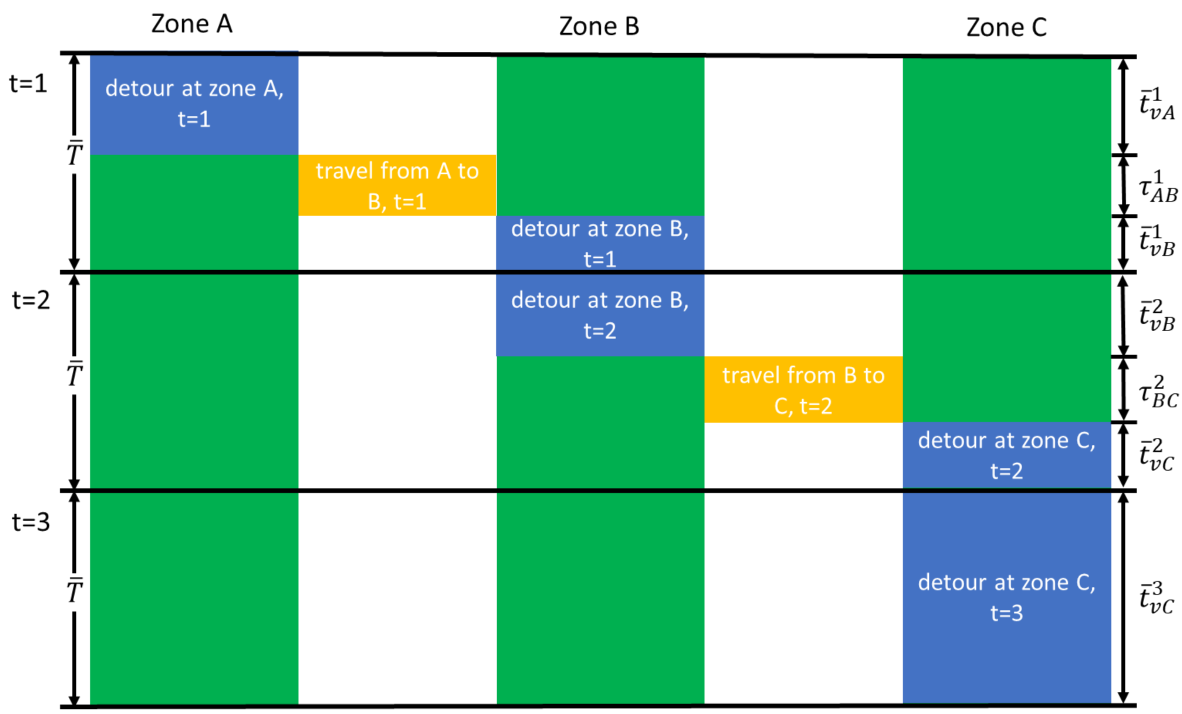

Figure 3. Illustration of allowable detour time (in blue) and travel time (in yellow) for route ABC1

2.3. *Detour Time across Time Intervals and Vehicle Seat Reservation*

To provide flexibility for the vehicles, the detour time to pick-up or drop-off ride requests may span across time intervals. In other words, consecutive pick-up (drop-off) time intervals can share the detour time, in the condition that the detour time required is sufficient to serve the passenger. For every scenario κ, we define $\mathbf{W} = \left\{ w_{rv}^{st,\kappa} \right\}$ as the portion of the detour time for pick-up and drop-off in the time interval pair (s,t). For example, for vehicle v and any time interval pairs s,t, the portion of detour time is defined as four variables $w_{rv}^{s,t,\kappa}, w_{rv}^{s+1,t,\kappa}, w_{rv}^{s,t-1,\kappa}, w_{rv}^{s+1,t-1,\kappa} \in [0,1]$. The pick-up detour time of ride request r in scenario κ allocated to TIs s and $s+1$ are denoted by $w_{rv}^{s,t,\kappa} + w_{rv}^{s,t-1,\kappa}$ and $w_{rv}^{s+1,t,\kappa} + w_{rv}^{s+1,t-1,\kappa}$, respectively; $w_{rv}^{s,t-1,\kappa} + w_{rv}^{s+1,t-1,\kappa}$ and $w_{rv}^{s,t,\kappa} + w_{rv}^{s+1,t,\kappa}$ of the drop-off detour time are allocated to each of TIs $t-1$ and t, respectively. As the service must be served entirely, if being served, the sum of the portions of detour time spent must be either 0 or 1.

Other than addressing the detour times, passengers waiting for pickup must be guaranteed a seat. To this end, $\bar{\mathbf{W}} = \left\{ \bar{w}_{rv}^{st,\kappa} \right\}$ is introduced to ensure a seat is reserved once the pick-up starts (i.e. at the first TI s where $w_{rv}^{st,\kappa} > 0$), and is released when the drop-off finishes (i.e. at the last TI t where $w_{rv}^{st,\kappa} > 0$). Each element $\bar{w}_{rv}^{st,\kappa}$ is a binary variable indicating whether a seat is reserved for ride request r on vehicle v from TI s to TI t in scenario κ. Under the condition that a vehicle is in the origin zone z in TIs s and $s+1$, and in the destination zone z' in TIs $t-1$ and t, Figure 4 provides two examples of seat reservation. The time span is indicated by the green lines on the right. To prevent the detour time from being allocated separately, which is not realistic, we assume the pick-up detour time must be allocated within one TI after the first TI of seat reservation, and the drop-off detour time must be allocated within one TI before the TI of seat reservation. For example, if $\bar{w}_{rv}^{st,\kappa} = 1$, the origin detour time can only be allocated in TIs s and $s+1$, and the destination detour time can only be allocated in TIs $t-1$ and t, as illustrated in Figure 4. This setting is formally expressed as follows,

$$w_{rv}^{st,\kappa} \le M(\bar{w}_{rv}^{s,t,\kappa} + \bar{w}_{rv}^{s-1,t,\kappa} + \bar{w}_{rv}^{s,t+1,\kappa} + \bar{w}_{rv}^{s-1,t+1,\kappa}), \forall v \in \mathcal{V}, (s,t) \in \mathcal{T}_r, r \in \mathcal{R}^\kappa, \kappa \in \mathcal{K}(\boldsymbol{\pi}) \quad (2),$$

where, for every ride request r, the set of feasible TI pairs $\mathcal{T}_r$ is defined as,

$$\mathcal{T}_r = \left\{(s,t): s,t \in \mathcal{T} \mid \max(s, e_r^{\text{arr}}) \le t \le \ell_r^{\text{arr}}, e_r^{\text{dep}} \le s \le \ell_r^{\text{dep}}\right\}, \forall r \in \mathcal{R}^{\kappa}, \kappa \in \mathcal{K}(\boldsymbol{\pi}) \quad (3),$$

where e_r^{arr}, ℓ_r^{arr}, e_r^{dep} and ℓ_r^{dep} denote the earliest arrival TI, the latest arrival TI, the earliest departure TI, and the latest departure TI, respectively.

From constraints (2) and (3), if $\overline{w}_{rv}^{st,\kappa}$ equals 1, only $w_{rv}^{s,t,\kappa}, w_{rv}^{s+1,t,\kappa}, w_{rv}^{s,t-1,\kappa}, w_{rv}^{s+1,t-1,\kappa}$ can be greater than zero to split the detour time required. For other TI pairs, the respective $w_{rv}^{s,t,\kappa}$ are always zero, as it is impossible to split the detour time into two or more inconsecutive TIs. Moreover, if the vehicle is not in the origin zone in TI s, or not in the destination zone in TI t, the respective $w_{rv}^{s,t,\kappa}$ is not specified. Also, s must be less than or equal to t according to (3) as time is irreversible. Specifically, under the condition of route and ride request of Figure 4, if $\overline{w}_{rv}^{s+1,t-1,\kappa}$, instead of $\overline{w}_{rv}^{st,\kappa}$, equals one and there are sufficient allowable detour times in TI $s+1$ and $t-1$, then $w_{rv}^{s+1,t-1,\kappa} = 1$ as it is the only feasible solution. Since vehicle is not in the origin zone in TI $s+2$, so that $w_{rv}^{s+2,t-2,\kappa}$ and $w_{rv}^{s+2,t-1,\kappa}$ equal 0. Also, as the vehicle is not in the destination zone in TI $t-2$, $w_{rv}^{s+1,t-2,\kappa}$ must be equal to 0.

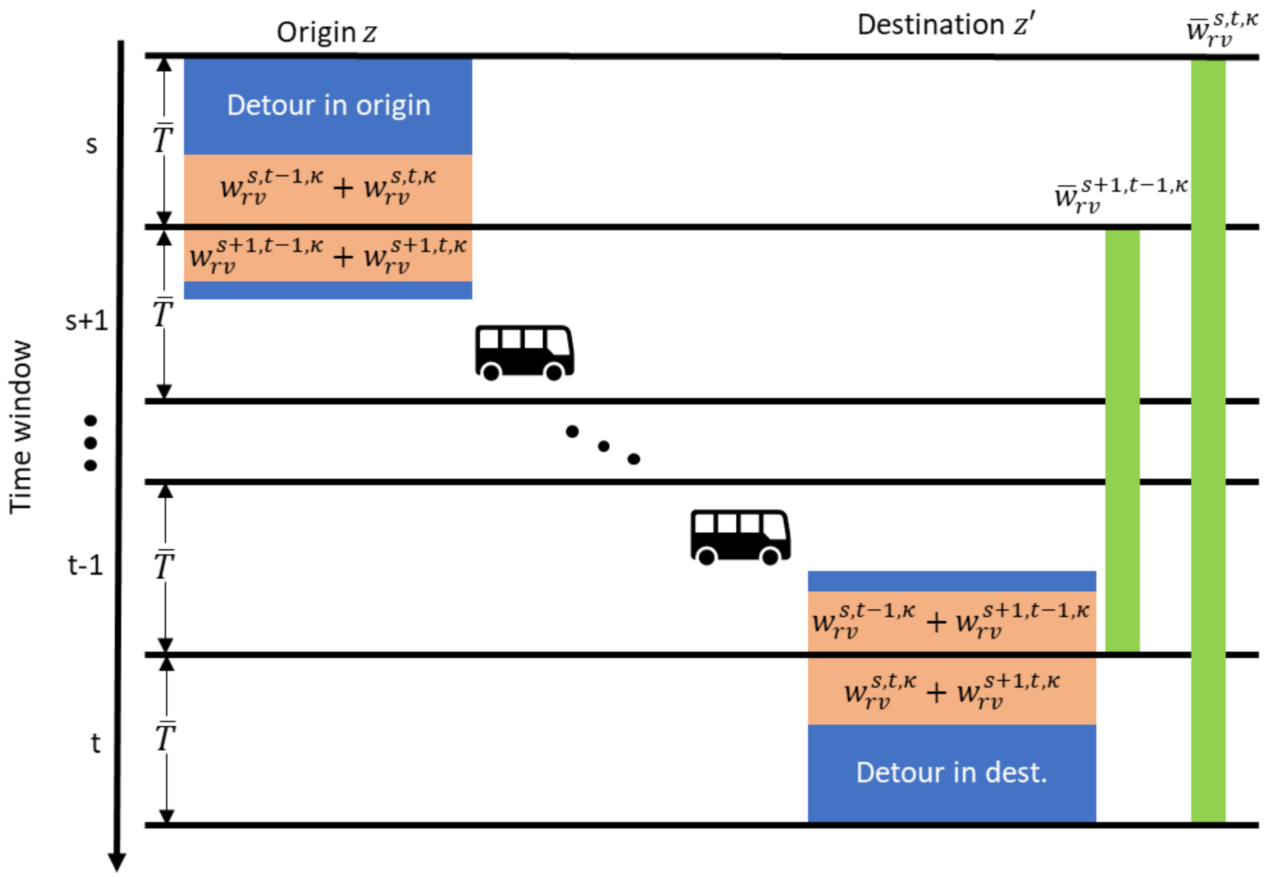

Figure 4. Illustration of splitting detour time of ride request for scenario κ when $\overline{w}_{rv}^{st,\kappa} = 1$

2.4. *Two-stage Stochastic Formulation with Recourse*

By using the time-space network structure, a two-stage stochastic problem with recourse and decision-dependent variables is formulated. It is also possible to formulate the problem as an integrated model by including all the conditions for each possible scenario as constraints. However, this will create a huge set of constraints as every scenario needs to have its own set of constraints, making it very difficult, if at all possible, to solve. In the two-stage stochastic problem formulated in this study, there are two stages of decisions: before the demand realization, a here-and-now decision is made on the vehicle-to-route assignment; and after the demand realization, a wait-and-see decision is made on the passenger-to-vehicle assignment.

$$\text{Stage 1: } \max_{\mathbf{X},\mathbf{W},\boldsymbol{\pi}} \bar{Q}(\mathbf{X}) - \sum_{v\in\mathcal{V}} \sum_{p\in\mathcal{P}} c_{pv}^{\text{fixed}} x_{pv} \tag{4}$$

subject to

$$\sum_{p\in\mathcal{P}} x_{pv} \le 1, \forall v\in\mathcal{V} \tag{5}$$

$$x_{pv} \in \{0,1\}, \forall p\in\mathcal{P}, v\in\mathcal{V} \tag{6}$$

where,

$$\text{Stage 2: } \bar{Q}(\mathbf{X}) = \max_{\mathbf{W},\bar{\mathbf{W}}} \sum_{\kappa\in\mathcal{K}(\boldsymbol{\pi})} p_\kappa \left\{ \sum_{r\in\mathcal{R}^\kappa} \left[g_r(\boldsymbol{\pi}) - c_r^{\text{adhoc}} \left(1 - \sum_{v\in\mathcal{V}} \sum_{(s,t)\in\mathcal{T}_r} w_{rv}^{st,\kappa} \right) \right] \right\} \tag{7}$$

subject to (1), (2), (3), and,

$$\sum_{r\in\mathcal{R}^\kappa} \left(\alpha_{r+}^z \tau_{r+} \sum_{t=\max(s,e_r^{\text{arr}})}^{\ell_r^{\text{arr}}} w_{rv}^{st,\kappa} + \alpha_{r-}^z \tau_{r-} \sum_{t=e_r^{\text{dep}}}^{\min(\ell_r^{\text{dep}},s)} w_{rv}^{ts,\kappa} \right) \le \bar{t}_{vz}^{\,s}, \forall (z,s)\in\mathcal{N}, v\in\mathcal{V}, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{8}$$

$$\zeta_v^{\text{IJ}st,\kappa} = \sum_{r\in\mathcal{R}^\kappa} n_r \alpha_{r+}^{\text{I}} \alpha_{r-}^{\text{J}} \bar{w}_{rv}^{st,\kappa}, \forall (\text{I},s),(\text{J},t)\in\mathcal{N}, v\in\mathcal{V}, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{9}$$

$$x_{pv}\mathbf{B}_p \boldsymbol{\zeta}_v^\kappa \le cap_v \mathbf{1}_{m_p-1}, \forall v\in\mathcal{V}, p\in\mathcal{P}, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{10}$$

$$\sum_{p\in\mathcal{P}} x_{pv} \left(1 - 1_{z_{ps}^{\text{first}}=O_r}\right)\left(1 - 1_{z_{ps}^{\text{second}}=O_r}\right) w_{rv}^{st,\kappa} = 0, \forall v\in\mathcal{V}, (s,t)\in\mathcal{T}_r, r\in\mathcal{R}^\kappa, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{11}$$

$$\sum_{p\in\mathcal{P}} x_{pv} \left(1 - 1_{z_{pt}^{\text{first}}=D_r}\right)\left(1 - 1_{z_{pt}^{\text{second}}=D_r}\right) w_{rv}^{st,\kappa} = 0, \forall v\in\mathcal{V}, (s,t)\in\mathcal{T}_r, r\in\mathcal{R}^\kappa, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{12}$$

$$\sum_{v\in\mathcal{V}} \sum_{(s,t)\in\mathcal{T}_r} w_{rv}^{st,\kappa} \le 1, \forall r\in\mathcal{R}^\kappa, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{13}$$

$$\sum_{v\in\mathcal{V}} \sum_{(s,t)\in\mathcal{T}_r} \bar{w}_{rv}^{st,\kappa} \le 1, \forall r\in\mathcal{R}^\kappa, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{14}$$

$$w_{rv}^{st,\kappa} \ge 0, \forall v\in\mathcal{V}, (s,t)\in\mathcal{T}_r, r\in\mathcal{R}^\kappa, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{15}$$

$$\sum_{(s,t)\in\mathcal{T}_r} w_{rv}^{st.\kappa} \in \{0,1\}, \forall v\in\mathcal{V}, r\in\mathcal{R}^\kappa, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{16}$$

$$\bar{w}_{rv}^{st,\kappa} \in \{0,1\}, \forall v\in\mathcal{V}, (s,t)\in\mathcal{T}_r, r\in\mathcal{R}^\kappa, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{17}$$

$$\bar{w}_{rv}^{st,\kappa} = 0, \forall v\in\mathcal{V}, (s,t)\notin\mathcal{T}_r, r\in\mathcal{R}^\kappa, \kappa\in\mathcal{K}(\boldsymbol{\pi}) \tag{18}$$

The objective function (4) is to maximize the expected total profit of the operator. The first term $\bar{Q}(\mathbf{X})$ is the stage 2 objective value as in (7). The second term is the total operating cost of flexible buses, where c_{pv}^{fixed} is the operating cost of vehicle v in route p and $\mathbf{X} = \{x_{pv}\}$ is the vehicle-to-route assignment matrix, with x_{pv} equal to 1 if vehicle v is used in route p. Constraint (5) states that each vehicle can serve at most one route. Condition (6) is the binary condition for x_{pv}.

Given elastic demand realizations $\kappa\in\mathcal{K}(\boldsymbol{\pi})$ based on price factor $\boldsymbol{\pi}$, the objective function (7) of Stage 2 is to maximize the expected average revenue minus the average ad hoc cost. For ride request $r\in\mathcal{R}^\kappa$, $g_r(\boldsymbol{\pi})$ and c_r^{adhoc} denote the revenue function given fare and the ad hoc service cost, respectively. The ad hoc cost c_r^{adhoc} for providing ad hoc service is incurred if ride request r is not served by flexible bus, i.e. $\sum_{v\in\mathcal{V}}\sum_{(s,t)\in\mathcal{T}_r} w_{rv}^{st,\kappa}=0$ for all scenario κ. Constraint (8) is the detour time constraint, defined for each zone visit per TI. The first term in the left hand side is the detour time for picking up, and the second term is the detour time of dropping off, at node (z,s) for ride request r. The origin (destination) detour time is denoted by $\tau_{r+(-)}$ for ride request r , and $\alpha_{r+(-)}^{z}$ indicates whether ride request r starts (ends) in zone $z\in\mathcal{Z}$. Flexible buses must detour within the allowed detour time $\bar{t}_{vz}^{s}$, which is a variable constrained by (1). Equation (9) calculates the passenger flow $\zeta_{v}^{\mathrm{IJ}st,\kappa}$ from node (I,s) to node (J,t). As explained in the previous section, $\bar{w}_{rv}^{st\kappa}$ indicates if vehicle space is reserved between TI s and t, or from node (O_r,s) to (D_r,t), exclusive of the last node, for ride request r in vehicle v. The last node is excluded because the passengers do not board at the last stop, which is the terminus stop of the route. n_r is the number of riders of ride request r. Constraint (10) is the capacity constraint defined for every zone or TI crossing on a schedule. $\mathbf{B}_p$ is the converting matrix from the passenger flow vector to the number of passenger on the vehicle while leaving each zone or closing TI, detailed in (Lee et al., 2021b). cap_v is the vehicle capacity of vehicle v, and m_p is the sum of the number of TIs and the number of zones visited by route p. It is assumed that routes do not revisit zone. Also, vehicles are assumed to first drop-off passengers before picking up for each zone-time node (z,s) to prevent overcapacity. Equation (11) ensures vehicles appear in the origin zone of a ride request in the designated TI, in order to pick the ride request up. Precisely, if a ride request r is picked up by vehicle v at time s, it must be in the origin zone O_r at TI s, i.e. $z_{ps}^{\text{first}}=O_r$, indicated by $1_{z_{ps}^{\text{first}}=O_r}$, or $z_{ps}^{\text{second}}=O_r$. Recall that $z_{ps}^{\text{first(second)}}$ is the zone which route p visits in the first (second) part of the TI s. Similarly, equation (12) enforces the destination location constraint. Constraints (13) and (14) specify that every ride request is served at most once, and constraint (15) ensures $w_{rv}^{st,\kappa}$ is non-negative. In other words, it is impossible to assign a ride request to an infeasible TI by the definition of $w_{rv}^{st,\kappa}$. To reduce the complexity, $w_{rv}^{st,\kappa}$ and equations (11)-(14) are defined only in feasible TIs. Constraint (16) requires each ride request to be served completely by the same vehicle, if being served. Coupling with (2) and (3), the space must be reserved for respective TI if the ride request is served. $\mathbf{W}$ is defined only in feasible TI pairs $\mathcal{T}_r$ and $\bar{\mathbf{W}}$ can only be positive for feasible pick-up and drop-off TI pairs $(s,t)\in\mathcal{T}_r$. In other words, enforcing the time window constraints, as specified in (17) and (18).

3. Solution Method

The two-stage stochastic problem above is a stochastic binary program, which is NP-hard. Therefore, a customized solution method is required to solve the problem. First, a decomposition method via the concept of reliability measures is developed in Section 3.1, which separates the problem into two phases to be solved by a gradient-based solution method. Section 3.2 describes problem simplifications, including elimination of time label and time window constraints, formulation of relaxed problems, and development of insertion heuristics, which allow the approach applicable for large-scale problems. The organization of this section is presented in Figure 5.

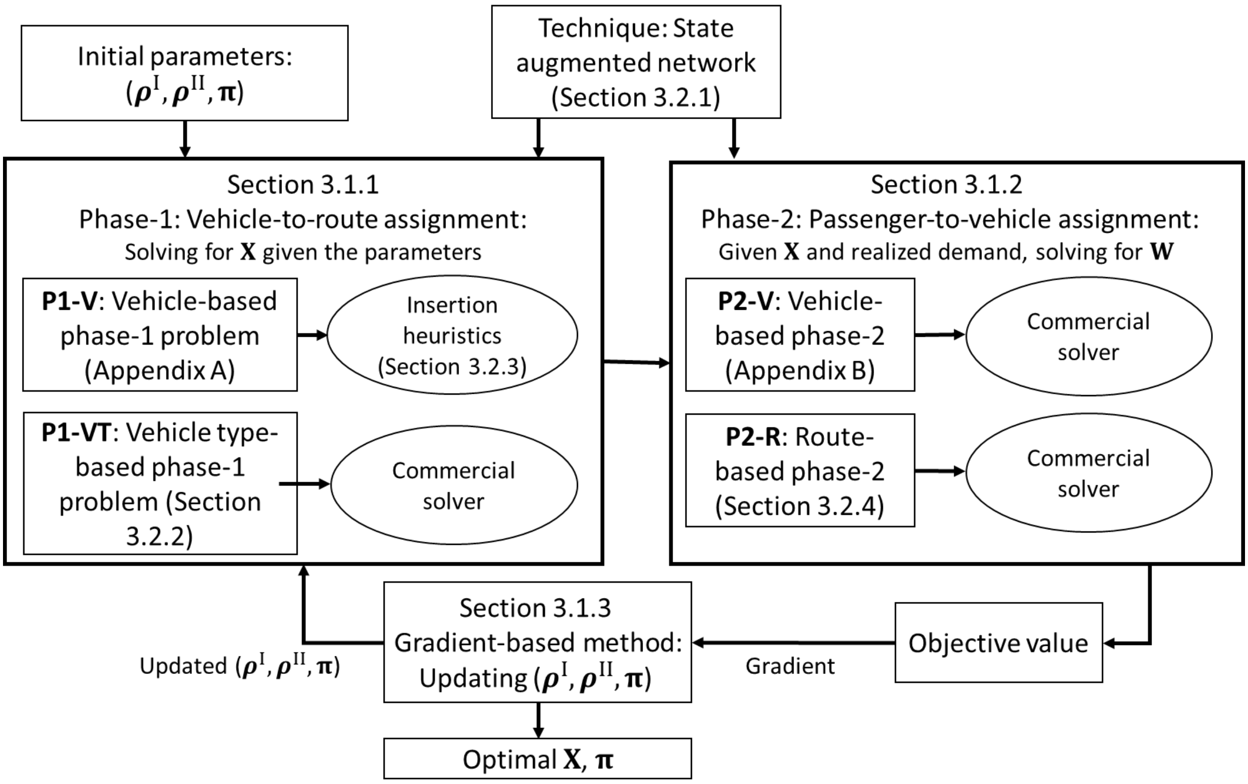

Figure 5. The gradient-based optimization algorithm and its improvements

3.1. *Service Reliability-based Solution Method*

Service reliability measures are introduced to decompose the two-stage stochastic problem into vehicle-to-route assignment (phase-1) and passenger-to-vehicle assignment (phase-2) problems. The ride requests on the same OD zones are separated into demand categories $c \in \mathcal{C}$ based on the number of passengers traveling together, their time window constraints, which are aggregated to form the total random demand. Volume reliability measure $\boldsymbol{\rho}^{\mathrm{I}}$ is introduced to determine the passenger volume, and detour time reliability measure $\boldsymbol{\rho}^{\mathrm{II}}$ is introduced to generate the provisional detour time for planning the service in phase 1.

3.1.1. Vehicle-to-route Assignment Problem (Phase-1)

The vehicle-to-route assignment problem assigns flexible bus vehicles to routes, while satisfying the designated passenger demands with provisional detour times, according to the corresponding specified reliability measures. To quantify the stochasticity of the ride requests, they are aggregated by demand categories, defined as follows.

Definition 1 (Demand category). A demand category c is an 8-tuple, defined by set of origin augmented zones $\mathcal{O}_c$, set of destination augmented zones $\mathcal{D}_c$, demand volume distribution Δ_c, number of passengers n_c, earliest departure TI e_c^{dep}, latest departure TI ℓ_c^{dep}, earliest arrival TI e_c^{arr}, and latest arrival TI ℓ_c^{arr}.

In practice, the ride request is categorized by origin zone, destination zone, number of passengers, and time window constraints. The demand volume distribution is then inferred from the historical data. Then, in the phase-1 problem, given the volume reliability ρ_c^{I}, the number of ride requests needed to be served is generated by inversing the random variable Δ_c for each ride request $c \in \mathcal{C}$,

$$\delta_c = \inf\left\{\eta \in \mathbb{Z} \mid \Pr(\Delta_c(\boldsymbol{\pi}) \leq \eta) \geq \rho_c^{\mathrm{I}}\right\}, \forall c \in \mathcal{C} \tag{19}$$

This equation is generic for any demand volume distribution as it rounds up the number of ride requests to the nearest integer, and is used in both versions of the phase-1 problem. The vehicle-based phase-1 problem directly is derived from the two-stage stochastic problem, as detailed in Appendix A, which is hard to solve for large-scale instances. Therefore, an insertion heuristic is proposed in Section 3.2.3. A relaxation by grouping vehicles with the same capacity, as *vehicle type*, is also formulated in Section 3.2.2.

3.1.2. Passsnger-to-vehicle Assignment Problem (Phase-2)

With the vehicle-to-route assignment **X** in phase-1 fixed, the passenger-to-vehicle assignment is optimized to minimize the ad hoc cost given the demand $\mathcal{R}^{\kappa}$ in phase-2. Note that price elasticity has been taken into account when the demand is realized, and the revenue function $g_r(\boldsymbol{\pi})$ is fixed for each ride request. Given the realized demand, each ride request $r \in \mathcal{R}^{\kappa}$ is either served by a flexible bus without any extra cost or assigned to an ad hoc service with an extra cost c_r^{adhoc} charged by the operator. The vehicle-based formulation P2-V, which is equivalent to stage 2 of the two-stage problem given a vehicle assignment, is provided in Appendix B. The equivalence between stage 2 of two-stage problem and P2-V is also discussed in Appendix B. A similar phase-2 formulation with integral relaxation is formulated based on routes in Section 3.2.4, which significantly reduces the solution time by reducing the complexity of the problem while transforming the problem into a linear program.

3.1.3. Gradient-based Solution Algorithm

Various methods have been applied to solve the route planning problem with passenger assignment on the time-space network. Some used a combination of column generation and Lagrangian relaxation (Steinzen et al., 2010); others used an integer programming heuristic to determine a feasible integer solution. To solve the VRP with stochastic ride request, Ulmer et al. (2018) used an anticipatory time budgeting heuristic, which simulated the problem realization to approximate the values for every vector of point of time and free time budget, and then achieved an approximation of an optimal decision policy. They used a dynamic lookup table to partition the vector space to the approximation process. Ulmer et al. (2019) solved a VRP with stochastic ride requests by an offline-online approach for approximate dynamic programming. They combined the offline value function approximation (VFA) with online rollout algorithms to anticipate the future value, which is stochastic in temporal and spatial ways.

In this paper, a gradient solution approach is applied to solve the route planning problem, as outlined in Algorithm 1 (An and Lo, 2016, 2014; Huang et al., 2018; Lee et al., 2021a; Li et al., 2018; Lo et al., 2013). This algorithm optimizes the reliability measures and price factor by the gradient-based Adam method. It is a stochastic gradient descent method that does not require preselected scenarios, and converges to a local optimum where the gradient is zero. Users may choose the appropriate formulations and solution methods for phase-1 and phase-2 subproblems after considering the trade-off between accuracy and solution speed. Note that the number of iterations is not fixed, and the algorithm will stop when the percentage change of the profit is less than 1%.

Algorithm 1. Reliability-based gradient solution approach

Input: Initial parameters $\boldsymbol{\rho}$ including reliability measures and price factor, iteration $k=0$

Output: Optimal vehicle assignment $\mathbf{X}$

Step 1: Solve the phase-1 problem to obtain vehicle assignment $\mathbf{X}$ based on the designated reliability measures $\boldsymbol{\rho}$ and price factor π.

Step 2: Solve the phase-2 problem for a batch of scenarios and obtain the average profit.

Step 2.1: Save $\mathbf{X}$ and the profit P_k if it is the highest among previous iterations.

Step 3: Convergence check: If the percentage change of total profit is less than a threshold ϵ, say 1%, terminate the algorithm and **return** the best solution $\mathbf{X}^*$.

Step 4: Update reliability measures and price factor

Step 4.1: Estimate the partial derivative of the parameter by perturbation analysis. For each parameter ρ_i,

Step 4.1.1: Increase the value of ρ_i by 0.05 and solve the vehicle-to-route assignment problem until having a ρ_{ik} that generates a schedule different from the current schedule.

Step 4.1.2: Solve phase-2 problem repetitively and obtain the total profit P_{ik}.

Step 4.1.3: Estimate the partial derivative by $\frac{\Delta P_k}{\Delta \rho_i} = \frac{P_k - P_{ik}}{\rho_i - \rho_{ik}}$

Step 4.2: Form a vector of gradient by combining the partial derivatives.

Step 4.3: Update the reliability measures and price factor by Adam method (Kingma and Ba, 2014), $k \leftarrow k+1$, go to step 1.

3.2. *Relaxations and Heuristics*

3.2.1. Eliminating Time Label and Time Window Constraints

A state augmented network is constructed to eliminate the time label in the formulations for simplicity. The zones are augmented by the number of TIs, forming a new set of augmented zones $\mathcal{Z}^{\text{aug}}$. For example, if zone A is visited in two consecutive TIs, it is represented by augmented zone A1 and A2 in the first and the second TIs, respectively. With respect to the route illustrated in Figure 2, the augmented route of ABC1 becomes A1, B1, B2, C2, C3. As the zones are augmented, passengers can access or egress vehicles in zones that are augmented from the origin or destination zones between their earliest departure times and latest arrival times. After the augmentation, the possible origin and destination of the demand category $c \in \mathcal{C}$, which is the aggregation of ride requests with the same OD zone, number of passengers, and time window constraint, are denoted by $\mathcal{O}_c$ and $\mathcal{D}_c$, respectively. Therefore, the locations at time window constraints (11) and (12) are dropped to reduce the complexity of the problem, as they are satisfied by defining $\mathcal{O}_c$ and $\mathcal{D}_c$ if and only if it is feasible. By deciding the route and the augmented origin and destination where augmented ride requests are picked up and dropped off, TIs of picking up and dropping off the ride request can be deduced. Thus, the time label is omitted.

The following formulations are pre-processed to ensure that the feasible decision variables satisfy location constraints (11), (12), and time window constraints. Moreover, the equations and constraints are limited to feasible TIs to capture the time window constraints, as opposed to the original two-stage stochastic formulation which explicitly states the zone visit constraints. Note that under the pre-processing procedure, the latest departure time and earliest arrival time constraints as in some dial-a-ride literature can be handled by removing zones that violate the time constraints from $\mathcal{O}_c$ and $\mathcal{D}_c$ (Cordeau and Laporte, 2007). Conversely, relaxations of any time window constraints can be handled by using a larger set of $\mathcal{O}_c$ and $\mathcal{D}_c$.

3.2.2. Vehicle Type-based Phase-1 Problem (P1-VT)

A relaxation based on vehicle type is formulated to significantly reduce the number of variables as the formulations for different vehicle types are similar except their vehicle capacity constraints and detour time constraints, which aggregate the demand and detour time.

$$\min_{\mathbf{X}',\mathbf{Y}_+,\mathbf{Y}_-,\bar{\mathbf{Y}}_+,\bar{\mathbf{Y}}_-} C_f(\boldsymbol{\rho}^{\mathrm{I}},\boldsymbol{\rho}^{\mathrm{II}},\boldsymbol{\pi}) = \sum_{u\in\mathcal{U}}\sum_{p\in\mathcal{P}} c_{pu}^{\text{fixed}} x_{pu} \tag{20}$$

subject to (19), and,

$$\tau_{z+(-)} = \inf\{\eta \mid \Pr(\Lambda_{z+(-)} \le \eta) \ge \rho_{z+(-)}^{\mathrm{II}}\}, \forall z \in \mathcal{Z}^{\text{aug}} \tag{21}$$

$$\sum_{p\in\mathcal{P}} x_{pu} \le N_u, \forall u \in \mathcal{U} \tag{22}$$

$$\sum_{u\in\mathcal{U}}\sum_{p\in\mathcal{P}}\sum_{z\in\mathcal{O}_{cp}} y_{cup+}^{z} = \delta_c, \forall c \in \mathcal{C} \tag{23}$$

$$\mathbf{B}_{p+}\bar{\mathbf{y}}_{up+} + \mathbf{B}_{p-}\bar{\mathbf{y}}_{up-} \le x_{pu} cap_u \mathbf{1}_{m_p-1}, \forall u \in \mathcal{U}, p \in \mathcal{P} \tag{24}$$

$$\sum_{c\in\mathcal{C}_p}\left(y_{cup+}^{z}\tau_{z+} + y_{cup-}^{z}\tau_{z-}\right) \le \bar{t}_{uzp}, \forall z \in \mathcal{Z}_p, u \in \mathcal{U}, p \in \mathcal{P} \tag{25}$$

$$\bar{t}_{u\mathrm{I}p} + \bar{t}_{uz_{p\mathrm{I}}p} = x_{pu}\left(\bar{T} - \tau_{\mathrm{I}p}\right), \forall \mathrm{I} \in \mathcal{Z}_p, u \in \mathcal{U}, p \in \mathcal{P} \tag{26}$$

$$\sum_{z\in\mathcal{O}_{cp}} y_{cup+}^{z} = \sum_{z\in\mathcal{D}_{cp}} y_{cup-}^{z}, \forall c \in \mathcal{C}_p, u \in \mathcal{U}, p \in \mathcal{P} \tag{27}$$

$$\sum_{z'\in\mathcal{O}_{cp,z}} y_{cup+}^{z'} \le \sum_{z'\in\mathcal{O}_{cp,z}} \bar{y}_{cup+}^{z'}, \forall z \in \mathcal{O}_{cp}, c \in \mathcal{C}_p, u \in \mathcal{U}, p \in \mathcal{P} \tag{28}$$

$$\sum_{z'\in\mathcal{D}_{cp,z}} y_{cup-}^{z'} \ge \sum_{z'\in\mathcal{D}_{cp,z}} \bar{y}_{cup-}^{z'}, \forall z \in \mathcal{D}_{cp}, c \in \mathcal{C}_p, u \in \mathcal{U}, p \in \mathcal{P} \tag{29}$$

$$\sum_{z\in\mathcal{O}_{cp}} y_{cup+}^{z} \in \mathbb{N}_0, \forall c \in \mathcal{C}_p, u \in \mathcal{U}, p \in \mathcal{P} \tag{30}$$

$$x_{pu} \in \mathbb{N}_0, \forall u \in \mathcal{U}, p \in \mathcal{P} \tag{31}$$

$$y_{cup+}^{z} \ge 0, \forall z \in \mathcal{O}_{cp}, c \in \mathcal{C}_p, u \in \mathcal{U}, p \in \mathcal{P} \tag{32}$$

$$y_{cup-}^{z} \ge 0, \forall z \in \mathcal{D}_{cp}, c \in \mathcal{C}_p, u \in \mathcal{U}, p \in \mathcal{P} \tag{33}$$

$$\bar{y}_{cup+}^{z}, \bar{y}_{cup-}^{z'} \in \mathbb{N}_0, \forall z \in \mathcal{O}_{cp}, z' \in \mathcal{D}_{cp}, c \in \mathcal{C}_p, u \in \mathcal{U}, p \in \mathcal{P} \tag{34}$$

$$\mathcal{O}_{cp,z} = \left\{i \in \mathcal{O}_{cp} \mid TW(z) - TW(i) \ge 0\right\}, \forall z \in \mathcal{O}_{cp}, c \in \mathcal{C}_p, p \in \mathcal{P} \tag{35}$$

$$\mathcal{D}_{cp,z} = \left\{i \in \mathcal{D}_{cp} \mid TW(z) - TW(i) \ge 0\right\}, \forall z \in \mathcal{D}_{cp}, c \in \mathcal{C}_p, p \in \mathcal{P} \tag{36}$$

This formulation optimizes the vehicle type-based assignment x_{pu} for the designated demand volume and detour time under given reliability measures while considering the passengers who are picked up $\mathbf{Y}_+ = \left\{y_{cup+}^{z}\right\}$ and dropped off $\mathbf{Y}_- = \left\{y_{cup-}^{z}\right\}$. $y_{cup+(-)}^{z}$ is the total number of passengers of demand category c picked up (dropped off) for all vehicles of type u and route p in zone z.

Equation (21) specifies the detour time given the detour time reliability. Note that the detour time distribution $\Lambda_{z+(-)}$ is equal for every augmented zone z with the same physical zone. Therefore, the detour time reliability measure $\rho_{z+(-)}^{\mathrm{II}}$ is in lockstep between every augmented zone z with the same physical zone. Constraint (22) ensures the vehicle assigned does not exceed the total number of vehicles. Equation (23) guarantees the number of ride requests required can be fulfilled by the flexible bus schedule planned, given the service reliability δ_c. Constraint (24) enforces capacity constraint, where $\mathbf{B}_{p+(-)}$ is the converting matrix from the number of passengers pick-up (drop-off) to the number of passengers boarded (egressed) per augmented zone. Matrix $\mathbf{B}_{p+(-)} = \left\{b_{zcp+(-)}^{z'}\right\}$ denotes the number of spaces reserved (released) from the start of the route p till zone $z \in \mathcal{Z}_p$, given $\bar{y}_{cup+(-)}^{z}$, the number of ride requests of demand category c requiring space reserved (released). $b_{zcp+}^{z'} = n_c$ if z' is after z for route p, 0 otherwise. Similarly, to calculate the cumulative passengers egressed $b_{zcp-}^{z'} = -n_c$ if z' is after z for route p, 0 otherwise. If zone z is not in $\mathcal{Z}_p$, a very large number is set for $b_{zcp+}^{z'}$, and a very small negative number is set for $b_{zcp-}^{z'}$ to ensure $\overline{y}_{cup+(-)}^{z} = 0$. Therefore, $\mathbf{B}_{p+}\overline{\mathbf{y}}_{up+} + \mathbf{B}_{p-}\overline{\mathbf{y}}_{up-}$ is a vector of space reserved per augmented zone visit. Constraint (25) limits the detour time to the maximum allowable detour time, while equation (26) defines the maximum detour time in the zones along routes. $z_{p\mathrm{I}}$ is the next augmented zone to be visited from zone I for route p. Constraint (27) ensures vehicles drop off the same number of passengers as they pick up. Constraints (28) and (29) require sufficient space reserved for the passengers. Constraints (30) and (31) are integral constraints. Equations (32) and (33) define the feasible region of the variable $y_{cup+(-)}^{z}$, where $\mathcal{C}_p$ is the set of ride requests that is feasible to be served by any vehicle of route p, and the OD of ride request should be traversed by route p within the time window specified. Condition (34) ensures the space reserved and released are integral. Condition (35) aggregates the origin zones and (36) aggregates the destination zones with earlier time windows. Correspondingly, the drop-off equation (29) is reversed to ensure the space released is aggregately smaller than the portion of ride requests dropped off. The relationship between P1-V and P1-VT is formally established in Proposition 1.

Proposition 1. Any feasible solution of P1-V has an equivalent solution in P1-VT.

Proof. See Appendix C.

Although P1-VT is a relaxation of P1-V, it can still be used since the reliability measures are decision variables. The reduction of the number of vehicles assigned can be compensated by increasing the reliability measures so that more vehicles are added back to respective routes.

3.2.3. Insertion Heuristic for Phase-1 Problem

Due to a large number of feasible routes and demand categories, P1-VT, in which the dimensions of decision variables depend on the feasible route set, still takes a long time to solve, if at all possible. Therefore, an insertion heuristic algorithm is devised to generate routes for vehicles while ensuring the feasibility of the vehicle-based formulation P1-V in Appendix A. The input of the algorithm includes *vehicle configurations*, which state the capacity for each vehicle $v \in \mathcal{V}$, and *demand category configurations*, which include every component of the demand category and the required demand volume. After preparing the respective input parameters, the outline of the algorithm is given as Algorithm 2.

Algorithm 2: Route generation and vehicle assignment algorithm

Input: Zonal network, vehicle configurations, demand category configurations, and pick-up and drop-off detour time $\tau_{z+(-)}$ per zone

Output: A set of route $\mathcal{P}$ and vehicle assignment x_{pv}.

for each demand category $c \in \mathcal{C}$
  for each vehicle $v \in \mathcal{V}$
    Determine the feasibility of inserting ride request of c to the vehicle v
    if feasible
      Calculate the maximum amount of ride request of c can be added to v
      Calculate the least insertion cost per ride request
      Determine additional zones required to be added if inserting some ride request of demand category c
      Assign the designated number of ride requests to vehicles with the least insertion costs per ride request, until all δ_c requests required are assigned.

Append the additional zones to the vehicle route, if necessary.

The main idea of Algorithm 2 is to assign the required demand of demand categories to the vehicle with the least additional cost. If both capacity and detour time constraints allow, vehicles of the following three types of routes can serve demand category $c \in \mathcal{C}$. First, if the vehicle has a route traversing some zones in $\mathcal{O}_c$ and $\mathcal{D}_c$, there is no cost to serve the additional demand of c. Second, if the vehicle traverses some zones in $\mathcal{O}_c$ but none of the zone in $\mathcal{D}_c$, it is required to travel to one of the $\mathcal{D}_c$ which incurs additional operating cost from the current last zone of the route to $\mathcal{D}_c$. Third, if the vehicle traverses neither the origin zones nor the destination zones, both origin and destination zones are required to be added to the route. The additional cost is comprised of the cost from the current last zone to the origin and the cost from the origin to the destination. Therefore, the third situation rarely happens, as the additional cost is higher than deploying idle vehicles. Only the cost from the origin to the destination is incurred to deploy an idle vehicle for ride requests of demand category c. Note that this insertion algorithm inserts demand categories to get onboard the latest feasible TI and egress in the earliest feasible TI to reduce the passenger load along the route.

Sometimes, a vehicle can serve more ride requests by staying longer in a zone, since the additional staying time provides more time for picking up or dropping off additional passengers. If so, by staying in a zone for two consecutive TIs, the average insertion costs per ride request can potentially be reduced. Therefore, the average insertion costs per ride request by staying for one and two TIs are calculated for each vehicle, and the duration of TI with the least insertion cost per ride request is picked. Finally, after comparison, the vehicles with the least average insertion costs per ride request are used to serve the ride requests, until every ride request is assigned.

3.2.4. Route-based Relaxation of Passenger-to-vehicle Assignment Problem

Integral constraints are relaxed, and the time window constraints are eliminated through the feasible passenger-to-vehicle assignment beforehand. Consequently, the problem becomes a linear program, which can be solved in a very short time. Since the integral requirement of space is relaxed, setting $\bar{\mathbf{W}} \approx 0$ already ensures that the capacity constraint is never violated. We use the portion of serving request $\mathbf{W}'' = \left\{ w_{rp}^{zz'} \right\}$ as the portion of space reserved to substitute the capacity constraint. $w_{rp}^{zz'}$ is only defined for a feasible set of origin zone $\mathcal{O}_{rp}$ and destination zone $\mathcal{D}_{rp}$. The route-based passenger-to-vehicle problem (P2-R) is formulated as:

$$\min_{\mathbf{W}''} Q_\kappa = \sum_{r \in \mathcal{R}^\kappa} c_r^{\text{adhoc}} \left(1 - \sum_{p \in \mathcal{P}} \sum_{z \in \mathcal{O}_{rp}} \sum_{z' \in \mathcal{D}_{rp}} w_{rp}^{zz'} \right) \tag{37}$$

subject to,

$$\sum_{r \in \mathcal{R}^\kappa} \left(\sum_{z' \in \mathcal{D}_{rp}} \tau_{r+} w_{rp}^{zz'} + \sum_{z' \in \mathcal{O}_{rp}} \tau_{r-} w_{rp}^{z'z} \right) \le \bar{t}_{pz}, \forall p \in \mathcal{P}, z \in \mathcal{Z}^{\text{aug}} \tag{38}$$

$$\bar{t}_{p\mathrm{I}} + \bar{t}_{pz_{p\mathrm{I}}} = \left(\bar{T} - \tau_{\mathrm{I}p} \right) \sum_{v \in \mathcal{V}} x_{pv}, \forall \mathrm{I} \in \mathcal{Z}_p, p \in \mathcal{P} \tag{39}$$

$$\zeta_p^{zz'} = \sum_{r \in \mathcal{R}^\kappa | i \in \mathcal{O}_{rp} \wedge j \in \mathcal{D}_{rp}} n_r w_{rp}^{ij}, \forall p \in \mathcal{P}, z, z' \in \mathcal{Z}^{\text{aug}} \tag{40}$$

$$\mathbf{B}_p \zeta'_p \le \mathbf{1}_{m_p - 1} \sum_{v \in \mathcal{V}} x_{pv} cap_v, \forall p \in \mathcal{P} \tag{41}$$

$$\sum_{p \in \mathcal{P}} \sum_{z \in \mathcal{O}_{rp}} \sum_{z' \in \mathcal{D}_{rp}} w_{rp}^{zz'} \le 1, \forall r \in \mathcal{R}^\kappa \tag{42}$$

$$w_{rp}^{zz'} \ge 0, \forall z \in \mathcal{O}_{rp}, z' \in \mathcal{D}_{rp}, r \in \mathcal{R}^\kappa, p \in \mathcal{P} \tag{43}$$

Objective function (37) minimizes the total ad hoc service cost by determining $\mathbf{W}''$, which is the portion of ride request r served by any vehicle of route p from zone z to z'. Constraint (38) is the detour time constraint to ensure that the total detour time at any zone is less than the maximum allowable detour time $\overline{t}_{pz}$. Equation (39) is defined to calculate the detour time at each augmented zone $z \in \mathcal{Z}_p$ along route p. Equation (40) calculates the passenger flow for each OD pair of the route. Constraint (41) ensures each zone-time node visit does not violate the capacity constraint after aggregation for each route p. Constraint (42) enforces every ride request to be picked up at most once. The integral constraint is relaxed to reduce the solution time. Constraint (43) specifies the non-negativity nature of the passenger flow. The relationship between P2-R and P2-V is established in Proposition 2.

After solving the phase-2 problem for scenario κ, the profit is given by,

$$\sum_{r \in \mathcal{R}^{\kappa}} g_r(\boldsymbol{\pi}) - Q_{\kappa} \tag{44}$$

Assume that all scenarios occur with the same probability, the total expected profit is calculated by,

$$\frac{1}{|\mathcal{K}|} \sum_{\kappa \in \mathcal{K}} \left[\sum_{r \in \mathcal{R}^{\kappa}} \left(g_r(\boldsymbol{\pi}) \right) - Q_{\kappa} \right] - C_f \qquad (45),$$

where $|\mathcal{K}|$ is the number of scenarios and C_f is the objective function of the phase-1 problem.

Proposition 2. The solution in P2-R is a lower bound of the solution in P2-V.

Proof. See Appendix C.

4. Numerical Studies

4.1. *Quality of the Solution Method in a Small Example*

To illustrate the problem, a small example is introduced in this subsection based on the simple zonal-based time-space network depicted in Figure 2. Three routes are considered, including ABC1, AABC1, and BC1. There is only one vehicle type, with vehicle capacity of 8 people. The flexible bus travel cost between zones is \$30, and the travel time between zones is 3 minutes. The ad hoc service cost per request is assumed to be 90% of the flexible bus cost. Each time interval is assumed to have a duration $\overline{T}$ of 10 minutes. Two demand categories are considered in this small example, as shown in Table 1. To illustrate the model flexibility while maintaining tractability, two possible detour times are available for each pick-up and drop-off operation, and three possible demand volumes may be generated for each demand category. The number of possible scenarios for this small example is 217404, which makes it hard to enumerate all the possible scenarios while calculating the optimal solution. For simplicity, fare and price elasticity of demand are not considered in this small example. Therefore, this small example only minimizes the operating expenses.

The global optimal vehicle type assignment is in the set $\{x = (x_1, x_2, x_3) : x_1, x_2, x_3 \in [0, 4]\}$. Even in the worst-case scenario, where the highest demand volume and maximum allowable detour time are realized, all passengers can be served by assigning 4 vehicles to ABC1 and 3 vehicles to BC1. As AABC1 is an extension of ABC1, any vehicle of route ABC1 can be replaced by AABC1. Therefore, the global optimal solution can be obtained approximately by enumerating the set. To illustrate the performance of the decomposed model, P2-V, which is equivalent to the original two-stage model, is used for benchmarking to find the global optimal solution. 300 instances of passenger realization are generated and solved for each vehicle type assignment to evaluate the solution.

Table 1. The OD, demand volume distribution and detour distribution for the problem

OD	Demand volume distribution (Δ_e)	Origin detour distribution (Λ_z)	Destination detour distribution (Λ_z)	Number of passengers (n_e)	Earliest departure TI	Latest arrival TI
AC	$p_{\Delta_1}(x)=\begin{cases}0.2, x=3\\0.4, x=8\\0.4, x=11\end{cases}$	$p_{\Lambda_A}(x)=\begin{cases}0.5, x=1\\0.5, x=2\end{cases}$	$p_{\Lambda_C}(x)=\begin{cases}0.5, x=1.5\\0.5, x=2.5\end{cases}$	1	1	3
BC	$p_{\Delta_2}(x)=\begin{cases}0.25, x=5\\0.5, x=7\\0.25, x=9\end{cases}$	$p_{\Lambda_B}(x)=\begin{cases}0.5, x=1.5\\0.5, x=3\end{cases}$	$p_{\Lambda_C}(x)=\begin{cases}0.5, x=1.5\\0.5, x=2.5\end{cases}$	2	1	2

The optimal solution with the lowest objective value is $x=[2,0,1]$, which schedules two vehicles to route ABC1 and one vehicle to route BC1, with an objective value of $214.6, with a 95% confidence interval of $6.6, or 3%. The second best vehicle schedule is $x=[3,0,0]$, yielding an objective value of $216.9 and a 95% confidence interval within $3.2 or 1.5%.

The proposed reliability-based gradient solution approach is then applied to this instance from several initial reliability measures. Note that the P2-R relaxation formulation proposed is used. When applying the solution approach, 50 independent passenger realizations are generated for calculating the objective value by (45). To test the effectiveness of the proposed solution approach under different initial reliability measures, 16 initial reliability measures are selected. Every run is converged within 180 seconds. From the result, nine instances yield a vehicle assignment of [2,0,1], and seven instances yield a vehicle assignment of [3,0,0]. Both [2,0,1] and [3,0,0] can be the optimal solution because of the randomness of the sample cost. In short, this small example demonstrates that the proposed reliability-based gradient solution approach can be used to solve the proposed scheduling problem with a good quality solution.

4.2. *Numerical Examples on five-zone Scenario*

To demonstrate the advantage of schedule-based planning, it is applied to a five-zone network as depicted in Figure 6. Unless specified otherwise, there are seven TIs, each with a duration of 10 minutes. The base travel time (in minutes) is 4 minutes for links AB, BC, CD, and AD; and 3 minutes between any zone and E. The total route cost between zones is the sum of the distance travel cost and the time cost. The time cost is $2/minute, which accounts for the driver's wage and vehicle idle cost. The time-dependent travel time is the base travel time times the travel time factor for each TI, as given in Table 2. The demand categories, with their OD pairs, numbers of passengers, and request volume distributions per 10 minutes are given in Table 3. Probability distribution $TN(\mu,\sigma^2,a,b)$ denotes the truncated normal distribution for a normal distribution of mean μ and standard deviation σ that is bounded by $[a,b]$. These demand categories are repeated for the first 4 TIs so that the earliest departure TIs and latest arrival TIs are [1,2], [2,3], [3,4], and [4,5], respectively. The detour time distributions per ride request for zones are given in Table 4. The total ad hoc service cost, including the time cost, is assumed to be 1.5 times of the distance travel cost of flexible bus. The elastic demand function with respect to price is taken to be $D=D_0e^{-0.6(\pi-1)}$, where D_0 is the demand volume variable in Table 3, and π is the price factor to be optimized. Note that to simplify the problem, the price factor is now a scalar. On the other hand, the fare revenue of ride request r is stated as,

$$g_r(\boldsymbol{\pi})=\pi c_r^{\text{adhoc}} \tag{46},$$

which uses the cost of ad hoc service c_r^{adhoc} as a factor for the fare. The demand volume distribution is assumed to be equal in every TI. For simplicity, all vehicles have 7 seats. The possible route set is predetermined, including AB, BA, AEC, BC, AD, DA, AE, EA, BE, CD, DE, ED, CE, EC, EB, DEB, BED, and CB. Also, vehicles have an option to stay in the first zone for one more TI, hence including routes AAB, BBA, AAEC, BBC, AAD, DDA, AAE, EEA, BBE, CCD, DDE, EED, CCE, EEC, EEB, DDEB, BBED, and CCB. This option allows vehicles to spend one more TI in the first zone, compensating the time taken to travel to the next zone, which otherwise would reduce the detour time allowed in the first TI. The additional staying time in the first TI could significantly enhance the flexibility of picking up requests. While optimizing the parameters, the route-based passenger-to-vehicle assignment problem is used to approximate the phase-2 objective to reduce the solution time. As before, 50 passenger realizations are generated to approximate the objective value by (45).

After the solution approach is terminated, the vehicle schedule with the maximal profit is selected for evaluation. The objective value for the optimal vehicle schedule, which is the expected value of some sample possible passenger realization scenarios, is usually biased toward a higher profit than the true profit. It is because the higher biased objective is more likely to be the maximal objective by its high value. Moreover, the objective value itself is higher since the phase-2 objective is approximated based on P2-R formulation, which relaxes the actual passenger-to-vehicle matching problem. To obtain the objective value of the original two-stage problem, 200 independent passenger realizations are generated and the corresponding passenger-to-vehicle assignment problems as formulated by P2-V are solved until a 2% gap. Note that this 2% gap is only used for this evaluation. The decomposition of the problem for routes and demand, which improves the efficiency of the solution approach as introduced in Lee et al. (2021b), is applied to all numerical studies.

Table 2. Travel time factor per time interval

Time interval	1	2	3	4	5	6	7
Travel time factor	1.6	0.4	1.6	0.5	0.9	1	1

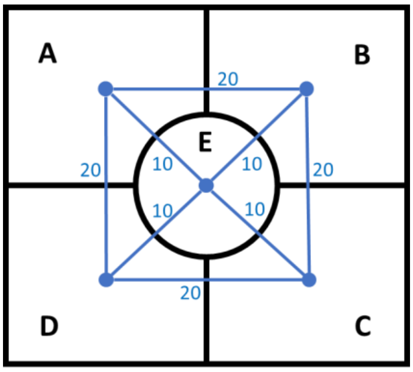

Figure 6. Five-zone network and the distance travel cost between zones (numbers in blue)

Table 3. OD, volume stochasticity, and number of passengers for demand categories

OD	Request volume distribution (Δ_e)	Number of passengers (n_e)	OD	Request volume distribution (Δ_e)	Number of passengers (n_e)
AB	$TN(3,4,0,+\infty)$	1	DE	$TN(5,4,0,+\infty)$	1
BA	$TN(3,4,0,+\infty)$	2	ED	$TN(4,9,0,+\infty)$	2
AC	$TN(3,4,0,+\infty)$	1	CE	$TN(2,4,0,+\infty)$	1
BC	$TN(3,4,0,+\infty)$	3	EC	$TN(2,1,0,+\infty)$	2
AD	$TN(4,4,0,+\infty)$	2	EB	$TN(2,9,0,+\infty)$	1
DA	$TN(4,4,0,+\infty)$	1	DB	$TN(6,1,0,+\infty)$	1
AE	$TN(4,9,0,+\infty)$	1	BD	$TN(4,4,0,+\infty)$	3
EA	$TN(2,1,0,+\infty)$	3	CB	$TN(5,9,0,+\infty)$	2
BE	$TN(6,16,0,+\infty)$	2	EA	$TN(2,1,0,+\infty)$	1

CD	$TN(3,4,0,+\infty)$	1	

Table 4. The detour time distributions per zone

Zone	Detour time distribution (Λ_z)
A	$TN(1.5,1,0,+\infty)$
B	$TN(2,1,0,+\infty)$
C	$TN(2,1.5,0,+\infty)$
D	$TN(1.5,1,0,+\infty)$
E	$TN(2,1,0,+\infty)$

4.2.1. Effect of Considering Time-dependent Travel Time

This section evaluates the effect of considering time-dependent travel time. Two instances for flexible bus schedule planning are considered: Instance 1 does not consider time-dependent travel time while scheduling flexible bus; i.e. it assumes the travel times are the same as those in TI 1; Instance 2 takes time-dependent travel time into account. As shown in Table 5, the average flexible bus profit by considering time-dependent travel time is $5944, while the counterpart is $5874 without considering time-dependent travel time. The fare revenue received is similar, but the total cost is reduced by 5.3% after considering time-dependent travel time. Considering time-dependent travel time yields a slight advantage while improving the authencity of the model. Therefore, time-dependent travel time is considered for the remaining numerical studies.

Table 5. Average service characteristics of Instance 1, Instance 2, and insertion heuristic

Results	Instance 1	Instance 2	Heuristic
Profit ($)	5874	5944	5909
Fare revenue ($)	10194	10037	10196
Flexible bus cost ($)	2463	2411	3036
Ad hoc cost ($)	1857	1681	1251
Price factor	1.98	2.12	2.02
Occupancy	49.3%	48.1%	41.3%
Computation time per iteration (s)	1902	2494	129
%1pax served by flexible bus	69.0%	67.9%	79.9%
%2pax served by flexible bus	70.5%	73.5%	75.1%

%1pax: The percentage of single passenger request; %2pax: The percentage of two passengers request

4.2.2. Result from the Insertion Heuristic and Convergence of Algorithm

The vehicle type-based formulation requires predefining a route set $\mathcal{P}$ before formulating the problem. As the set of all possible routes is too large, it is formidable to enumerate every route, resulting in a suboptimal solution. In this section, the insertion heuristic proposed in Section 3.2.3 is used to solve the illustrative example again, with the demand volume reliability measures updated in lockstep per OD. The optimized profits, revenues, costs, price factors, and occupancies of Instance 1, Instance 2, and insertion heuristics are shown in Table 5. The resultant profit by using the insertion heuristic is $5909, which is slightly lower than the profit obtained from the vehicle type-based formulation. A smaller ad hoc cost is incurred for the heuristic than those of Instance 1 and Instance 2, as more flexible buses are deployed so that fewer passengers required ad hoc services. The computation time per iteration is about 5.2% of the computing time of vehicle-based formulation and commercial solver, enabling exploration of larger applications. Figure 7 depicts the convergence of scores for heuristics and normal instances. It is found that all instances improved the objective value and converged. Although some of the instances converged to local minima with a lower objective value, 17 of 20 instances converged in a local optimum within 10% of the best optimal profit found. Note that the score shown in Figure 7 is taken while performing the solution method, which is not adjusted by another 200 scenarios as other tables reported.

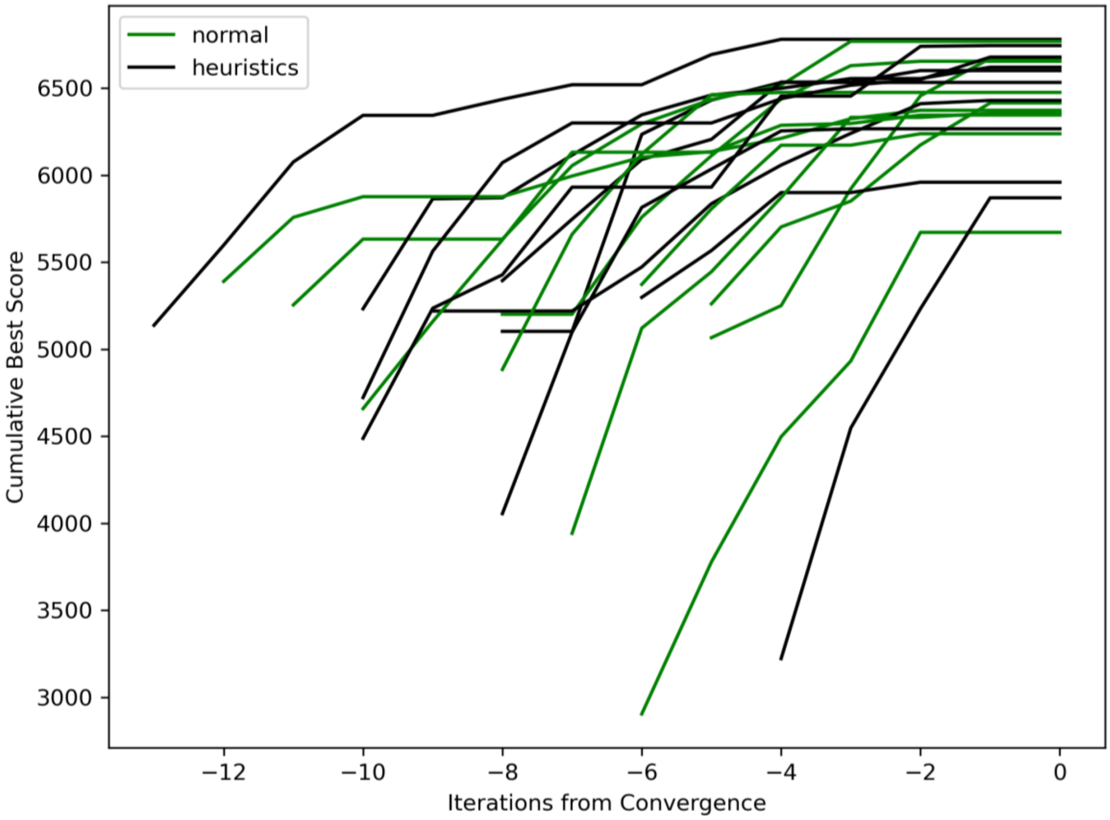

Figure 7. Convergence of the proposed solution algorithm

4.2.3. Comparison between Frequency-based vs Schedule-based Approaches

This subsection compares the frequency-based approach in Lee et al. (2021a) and the schedule-based approach proposed by planning the zonal-based flexible bus under the same constraints, except that the frequency-based approach omits the time window constraints, dynamic demand, and time-dependent travel time. The frequency-based flexible bus route is solved once for the demand of the entire period, the result is then converted to a schedule uniformly throughout the service time to observe the difference between the planned service under different circumstances. For the schedule-based approach, demand from eight consecutive time intervals, each with 10 minutes duration, is considered. Each ride request must be served within two TIs for traveling between adjacent zones, i.e. at most 20 minutes between pick-up and drop-off, and within three TIs for traveling between non-adjacent zones. Time-dependent travel time is considered in both case studies. The price elasticity of demand follows the previous examples. The first instance consists of uniform demand per time interval, under the demand distributions as shown in Table 3 for each time interval. The second instance considers a variable demand over time intervals, and the mean of the demand volume distributions are multiplied by the demand factors. The demand factors and time-dependent travel time are given in Table 6. The variable demand multipliers and travel time factors emulate a peak hour in which both the demand and travel time increase.

Table 6. Demand multipliers and travel time factors per time interval

Time interval	1	2	3	4	5	6	7	8	9	10	11	12
Demand multiplier	0.1	0.3	1.8	2.4	1.7	0.5	0.7	2	-	-	-	-
Travel time factor	1	1.3	1.5	1.6	1.7	1.7	1.6	1.4	1	1	1	1

P1-VT is used in this case study, with the same predetermined path set as in Section 4.2.1. To ensure fairness of the comparison, the predetermined path set is the same for both the frequency-based and schedule-based formulations. Before the comparison, the frequency-based vehicle assignment is divided into different TIs uniformly to ensure consistency of the service. For example, if there are 4 vehicles assigned to route AB during the frequency-based vehicle assignment, the route would be converted evenly to AB1, AB3, AB5, and AB7. However, if there is only 1 vehicle assigned to a route, it would be scheduled to depart in the first time interval. The adjusted schedules are then simulated for the passenger realizations generated under uniform distribution and time-varying demand distributions to obtain profits and costs.

The resultant objective value and departure per TI are shown in Table 7. The schedule-based result is better than the frequency-based result. Under the uniform demand, the schedule-based result achieves relatively even departures, with 753% higher profit. On the other hand, under time-varying dynamic demand, the profit of schedule-based result is 256.4% higher, while the total cost is 51.4% lower. Therefore, both schedule-based planning results outperform their frequency-based counterparts. The result also shows that the frequency-based result under uniform demand is worse than that under time-varying dynamic demand. It is because the frequency-based model result is designed for a higher demand volume generated from a lower price factor, but fails to strategically distribute the vehicles and address the additional time window constraint.

Table 7. Objective values and departures per time interval

	Uniform demand		Time-varying dynamic demand	
	Frequency-based	Schedule-based	Frequency-based	Schedule-based
Profit ($)	1100	8278	3375	8655
Fare Revenue ($)	15068	14487	16016	15158
Total cost ($)	16168	6209	12641	6502
Flexible bus cost ($)	13840	2645	9226	2202
Ad hoc cost ($)	2328	3564	3415	4300
Price factor	1.44	2.20	1.87	2.34
Departure at 1	22	5	19	7
Departure at 2	14	6	11	1
Departure at 3	20	5	11	3
Departure at 4	14	7	9	5
Departure at 5	20	5	16	7
Departure at 6	16	2	10	7
Departure at 7	18	7	12	6
Departure at 8	11	8	7	4

4.2.4. Sensitivity Analysis of the Time Interval

To benchmark the effect of the time interval, in the 40 minutes duration, we run a sensitivity analysis on the TI duration from 7 to 12 minutes. μ and σ of the request volume distribution are adjusted linearly to the TI duration. For example, the mean and S.D. of 8 minutes are 0.8μ and 0.8σ, respectively. If the 40-minute mark is not exactly between the TI, the remainder is also treated linearly. For example, for the TI duration of 9 minutes, the fifth TI, which is from 36th minute to 45th minute, has 4 minutes of demand, which makes the mean and S.D. to be 0.4μ and 0.4σ, respectively. On the other hand, the travel time between zones is set as 3 minutes for each pair of zones. For the computation time, a desktop computer with i9-10900 CPU with 32Gb ram is used. The insertion heuristic approach proposed is used. Note that the flexible bus cost is $2 per minute. The result is shown in Table 8.

Table 8. The service characteristics of different time interval durations

Time Interval Duration (mins)	7	8	9	10	11	12
Profit ($)	5300	6577	6735	7135	7106	7124
Fare Revenue ($)	9428	10570	10516	10609	10524	10604
Flexible bus cost ($)	3097	2424	2362	2357	2366	2524
Ad hoc cost ($)	1031	1569	1419	1117	1052	956
Number of vehicles deployed	80	59	53	50	45	51

Average detour time per zone (min)	2.07	2.68	3.03	3.49	3.71	4.20
Computation Time(s/iter)	96	108	74	100	104	94

A longer time interval allows more passengers to be picked up or dropped off. However, the time cost also increases, which increases the flexible bus cost, especially if the time interval is not utilized. Therefore, the profit first increases for as the time interval increases from 7 to 10 min, then the profit remains stable as the time intervals vary from 10 to 12 min. Moreover, the number of vehicles deployed decreases gradually. However, as the time interval is becoming longer, the average detour time per zone increases, thus the passengers experienced a longer trip. In short, longer time intervals increase the profit, but shorter time intervals reduce passengers' trip times. The computation time is about 100 seconds per iteration for all time intervals tested.

4.2.5. Comparison with Bus and Pure Demand-responsive Service

To illustrate their performance differences, we compare the detour time and cost with traditional bus and demand-responsive service such as taxi. Two scenarios are considered: traveling only by bus and traveling only by demand-responsive service. For the first setting, it is assumed that the bus station is located in the centroid of each zone, and the bus routes are provided with sufficient capacity. The total walking time is calculated by 5 times the vehicle detour time of the passengers. For demand-responsive service, we just apply the model to solve P2-V without having any flexible bus.

For bus service, the operating cost is $1489 to cover all the OD pairs, but the total walking time is 7675 minutes. For demand-responsive service with no walking or waiting time, the total cost to the operator is $14081. In the same condition, recap the result of Section 4.2.1, with a time interval duration of 10, the total cost is just $4320. The average total detour time is 498 minutes, which is significantly less than the access time to the traditional bus station. Therefore, it is shown that the flexible bus has a shorter total detour time required than the bus walking time, and also lower operating cost than pure demand-responsive service.

4.3. *Numerical Studies on DARP Instances*

As the schedule-based model proposed in this paper considers the location and TI of every ride request, it can be applied to different variations of DARP. To illustrate the applicability of the model to DARP instances, the benchmark instances provided by Ropke et al. (2007) are formulated as a deterministic vehicle assignment problem, and solved accordingly. As the demand volume per time unit is small in the benchmark instances, the specified earliest departure time, earliest arrival time, latest departure time, and latest arrival time are reduced by 90% to suit the flexible bus planning which is designed for higher demand volume. Each TI has a duration of 15 time units. For example, if the earliest departure time is 100 in the original setting, it becomes 10 in this numerical example, which lies in the first TI. As the benchmark instances are deterministic, the demand and detour distributions are deterministic while applying to the flexible bus model, so only the phase-1 problem is required. The service area is divided into four equal zones as shown in Figure 8. Only one vehicle type with capacity 7 is used. Similar to the benchmark instances that use the Euclidean distance as the travel time, the detour time is calculated by the Euclidean distance between zone centroid and the coordinate of the ride request location. Except for the intrazonal ride request which must be served by ad hoc service, the ride requests are assigned to vehicles to minimize the vehicle travel cost. The resultant cost and number of flexible buses used are shown in Table 9. It is found that the resultant cost is smaller compared to the benchmark result. Note that compared to the cost reported from DARP instances, in our model, the fleet size is larger and travel costs for detouring are not included.

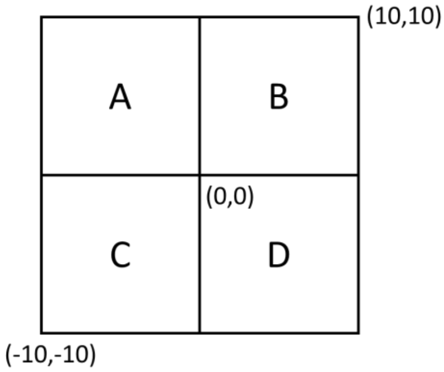

Figure 8. The zone division of the DARP instances

Table 9. The cost and number of vehicles used for each DARP instances

Instance	Vehicle Used	Cost	Instance	Vehicle Used	Cost
a2-16	9	186.02	b2-16	9	163.24
a2-20	7	172.76	b2-20	10	214.13
a2-24	10	216.33	b2-24	14	263.51
a3-24	7	203.06	b3-24	10	218.62
a3-30	11	234.58	b3-30	17	342.46
a3-36	14	331.42	b3-36	22	398.31
a4-32	12	288.09	b4-32	15	312.11
a4-40	15	341.04	b4-40	24	408.61
a4-48	16	365.38	b4-48	16	413.04
a5-40	13	297.46	b5-40	16	335.07
a5-50	14	391.05	b5-50	23	454.08
a5-60	21	470.47	b5-60	22	499.61
a6-48	20	408.27	b6-48	23	519.77
a6-60	25	510.36	b6-60	23	533.16
a6-72	25	554.75	b6-72	31	590.61
a7-56	20	443.33	b7-56	26	501.91
a7-70	24	528.79	b7-70	29	578.07
a7-84	28	618.84	b7-84	36	753.11
a8-64	22	484.46	b8-64	24	520.97
a8-80	31	627.00	b8-80	37	710.36
a8-96	33	786.69	b8-96	41	813.01

4.4. *Application for Chengdu, China Based on Real Data*

4.4.1. Improvement of the Schedule-based Model under Large Real Instances

To demonstrate real-world applicability of the model, we apply the zonal-based flexible bus scheduling model using ride-sourcing service data for Chengdu, China, between 6:30 am and 8:30 am, every Tuesday, Wednesday, and Thursday between 1st and 30th November 2016. Each ride request must be picked up and dropped off within 3 TIs, as the latest arrival TI is assumed to be two TIs from the earliest departure TI. The study area is within Second Ring Road, and the area is divided into 9 zones, as illustrated in Figure 9. Eight TIs, each of 15 minutes, are considered in the first example. The sensitivity analysis for other zone sizes and time interval durations is provided in the next subsection. 8-seater vehicles and 10-seater vehicles are provided, with a cost factor given in Table 10. To increase the solution speed, the reliability measures of each OD are optimized in lockstep. Precisely, the reliability measures of demand volume of the demand categories with the same OD pairs, across TIs and number of passengers, are synchronized and optimized as one variable. Also, the insertion heuristic approach and the route-based phase-2 formulation are used while optimizing the schedule. Following (Lee et al., 2021a), the parameters are listed in Table 10, with justification provided in Appendix D. The scheduling result is compared with the adjusted schedule obtained through the frequency-based approach.

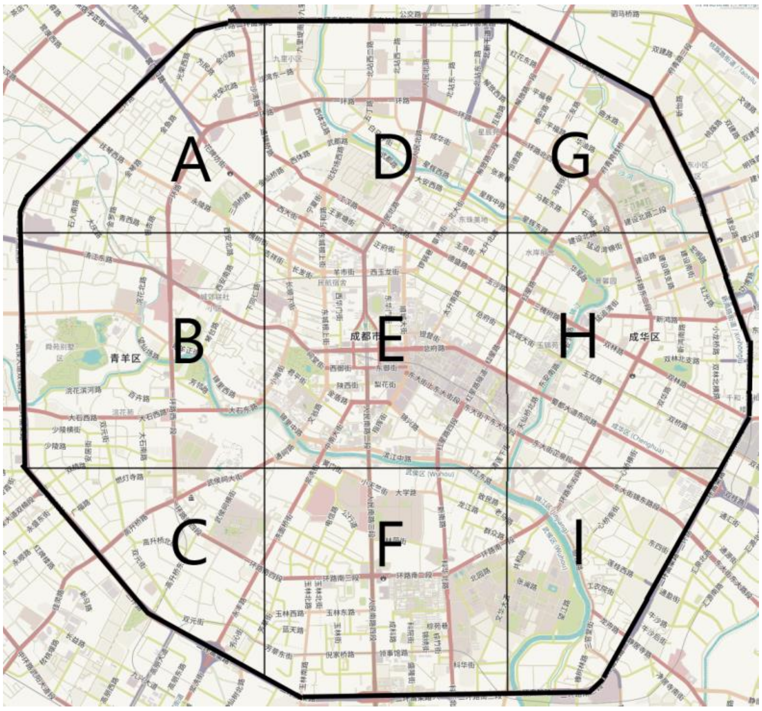

Figure 9. The study area in Chengdu, China, with a 3x3 zone division. Source: Map data OpenStreetMap contributors, CC BY-SA

Table 10. Parameters in the Chengdu scenario

Parameter	Value
Cost factor for 8-seaters α_8	1
Cost factor for 10-seaters α_{10}	1.1
Flexible bus cost per km (RMB¥)	0.88
Flexible bus cost per minute (RMB¥)	0.29
Ad hoc cost per km (RMB¥)	1.24
Ad hoc time cost per minute (RMB¥)	0
Travel time from zone z to z', $\tau_{szz'}$,	6.2 minutes, if zone z and z' has a common border;
for any time interval s	8.7 minutes, if zone z and z' only has a common corner
Fare revenue for ride request	$2\pi c_r^{\text{adhoc}}$

The resultant profit, revenue, cost, number of vehicles, percentage of passengers served, and solution time by using the frequency-based model and the schedule-based model are shown in Table 11. Based on the approximation by P2-R, the resultant profit is RMB¥13466. The flexible bus operating cost is RMB¥3931, and the ad hoc service cost is RMB¥921. 79.7% of the passengers are served by the flexible bus, while the rest are served by ad hoc services. The optimized flexible bus schedule deploys 356 vehicles to serve the passengers, and the resultant schedule is shown in Appendix E. The proposed scheduling model achieves a schedule that matches the demand volume better than the frequency-based model, as illustrated in Figure 10. For example, in the first TI, although there are only a few demands, the frequency-based model assigns a lot of vehicles in it and wastes their capacity. Despite deploying fewer vehicles, the portion of passengers served by flexible bus from the schedule-based result is higher than that of the frequency-based plan. Flexible bus cost increases by 16.4%, but 61.6% of ad hoc service cost is saved by using the schedule-based model. Therefore, the advantage of increasing profit and reducing ad hoc costs by the proposed schedule-based model is demonstrated under a real scenario. Note that the schedule-based model solution time is shorter, achieved by the proposed insertion heuristic algorithm and improved P2-R coding.

Table 11. The results of frequency-based and schedule-based model

Results	Frequency-based	Schedule-based	% Difference
Profit (RMB¥)	12849	13466	+4.8%
Fare revenue (RMB¥)	18624	18319	-1.7%
Flexible bus cost (RMB¥)	3376	3931	+16.4%
Ad hoc service cost (RMB¥)	2398	921	-61.6%
8-seater vehicles deployed	210	300	42.9%
10-seater vehicles deployed	0	56	-
%pax served by FB	54.8%	79.7%	-
Solution time (s)	11224	8841	-21.2%

%pax: Percentage of passengers

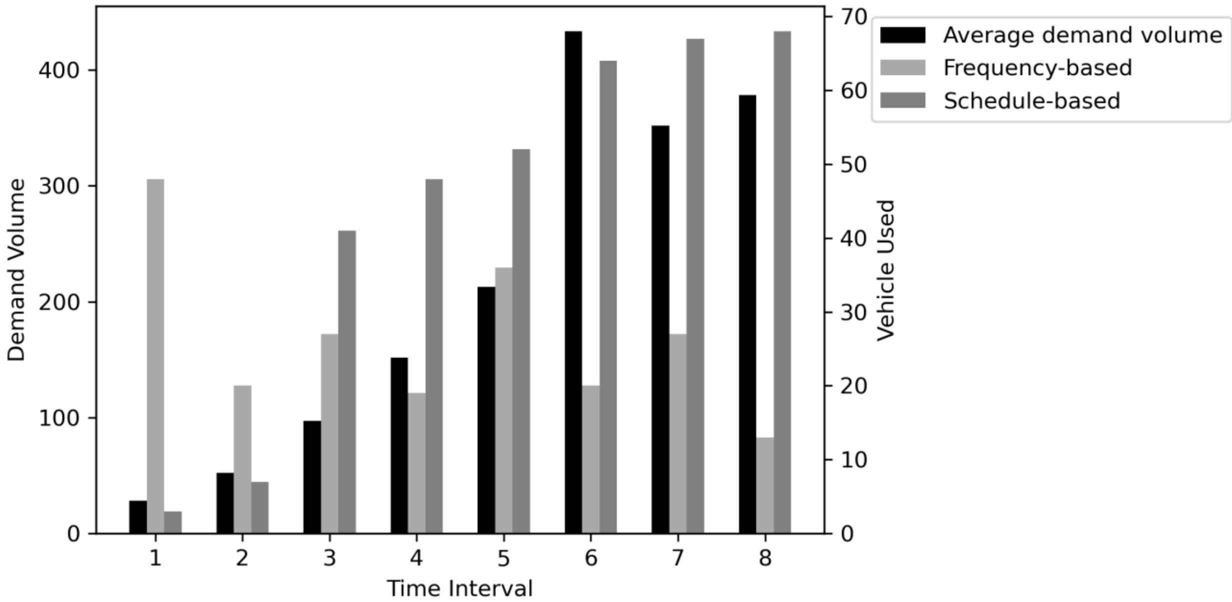

Figure 10. The demand volume and number of vehicles used in different time intervals

4.4.2. Sensitivity Analysis of the Time Interval Duration and Zonal Division

Four additional instances based on the same timeframe and area are created and solved. The instances include 4x3 zonal division with 15 minutes time intervals (4x3 zonal), 3x2 zonal division with 15 minutes time intervals (3x2 zonal), 3x3 zonal division with time interval durations $\overline{T}$ of 12 minutes (10 TIs), and 3x3 zonal division with time interval durations of 20 minutes (6 TIs). The results are summarized in Table 12. It was found that the 3x2 zonal division achieved the best profit, with the highest fare revenue. A lot more vehicles are deployed for the 10 TIs instance since the time interval is shorter so that more vehicles are required. Moreover, the profit is lower than the 6 TIs and the original case. It is because the time constraint for passengers depends on the duration of time interval, so for 10 TIs instance, the time constraint is tighter than the baseline scenario, which is tighter than 6 TIs instance. The demand is better aggregated with larger zones and longer TI durations, reflected by the fewer vehicles deployed, therefore, the profit of respective instances is higher than the base instance.

Table 12. Results from different instances

Results	4x3 zonal	3x2 zonal	10 TIs	6 TIs
Profit (RMB¥)	12009	15365	13058	14923
Fare revenue (RMB¥)	16909	19648	17809	19179
Flexible bus cost (RMB¥)	3902	2411	2352	3678
Ad hoc service cost (RMB¥)	997	1871	2400	579
8-seater vehicles deployed	261	164	258	185

10-seater vehicles deployed	4	3	6	9
%pax served by FB	85.2%	57.2%	50.0%	89.7%

%pax: Percentage of passengers

5. Conclusion

This paper proposed a schedule-based zonal-based flexible bus model, extending the frequency-based approach in Lee et al. (2021a), considering stochastic elastic demand volume, stochastic location, time-variant travel time, and time window constraints. The ride requests were categorized by OD zones, departure TI and arrival TI. The flexible bus schedules were determined by using a zonal-based time-space network approach with each node representing a time interval and a zone. To solve the problem, a gradient-based solution approach was implemented, which first separated the problem into vehicle-to-route assignment and passenger-to-vehicle assignment by reliability measures, and then optimized the reliability measures and prices to maximize the operating profit of the zonal-based flexible bus service. To create routes for large networks and increase the solution speed, an insertion heuristic was proposed to generate routes and match ride requests up to the required detour time and demand volume specified by the reliability measures. For small time-space networks, the route set can be predefined and solved by the vehicle type-based formulation of the vehicle-to-route assignment problem, but for large instances, the insertion heuristic is required to solve the problem. Relaxing the integral constraints of vehicle-based passenger-to-vehicle assignment formulation (P2-V), a linear route-based passenger-to-vehicle assignment problem (P2-R) was formulated, thereby improving the solution speed. By introducing a novel zonal-based time-space network, both formulations allow flexibility for vehicles to alter the time staying in zones, improving the adaptivity to stochastic demand, which is not allowed in most existing studies using normal time-space networks.

We evaluated the effectiveness of the schedule-based model through a few examples. First, an illustrative example demonstrated the ability of the proposed algorithm to converge to the optimal solution. Then, a five-zone example demonstrated that considering time-dependent travel time alone might not bring about substantial changes to the schedules produced. However, including time-dependent dynamic demand brought substantial reductions in the total service cost, hence improving the profitability of the service. More notably, the usage of flexible bus was significantly reduced. After that, the sensitivity analysis of TI shows that increasing TI duration from 7 to 10 increases the profit, but the profit is about the same from 10 to 12, due to the fact that the bus needs to wait until TI is over, wasting some time, which incurred additional time cost. Then, a comparison between ZBFBS, traditional bus, and demand-responsive service shows that ZBFBS yields a shorter detour time than the walking time for the traditional bus, and lower cost than the demand-responsive service. This model allows schedule-based planning over the designated service time, customizing the departure times for different vehicles. In another example, we demonstrated that DARP instances can be solved by this schedule-based model as well. Last but not least, we demonstrated the applicability of the model for real-world problems by formulating and solving instances created for Chengdu, China, with real data, under different time interval durations and zone sizes. It was found that longer durations and larger zones yield higher profits. In general, this paper provided a model for strategic planning of flexible bus service by determining the zonal flexible bus schedule and service price under dynamic stochastic elastic demand volume, stochastic detour time, passengers’ time constraints, and time-dependent travel time. Future research can be conducted for dynamic vehicle assignments for real-time ride requests. For example, larger instances can be formulated and solved by approximate dynamic programming. Moreover, the demarcation of zones and duration of time interval for planning the service can be optimized by considering the demand patterns and network configurations.

CRediT authorship contribution statement

Enoch Lee: Writing - original draft, Writing - review & editing, Conceptualization, Methodology, Software, Visualization. **Xuekai Cen:** Writing - review & editing, Methodology, Supervision. **Hong K. Lo:** Writing - review & editing, Conceptualization, Methodology, Resource, Supervision

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is supported by the General Research Fund #16212819 of the Research Grants Council of the Hong Kong Special Administrative Region Government, and National Natural Science Foundation of China (No. 71890970, No. 71890974). The authors thank two anonymous reviewers for their valuable comments. Data source: DiDi Chuxing, available at: https://gaia.didichuxing.com

Appendix A: Vehicle-based flexible bus scheduling problem (P1-V) and explanation of vehicle seat reservation

Naturally, following the two-stage stochastic problem, the phase-1 problem is a flexible bus scheduling problem formulated for every vehicle to minimize the cost while meeting the designated service level specified by the reliability measures. The scheduling problem considering reliability $\boldsymbol{\rho}^{\mathrm{I}}$, $\boldsymbol{\rho}^{\mathrm{II}}$ and price factor $\boldsymbol{\pi}$ is formulated as follows,

$$\min_{\mathbf{X}',\mathbf{Y},\bar{\mathbf{Y}}} C_f(\boldsymbol{\rho}^{\mathrm{I}},\boldsymbol{\rho}^{\mathrm{II}},\boldsymbol{\pi}) = \sum_{v\in\mathcal{V}}\sum_{p\in\mathcal{P}} c_{pv}^{\text{fixed}} x_{pv} \tag{47}$$

subject to (19), (21), and,

$$\sum_{z\in\mathcal{O}_c}\sum_{z'\in\mathcal{D}_c}\sum_{v\in\mathcal{V}} y_{cv}^{zz'} = \delta_c, \forall c\in\mathcal{C} \tag{48}$$

$$\zeta_v^{zz'} = \sum_{c\in\mathcal{C}|z\in\mathcal{O}_c\wedge z'\in\mathcal{D}_c} n_c \bar{y}_{cv}^{zz'}, \forall v\in\mathcal{V}, z,z'\in\mathcal{Z}^{aug} \tag{49}$$

$$x_{pv}\mathbf{B}_p\boldsymbol{\zeta}_v \le cap_v \mathbf{1}_{m_p-1}, \forall v\in\mathcal{V}, p\in\mathcal{P} \tag{50}$$

$$\sum_{c\in\mathcal{C}}\left(\sum_{z'\in\mathcal{D}_c} y_{cv}^{zz'}\tau_{z+} + \sum_{z'\in\mathcal{O}_c} y_{cv}^{z'z}\tau_{z-}\right) \le \bar{t}_{vz}, \forall v\in\mathcal{V}, z\in\mathcal{Z}^{\text{aug}} \tag{51}$$

$$\bar{t}_{v\mathrm{I}} + \bar{t}_{vz_{p\mathrm{I}}} = x_{pv}\left(\bar{T} - \tau_{\mathrm{I}p}\right), \forall \mathrm{I}\in\mathcal{Z}_p, p\in\mathcal{P}, v\in\mathcal{V} \tag{52}$$

$$\sum_{i\in\mathcal{O}_{c,z}}\sum_{j\in\mathcal{D}_{c,z'}} y_{cv}^{ij} \le \sum_{i\in\mathcal{O}_{c,z}}\sum_{j\in\mathcal{D}_{c,z'}} \bar{y}_{cv}^{ij}, \forall z\in\mathcal{O}_c, z'\in\mathcal{D}_c, c\in\mathcal{C}, v\in\mathcal{V} \tag{53}$$

$$\sum_{z\in\mathcal{O}_c}\sum_{z'\in\mathcal{D}_c} y_{cv}^{zz'} \in \mathbb{N}_0, \forall c\in\mathcal{C}, v\in\mathcal{V} \tag{54}$$

$$\sum_{p\in\mathcal{P}} x_{pv} \le 1, \forall v\in\mathcal{V} \tag{55}$$

$$x_{pv} \in \{0,1\}, \forall p\in\mathcal{P}, v\in\mathcal{V} \tag{56}$$

$$y_{cv}^{zz'} \ge 0, \forall z\in\mathcal{O}_c, z'\in\mathcal{D}_c, c\in\mathcal{C}, v\in\mathcal{V} \tag{57}$$

$$\bar{y}_{cv}^{zz'} \in \mathbb{N}_0, \forall z\in\mathcal{O}_c, z'\in\mathcal{D}_c, c\in\mathcal{C}, v\in\mathcal{V} \tag{58}$$

$$\mathcal{O}_{c,z} = \{i\in\mathcal{O}_c \mid TW(z) - TW(i) \ge 0\}, \forall z\in\mathcal{O}_c, c\in\mathcal{C}, v\in\mathcal{V} \tag{59}$$

$$\mathcal{D}_{c,z} = \{i\in\mathcal{D}_c \mid TW(i) - TW(z) \ge 0\}, \forall z\in\mathcal{D}_c, c\in\mathcal{C}, v\in\mathcal{V} \tag{60}$$

The objective (47) is to minimize the flexible bus service cost. Equation (48) ensures the service is provided for the number of passengers δ_c required as of (19). $\mathbf{Y} = \{y_{cv}^{zz'}\}$ denotes the portion of ride requests of category c picked up in zone z and dropped off in zone z' by vehicles v. It is equivalent to $\mathbf{Y}_+$ and $\mathbf{Y}_-$ in P1-V, but this formulation combines the pick-up and drop-off variables to reduce the number of constraints.

Equation (49) calculates $\zeta_v^{zz'}$, the passenger flow per OD pair per vehicle. Similar to $\bar{\mathbf{W}}$ in the two-stage stochastic problem, $n_c \bar{y}_{cv}^{zz'}$ represents the number of spaces reserved for ride requests of category c in vehicle v from zone z to zone z', as n_c is the number of passengers of ride requests of category c. Constraint (50) is the capacity constraint defined for each vehicle and each route. This equation is formulated in vector form defined for every zone-time node along the route. Constraint (51) is the detour time constraint for each vehicle and augmented zone. Equation (52) distributes the TI duration to allowable detour time $\bar{t}_{v\text{I}}, \bar{t}_{vz_{p\text{I}}}$ and travel time $\tau_{\text{I}p}$. $z_{p\text{I}}$ is the next zone of I in route p, and is a dummy variable for the last zone in route p to keep the equation structure. $\tau_{\text{I}p}$ is the travel time from I to the next zone on route p. Constraint (53) has the same effect as constraint (2), ensuring enough vehicle spaces are reserved for each demand category served in every vehicle. Similar to (13), (54) ensures the number of ride requests served per vehicle must either be a natural number or zero. (55) constraints each vehicle to be assigned to at most one route. (56), (57) and (58) define x_{pv}, $y_{cv}^{zz'}$ and $\bar{y}_{cv}^{zz'}$ as binary, non-negative variables, and natural number, respectively. Note that after pre-processing, $y_{cv}^{zz'}$ and $\bar{y}_{cv}^{zz'}$ are defined only on feasible augmented zone pairs, as the origin zone set $\mathcal{O}_c$ and the destination zone set $\mathcal{D}_c$ contain only augmented origin and destination zones with feasible TI. Therefore, the time window constraints and location constraints can be omitted to reduce the complexity of the problem. Finally, (59) and (60) defines the sets of augmented zones to be aggregated for the constraint (53), where $TW(z)$ is the TI of augmented zone z.

Analogous to the notion of space reserved in Section 2.3, vehicles are required to reserve enough space for ride requests served, as enforced by (53), (59) and (60). Precisely, for each origin and destination augmented zone pair, the sum of the space reserved must be more than the portion of ride request served between every earlier origin augmented zone and every later destination augmented zone. Slight different from the definition in Section 2.3, the portion of ride requests served is now defined per demand category, therefore, ride requests may now be served through multiple augmented OD pairs, regardless of the consecutiveness of TIs. For example, for a demand category c, if there are three origin augmented zones R_1, R_2, R_3 in $\mathcal{O}_c$ and two destination augmented zones S_1, S_2 in $\mathcal{D}_c$, in ascending order of TI, for each $v \in \mathcal{V}$, the following constraints hold by expanding (53) under conditions (59) and (60),

$$y_{cv}^{R_1S_2} \le \bar{y}_{cv}^{R_1S_2} \tag{61}$$

$$y_{cv}^{R_1S_2} + y_{cv}^{R_2S_2} \le \bar{y}_{cv}^{R_1S_2} + \bar{y}_{cv}^{R_2S_2} \tag{62}$$

$$y_{cv}^{R_1S_2} + y_{cv}^{R_2S_2} + y_{cv}^{R_3S_2} \le \bar{y}_{cv}^{R_1S_2} + \bar{y}_{cv}^{R_2S_2} + \bar{y}_{cv}^{R_3S_2} \tag{63}$$

$$y_{cv}^{R_1S_2} + y_{cv}^{R_1S_1} \le \bar{y}_{cv}^{R_1S_2} + \bar{y}_{cv}^{R_1S_1} \tag{64}$$

$$y_{cv}^{R_1S_2} + y_{cv}^{R_2S_2} + y_{cv}^{R_1S_2} + y_{cv}^{R_2S_1} \le \bar{y}_{cv}^{R_1S_2} + \bar{y}_{cv}^{R_2S_2} + \bar{y}_{cv}^{R_1S_1} + \bar{y}_{cv}^{R_2S_1} \tag{65}$$

$$y_{cv}^{R_1S_2} + y_{cv}^{R_2S_2} + y_{cv}^{R_3S_2} + y_{cv}^{R_1S_1} + y_{cv}^{R_2S_1} + y_{cv}^{R_3S_1} \le \bar{y}_{cv}^{R_1S_2} + \bar{y}_{cv}^{R_2S_2} + \bar{y}_{cv}^{R_3S_2} + \bar{y}_{cv}^{R_1S_1} + \bar{y}_{cv}^{R_2S_1} + \bar{y}_{cv}^{R_3S_1} \tag{66}$$

R_1, R_2, R_3 denote the origin zone in the first, second, and third TIs, respectively, and S_1, S_2 denote the destination zone in the first and second TIs, respectively. Constraint (61) establishes the condition of space reserved for the earliest origin R_1 and the latest destination S_2, guaranteeing more space is reserved for each ride request than its portion served. Constraint (62) ensures enough space is reserved between OD pair (R_2, S_2) given the portion of ride request served; the space reserved and the portion of ride request served in the earlier origin zone R_1 are included in the constraint, since the passengers boarded the flexible bus in origin zone R_1 is still on the bus in origin zone R_2. Similarly, space constraint (63) of OD pair (R_3, S_2) needs to consider earlier origin zones R_1 and R_2. However, the space reserved for destination zone S_1 is yet to be considered in (61), (62) and (63) since it is earlier than S_2. Space reserved for OD pair (R_1, S_1) is guaranteed in constraint (64), which also takes into account the space reserved until the later destination zone S_2. Similarly, constraint (65) extends (62) for OD pairs (R_1, S_2), and constraint (66) extends (63) for (R_1, S_3) to incorporate both destinations S_1 and S_2.

Appendix B: Vehicle-based passenger-to-vehicle assignment problem (P2-V) and Its Equivalence to Stage 2 Problem of Two-stage Stochastic Formulation

The vehicle-based phase-2 problem (P2-V) for demand realization $\mathcal{R}^\kappa$ in scenario κ is given by,

$$\min_{\mathbf{W}', \bar{\mathbf{W}}'} Q_\kappa = \sum_{r \in \mathcal{R}^\kappa} c_r^{\text{adhoc}} \left(1 - \sum_{v \in \mathcal{V}} \sum_{z \in \mathcal{O}_{rv}} \sum_{z' \in \mathcal{D}_{rv}} w_{rv}^{zz'} \right) \tag{67}$$

subject to,

$$\sum_{r \in \mathcal{R}^\kappa} \left(\sum_{z' \in \mathcal{D}_{rv}} \tau_{r+} w_{rv}^{zz'} + \sum_{z' \in \mathcal{O}_{rv}} \tau_{r-} w_{rv}^{z'z} \right) \le \bar{t}_{vz}, \forall v \in \mathcal{V}, z \in \mathcal{Z}^{\text{aug}} \tag{68}$$

$$\bar{t}_{v\mathrm{I}} + \tau_{\mathrm{I}v} + \bar{t}_{vz_{v\mathrm{I}}} = \bar{T}, \forall \mathrm{I} \in \mathcal{Z}_v, v \in \mathcal{V} \tag{69}$$

$$\zeta_v^{zz'} = \sum_{r \in \mathcal{R}^\kappa | z \in \mathcal{O}_{rv} \wedge z' \in \mathcal{D}_{rv}} n_r \bar{w}_{rv}^{zz'}, \forall v \in \mathcal{V}, z, z' \in \mathcal{Z}^{\text{aug}} \tag{70}$$

$$x_{pv} \mathbf{B}_p \boldsymbol{\zeta}'_v \le \mathbf{1}_{m_p - 1} cap_v, \forall p \in \mathcal{P}, v \in \mathcal{V} \tag{71}$$

$$w_{rv}^{zz'} \le M \left(\sum_{i \in \mathcal{O}_{rv,z}} \sum_{j \in \mathcal{D}_{rv,z'}} \bar{w}_{rv}^{ij} \right), \forall z \in \mathcal{O}_{rv}, z' \in \mathcal{D}_{rv}, r \in \mathcal{R}^\kappa, v \in \mathcal{V} \tag{72}$$

$$\sum_{z \in \mathcal{O}_{rv}} \sum_{z' \in \mathcal{D}_{rv}} w_{rv}^{zz'} \in \{0,1\}, \forall r \in \mathcal{R}^\kappa, v \in \mathcal{V} \tag{73}$$

$$\sum_{v \in \mathcal{V}} \sum_{z \in \mathcal{O}_{rv}} \sum_{z' \in \mathcal{D}_{rv}} w_{rv}^{zz'} \le 1, \forall r \in \mathcal{R}^\kappa \tag{74}$$

$$\sum_{v \in \mathcal{V}} \sum_{z \in \mathcal{O}_{rv}} \sum_{z' \in \mathcal{D}_{rv}} \bar{w}_{rv}^{zz'} \le 1, \forall r \in \mathcal{R}^\kappa \tag{75}$$

$$\bar{w}_{rv}^{zz'} \in \{0,1\}, \forall z \in \mathcal{O}_{rv}, z' \in \mathcal{D}_{rv}, r \in \mathcal{R}^\kappa, v \in \mathcal{V} \tag{76}$$

$$w_{rv}^{zz'} \ge 0, \forall z \in \mathcal{O}_{rv}, z' \in \mathcal{D}_{rv}, r \in \mathcal{R}^\kappa, v \in \mathcal{V} \tag{77}$$

$$\mathcal{O}_{rv,z} = \{ i \in \mathcal{O}_{rv} \mid 0 \le TW(z) - TW(i) \le 1), \forall z \in \mathcal{O}_{rv}, r \in \mathcal{R}^\kappa, v \in \mathcal{V} \tag{78}$$

$$\mathcal{D}_{rv,z} = \{ i \in \mathcal{D}_{rv} \mid 0 \le TW(i) - TW(z) \le 1 \}, \forall z \in \mathcal{D}_{rv}, r \in \mathcal{R}^\kappa, v \in \mathcal{V} \tag{79}$$

This formulation optimizes $\mathbf{W}' = \{ w_{rv}^{zz'} \}$, which is the portion of ride request r served per vehicle v from augmented zone z to z', between 0 and 1 as confined by (73). Note that any ride request r can only appear in one scenario κ, so that the notation κ is omitted. Also, (73) ensures that if one ride request is served by vehicle v, the entire ride request must be served. The objective function (67) minimizes the total ad hoc service cost. Constraint (68) is the detour time constraint. Equation (69) splits the time of each window to travel time and detour time. Equation (70) calculates the passenger flow from augmented zones z to z'. Constraint (71) enforces the capacity constraint. Constraint (72) requires the vehicle space to be reserved for ride requests being served. (73) ensures the ride request must be served, if being served, entirely by the same vehicle, although the detour time required can be split in two consecutive TIs. Constraint (74) ensures that each ride request can be served at most once. Constraint (75) specifies the space reserved for at most once for each ride request. (76) requires that vehicle space be reserved completely. Similar to the phase-1 problem, as stated in constraint (77), $w_{rv}^{zz'}$ is defined only for the feasible augmented origin and augmented destination, substituting the time window constraints. (78) defines $\mathcal{O}_{rv,z}$, includes augmented zone z and the origin zone one TI earlier than z, if it is feasible for vehicle v knowing its schedule. Similarly, (79) defines $\mathcal{D}_{rv,z}$, includes augmented zone z, and if feasible, the destination zone that is one TI later than z.

We claim that this formulation is equivalent to stage 2 of the two-stage stochastic problem given a fixed x_{pv}. First, we note that in the objective value, $w_{rv}^{zz'}$ and is one-to-one dependent, since each augmented origin zones and destination zones have a specific TI defined. Therefore, as $g_r(\boldsymbol{\pi})$ is the same, the objective functions (7) and (67) have the same optimal solution. As for the constraints, equation (1) is equivalent to (69); the space constraint (2) is represented jointly by (72), (78) and (79), as (78) handles the origin detour time allocated to the space reserved for the TI and the one after, and (79) ensures the destination detour time is allocated to the space $w_{rv}^{st,\kappa}$ released in the TI and the one before. The feasible TI set (3) and time window constraints of ride requests (11)-(12) are handled through preprocessing. With the location indicators $\alpha_{r+(-)}^{z}$, feasible augmented origin and destination zones $\mathcal{O}_{rv}, \mathcal{D}_{rv}$ are created on the augmented network for each $r \in \mathcal{R}^{\kappa}, v \in \mathcal{V}$. Therefore, constraints (8), (9) and (10) are identical to (68), (70) and (71), respectively. Equations (13), (14), (15), (16) and (17) are represented by (74), (75), (77), (73) and (76), respectively, and vice versa. Finally, $\bar{w}_{rv}^{st,\kappa}$ forced to 0 in (18) is again handled by the preprocessing, and is undefined, or can be considered as 0 in the P2-V. Therefore, every constraint in P2-V has a replacement in stage 2, and vice versa, so they are equivalent.

Appendix C: Proof of Proposition 1 and Proposition 2

Proposition 1. Any feasible solution of P1-V has an equivalent solution in P1-VT

Proof. We prove by demonstrating that any feasible solution $(\mathbf{X}^*, \mathbf{Y}^*)$ of P1-V satisfying constraints (48)-(57) has an equivalent solution $(\mathbf{X}'^*, \mathbf{Y}_+^*, \mathbf{Y}_-^*)$ that satisfies constraints (22)-(33).

Denote $\mathcal{V}_u$ to be set of vehicles of type $u \in \mathcal{U}$, we choose $y_{cup+(-)}^{z}$ and x_{pu} based on below equations.

$$y_{cup+}^{z} = \sum_{z' \in \mathcal{D}_{cp}} \sum_{v \in \mathcal{V}_u} x_{pv} y_{cv}^{zz'}, \forall z \in \mathcal{O}_{cp}, p \in \mathcal{P}, c \in \mathcal{C}, u \in \mathcal{U} \tag{80}$$

$$y_{cup-}^{z} = \sum_{z' \in \mathcal{O}_{cp}} \sum_{v \in \mathcal{V}_u} x_{pv} y_{cv}^{z'z}, \forall z \in \mathcal{D}_{cp}, p \in \mathcal{P}, c \in \mathcal{C}, u \in \mathcal{U} \tag{81}$$

$$\bar{y}_{cup+}^{z} = \sum_{z' \in \mathcal{D}_{cp}} \sum_{v \in \mathcal{V}_u} x_{pv} \bar{y}_{cv}^{zz'}, \forall z \in \mathcal{O}_{cp}, p \in \mathcal{P}, c \in \mathcal{C}, u \in \mathcal{U} \tag{82}$$

$$\bar{y}_{cup-}^{z} = \sum_{z' \in \mathcal{O}_{cp}} \sum_{v \in \mathcal{V}_u} x_{pv} \bar{y}_{cv}^{z'z}, \forall z \in \mathcal{D}_{cp}, p \in \mathcal{P}, c \in \mathcal{C}, u \in \mathcal{U} \tag{83}$$

$$x_{pu} = \sum_{v \in \mathcal{V}_u} x_{pv}, \forall p \in \mathcal{P}, u \in \mathcal{U} \tag{84}$$

First, we state the following properties of P1-V.

(A) $\sum_{p\in\mathcal{P}} x_{pv} y_{cv}^{zz'} = y_{cv}^{zz'}$. It is because $\sum_{p\in\mathcal{P}} x_{pv} = 1$ if some $y_{cv}^{zz'} > 0$, induced from constraints (51) and (52).

(B) $\bigcup_{u\in\mathcal{U}} \mathcal{V}_u = \mathcal{V}$, since it is the sum of all vehicles across vehicle types.

(C) $\sum_{z\in\mathcal{O}_{cp}} \sum_{z'\in\mathcal{D}_{cp}} x_{pv} y_{cv}^{zz'} = \sum_{z\in\mathcal{O}_c} \sum_{z'\in\mathcal{D}_c} x_{pv} y_{cv}^{zz'}$. If $x_{pv} = 0$, the equality is trivial. If $x_{pv} = 1$, $y_{cv}^{zz'}$ is always zero in the augmented zone that cannot be visited by route p, induced from (51) and (52). Similarly, $\sum_{z\in\mathcal{O}_{cp}} \sum_{z'\in\mathcal{D}_{cp}} x_{pv} \bar{y}_{cv}^{zz'} = \sum_{z\in\mathcal{O}_c} \sum_{z'\in\mathcal{D}_c} x_{pv} \bar{y}_{cv}^{zz'}$.

Taking summation across every path and every origin and every vehicle type in both sides of equation (80), we have $\sum_{u\in\mathcal{U}} \sum_{p\in\mathcal{P}} \sum_{z\in\mathcal{O}_{cp}} y_{cup+}^{z} = \sum_{u\in\mathcal{U}} \sum_{p\in\mathcal{P}} \sum_{z\in\mathcal{O}_{cp}} \sum_{z'\in\mathcal{D}_{cp}} \sum_{v\in\mathcal{V}_u} x_{pv} y_{cv}^{zz'}$. Given properties (A), (B), (C), we have the RHS equals to $\sum_{z\in\mathcal{O}_c} \sum_{z'\in\mathcal{D}_c} \sum_{v\in\mathcal{V}} y_{cv}^{zz'}$, which is the LHS of equation (48). Therefore, the equation (23) is satisfied.

Equation (24) is an aggregated version for each vehicle type and route, with respect to (49) and (50) for the capacity. The LHS of (24) is equivalent to the aggregation of the LHS of (50) across every vehicle with the same type u and route p. As constraint (50) is satisfied, aggregating across all vehicles of type u assigned to route p, LHS of (24) must be smaller than $x_{pu} cap_u \mathbf{1}_{m_p - 1}$, which is the RHS. Hence, equation (24) is satisfied.

For constraint (51), we have $\sum_{v\in\mathcal{V}_u} x_{pv} \sum_{c\in\mathcal{C}} \left(\sum_{z'\in\mathcal{D}_c} y_{cv}^{zz'} \tau_{z+} + \sum_{z'\in\mathcal{O}_c} y_{cv}^{z'z} \tau_{z-} \right) \le \sum_{v\in\mathcal{V}_u} x_{pv} \bar{t}_{vz}$, by rearranging the term, we have LHS equal to $\sum_{c\in\mathcal{C}} \left(\sum_{z'\in\mathcal{D}_c} \tau_{z+} \sum_{v\in\mathcal{V}_u} x_{pv} y_{cv}^{zz'} + \sum_{z'\in\mathcal{O}_c} \tau_{z-} \sum_{v\in\mathcal{V}_u} x_{pv} y_{cv}^{z'z} \right)$, which is equivalent to LHS of equation (25). The usage of $\mathcal{C}_p$ further simplifies the constraint. Let RHS equal to $\bar{t}_{uzp}$, we get equation (25). Therefore, the constraint (26) is automatically satisfied as it is the aggregated version of (52) but shares the $\left(\bar{T} - \tau_{1p}\right)$ across vehicles with the same type and route.

For the relationship between space reserved and ride request served, by aggregating (53) across vehicles of $\mathcal{V}_u$ and path p, we have $\sum_{v\in\mathcal{V}_u} x_{pv} \sum_{i\in\mathcal{O}_{cp,z}} \sum_{j\in\mathcal{D}_{c,z'}\cap\mathcal{D}_{cp}} y_{cv}^{ij} \le \sum_{v\in\mathcal{V}_u} x_{pv} \sum_{i\in\mathcal{O}_{cp,z}} \sum_{j\in\mathcal{D}_{c,z'}\cap\mathcal{D}_{cp}} \bar{y}_{cv}^{ij}$. As the former equation is satisfied for all z', selecting z' as the destination zone with the latest time window, we have $\sum_{i\in\mathcal{O}_{cp,z}} \sum_{v\in\mathcal{V}_u} \sum_{j\in\mathcal{D}_{cp}} x_{pv} y_{cv}^{ij} \le \sum_{i\in\mathcal{O}_{cp,z}} \sum_{v\in\mathcal{V}_u} \sum_{j\in\mathcal{D}_{cp}} x_{pv} \bar{y}_{cv}^{ij}$. From definition (82), substituting $\bar{y}_{cup+}^{z} = \sum_{z'\in\mathcal{D}_{cp}} \sum_{v\in\mathcal{V}_u} x_{pv} \bar{y}_{cv}^{zz'}$ into the former equation, the constraint (28) is satisfied. As $\mathcal{D}_{cp}$ is flipped as in (36), by similar substitutions of (83) and flipping the equation, constraint (29) is also satisfied.

From the definition of (54) and (57), we can also see that the equation (27) and (30) are satisfied by taking summation across origins for (80) and summation across destinations for (81). Conditions (31), (32), and (33) are satisfied by the definitions of (84), (80), and (81), respectively. Finally, condition (34) is satisfied by combining condition (58) with (82) and (83). As every constraint and every condition is now satisfied, this completes the proof. □

Proposition 2. The solution in P2-R is a lower bound of the solution in P2-V.

Proof. We prove the proposition by showing that every feasible solution of P2-V is feasible in P2-R. First, we formally establish the relationship between $w_{rp}^{zz'}$ and $w_{rv}^{zz'}$ by,

$$w_{rp}^{zz'} = \sum_{v\in\mathcal{V}} x_{pv} w_{rv}^{zz'}, \forall p \in \mathcal{P}, z, z' \in \mathcal{Z}^{\text{aug}} \tag{85}$$

Therefore, the objective function (67) is equivalent to the objective function (37).

Aggregating equation (68) for all vehicles, we have

$$\sum_{v\in\mathcal{V}} x_{pv} \sum_{r\in\mathcal{R}^{\kappa}} \left(\sum_{z'\in\mathcal{D}_{rv}} \tau_{r+} w_{rv}^{zz'} + \sum_{z'\in\mathcal{O}_{rv}} \tau_{r-} w_{rv}^{z'z} \right) \le \sum_{v\in\mathcal{V}} x_{pv} \overline{t}_{vz} \ .$$

Rearranging the term, the LHS equals to

$$\sum_{r\in\mathcal{R}^{\kappa}} \left(\sum_{z'\in\mathcal{D}_{rv}} \tau_{r+} \sum_{v\in\mathcal{V}} x_{pv} w_{rv}^{zz'} + \sum_{z'\in\mathcal{O}_{rv}} \tau_{r-} \sum_{v\in\mathcal{V}} x_{pv} w_{rv}^{z'z} \right),$$

which is equivalent to the LHS of (38). Let $\sum_{v\in\mathcal{V}} x_{pv} \overline{t}_{vz}$ equal to $\overline{t}_{pz}$, we got (38). Then, by summing across vehicles and multiplying x_{pv} in both sides of (69), (39) is satisfied.

By (72), (78) and (79), vehicle space is reserved between the earliest TI of z where $w_{rv}^{zz'} > 0$ and the latest TI of z' where $w_{rv}^{zz'} > 0$. Therefore, if we define a 'minimum' passenger flow by,

$$\hat{\zeta}_v^{zz'} = \sum_{r\in\mathcal{R}^{\kappa} | z\in\mathcal{O}_{rv} \wedge z'\in\mathcal{D}_{rv}} n_r w_{rv}^{zz'}, \forall v \in \mathcal{V} \tag{86}$$

Denote $\hat{\boldsymbol{\zeta}}'_v$ as a vector of $\hat{\zeta}_v^{zz'}$, we always have,

$$x_{pv} \mathbf{B}_p \hat{\boldsymbol{\zeta}}'_v \le x_{pv} \mathbf{B}_p \boldsymbol{\zeta}'_v \tag{87}$$

Let $\zeta_p^{zz'}$ of (40) equals to $\sum_{v\in\mathcal{V}} x_{pv} \hat{\zeta}_v^{zz'}$, by taking summation on both sides of equation (71) across vehicles and considering the inequality (87), the equation (41) is satisfied for every route $p \in \mathcal{P}$. Then, as equation (73) and (74) are tighter than constraint (42) trivially, (42) is satisfied. Last but not least, from (85), (77), and non-negativity condition of x_{pv}, we have $\sum_{v\in\mathcal{V}} x_{pv} w_{rv}^{zz'} \ge 0$, so (43) is satisfied.

Therefore, every equation and constraint in P2-R are satisfied, so every feasible solution of P2-V is feasible in P2-R. The proof is completed. □

Appendix D: Explanation of the parameters in Chengdu scenario

In Han et al. (2019), they found out that the fuel cost per kilometer of flexible bus is RMB¥0.88, and the fuel cost per kilometer of online car-hailing service is RMB¥0.56. The zones are divided uniformly in the 9km x 8.3km area confined by Second Ring Road, therefore, the flexible bus cost and ad hoc service cost are calculated accordingly.

The Didi Chuxing driver time cost of Chengdu used in Sun et al. (2020) is RMB¥17.24 per hour, therefore, it is RMB¥0.29 per minute. Moreover, the average driving speed is about 25.7km/h, so it takes about 6.2 minutes between zones with a common border, and 8.7 minutes between zones with only a common corner. The driver cost per km is RMB¥0.68. Therefore, the total ad hoc cost is RMB¥1.24/km.

Appendix E: Result vehicle assignment in Chengdu scenario

Path	Veh	Path	Veh	Path	Veh	Path	Veh	Path	Veh	Path	Veh
ABC6	1	CEH7	1	EH5	1	GH7	1	IF3*	1	IE8*	1
ABC7	1	CF3	1	EH6	1	GH8	2	IF4*	1	AAE6	1
ABF3	1	CF4	1	EH7	1	GHI7	1	IF5*	1	AB6	1
ABF4	2	CF5	1	EH8	1	GHI8	1	IF6*	2	AB8	1
ABF5	1	CF6	1	EI5	1	HD3	1	IF7*	1	AE8	2
ABF6	1	CF7	1	EI6	1	HD4	1	IF8*	1	BBD4	1
ABF7	1	CF8	1	EI7	1	HD5	1	IFC4*	1	BBD7	1
ABF8	1	CFI6	1	EI8	1	HD7	1	IFC5*	1	BDG7	1
AE4	1	CFI7	1	FB2	1	HD8	1	IFC6*	1	BDG8	1
AE5	1	DB6	2	FB3	1	HE3	1	IFC7*	1	BEF3	1
BA7	1	DB7	2	FB4	1	HE4	2	IFC8*	1	BF7	2
BBE6	2	DB8	1	FB5	1	HE5	2	IH3*	1	BF8	2
BBE7	2	DBC6	2	FBA4	1	HE6	5	IH4*	1	CE4	2
BBE8	2	DBC7	1	FC3	1	HE8	5	IH5*	1	CEG4	1
BBF6	1	DBC8	1	FC5	1	HEB4	1	IH6*	1	CEH8	1
BC6	2	DDE6	1	FC6	1	HEB5	1	IH7*	1	DB5	1
BC7	1	DE4	1	FC7	1	HEB6	1	IH8*	1	DBC3	1
BC8	2	DE5	2	FC8	1	HEB7	1	IHG6*	1	DDH7	1
BD3	1	DEF4	2	FE4	1	HEB8	1	IHG8*	1	DEF2	1
BD8	2	DEF8	2	FE5	1	HEC4	1	AB5	1	EAB3	1
BEH5	1	DEI5	1	FE6	3	HEC5	1	AB7	1	FB1	1
BEH6	1	DEI8	1	FE7	4	HEC6	1	ADH6	1	FB6	2
BEH7	1	DH5	1	FE8	4	HEC7	1	ADH8	1	FB7	2
BEH8	1	EA5	1	FED2	1	HEC8	1	BDG3	1	FB8	2
BF4	1	EA6	1	FED3	1	HF3	1	BEI6	1	FH6	2
BF5	1	EA7	1	FED4	1	HF4	1	BEI7	1	FH7	2
CB2	1	EB4	1	FED5	1	HF5	1	FED1	1	FH8	2
CB3	1	EB5	1	FED6	1	HG3*	1	CEG8	1	HD6	1
CB4	1	EB6	1	FED7	1	HG4*	1	DBC5	1	HEB2	1
CB5	1	EB7	1	FED8	1	HG6*	1	DEF3	1	HF7	2
CB6	1	EB8	1	FEG6	1	HG7*	1	DDH6	1	HF8	2
CB7	1	EC5	1	FEG7	1	HG8*	1	DDH8	1	HI3*	1
CB8	1	EC6	1	FEG8	1	HHF6	1	DH4	1	IE5*	1
CBA6	1	EC7	1	FH4	1	HI4*	1	EA8	1	IE7*	1
CBD3	1	EC8	1	FH5	1	HI5*	1	FIE3	1	IEB2*	1
CBD4	1	ED3	1	FHE3	1	HI6*	1	GGE6	1	IEF3*	1
CBD5	1	ED4	1	FI5	1	HI7*	1	GEF4	1	BFE3	1
CBD6	1	ED5	1	FI6	1	HI8*	1	GHI4	1	DDEF5	1
CBD7	1	ED6	1	FI7	1	IE4*	1	GHI6	1	DDEF6	1
CBD8	1	ED7	1	FI8	1	IE6*	2	HDA6	1	DDEF7	1
CCE5	1	ED8	1	GE4	1	IEB3*	1	HEB3	1	BBE4	1
CCE6*	2	EF3	1	GE5	1	IEB4*	1	HEC2	1	BBE5	1
CCE7*	1	EF4	1	GEF5	2	IEB5*	1	HG5*	1	BCE3	1
CCE8*	1	EF5	1	GEF6	1	IEB6*	1	IEB8*	1	CBE1	1
CE3	2	EF6	1	GEF7	2	IEB7*	1	AAE7	1	EHF3	1
CE7	2	EF7	1	GEF8	1	IED3*	1	BBD5	1	FEH3	1
CE8	1	EF8	1	GGE7	1	IED4*	1	BBD6	1		
CEH3	1	EG6	1	GGE8	1	IED5*	1	DDE7	1		
CEH4	1	EG7	1	GH4	1	IED6*	1	DDE8	1		
CEH5	1	EG8	1	GH5	1	IED7*	1	BE8	1		
CEH6	1	EH4	1	GH6	1	IED8*	1	HE7	5		

* 10-seaters are used

Appendix F: Parameters and variables for different models

For P1-V:

Notation	Description
Sets	
$\mathcal{Z}^{\text{aug}}$ / $\mathcal{Z}_p$	Set of augmented zone/ Set of augmented zones traversed by route $p \in \mathcal{P}$
$\mathcal{P}$	Set of routes

$\mathcal{V}$	Set of vehicles
$\mathcal{C}$	Set of demand categories
$\mathcal{O}_c, \mathcal{D}_c \subseteq \mathcal{Z}^{\text{aug}}$	Set of possible origins and destinations for demand category $c \in \mathcal{C}$
$\mathbb{N}_0$	Set of non-negative integers
Parameters	
$\tau_{\text{I}p}$	Travel time from zone I to next zone in route $p \in \mathcal{P}$
$\bar{T}$	Duration of time intervals
c_{pv}^{fixed}	The fixed cost for vehicle v serving route p, $v \in \mathcal{V}, p \in \mathcal{P}$
$\tau_{r+(-)}$	The detour time for picking up (dropping off) any ride requests in zone $r \in \mathcal{Z}$
n_r	Number of passengers for demand category $r \in \mathcal{C}$
$\mathbf{B}_p$	A converting matrix from passenger flow to the number of in-vehicle passengers for vehicles of route $p \in \mathcal{P}$
cap_v	The vehicle capacity of vehicle $v \in \mathcal{V}$
M	A very large number
m_p	Sum of the number of time intervals and the number of zones visited by route $p \in \mathcal{P}$
Δ_c	Demand volume distribution of demand category $c \in \mathcal{C}$
$\Lambda_{z+(-)}$	Detour time distribution in zone z for access (egress), $z \in \mathcal{Z}$
$TW(z) \in \mathcal{T}$	Time interval of the augmented zone $z \in \mathcal{Z}^{\text{aug}}$
Decision variables	
$\bar{t}_{vz}$	Maximum allowable detour time in augmented zone $z \in \mathcal{Z}^{aug}$ for vehicle $v \in \mathcal{V}$
$\mathbf{X} = \{x_{pv}\}$	A matrix indicates if vehicle v is serving route p, $v \in \mathcal{V}, p \in \mathcal{P}$
$\boldsymbol{\rho}^{\text{I}} = \{\rho_c^{\text{I}}\}$	Vector of volume reliability measure for demand category $c \in \mathcal{C}$
$\boldsymbol{\rho}^{\text{II}} = \{\rho_{z+}^{\text{II}}, \rho_{z-}^{\text{II}}\}$	Vector of detour time reliability measure, defined for each zone $z \in \mathcal{Z}$
$\mathbf{Y} = \{y_{cv}^{ij}\}$	The portion of ride requests of category c picked up by vehicle v in augmented zone i and dropped off in augmented zone j, $c \in \mathcal{C}, v \in \mathcal{V}, i, j \in \mathcal{Z}^{\text{aug}}$
$\bar{\mathbf{Y}} = \{y_{cv}^{ij}\}$	The number ride requests of category c requires space reserved between augmented zone i and augmented zone j in vehicle v, $c \in \mathcal{C}, v \in \mathcal{V}, i, j \in \mathcal{Z}^{\text{aug}}$
$\boldsymbol{\pi}$	Price factor to be optimized.
δ_c	The number of ride requests of demand category $c \in \mathcal{C}$ required to be served in phase-1
$\boldsymbol{\zeta}'_v = \{\zeta_v^{zz'}\}$	The number of passengers travelling from zone z to z' by vehicle $v \in \mathcal{V}$, $z, z' \in \mathcal{Z}^{\text{aug}}$

For P2-V:

Notation	Description
Sets	
$\mathcal{Z}^{\text{aug}} / \mathcal{Z}_v$	Set of augmented zone/ Set of augmented zones traversed by vehicle $v \in \mathcal{V}$
$\mathcal{P}$	Set of routes
$\mathcal{V}$	Set of vehicles
$\mathcal{R}^{\kappa}$	Set of ride requests under scenario $\kappa \in \mathcal{K}$
$\mathcal{O}_{rv}, \mathcal{D}_{rv} \subseteq \mathcal{Z}^{\text{aug}}$	Set of possible origins and destinations for ride request $r \in \mathcal{R}$ by vehicle $v \in \mathcal{V}$
$\mathcal{O}_{rv,z}, \mathcal{D}_{rv,z}$	Set of augmented zone with origin zone one TI earlier than $z \in \mathcal{O}_{rv}$, and destination zone one TI later than $z \in \mathcal{D}_{rv}$
$\mathbb{N}_0$	Set of non-negative integers
Parameters	
$(z,s) \in \mathcal{N}$	Zone-time node of zone z in time interval s, $z \in \mathcal{Z}, s \in \mathcal{T}$
$\tau_{\text{I}v}$	Travel time from zone I to next zone for vehicle $v \in \mathcal{V}$, $\text{I} \in \mathcal{Z}$
$\bar{T}$	Duration of time intervals
c_r^{adhoc}	The ad hoc service cost for ride request $r \in \mathcal{R}$
$\tau_{r+(-)}$	The detour time for picking up (dropping off) ride request $r \in \mathcal{R}$

n_r	Number of passengers for ride request $r \in \mathcal{R}$
$\mathbf{B}_p$	A converting matrix from passenger flow to the number of in-vehicle passengers for vehicles of route $p \in \mathcal{P}$
cap_v	The vehicle capacity of vehicle $v \in \mathcal{V}$
m_p	Sum of the number of time intervals and the number of zones visited by route $p \in \mathcal{P}$
M	A very large number
$\mathbf{X} = \{x_{pv}\}$	A matrix indicates if vehicle v is serving route p, $v \in \mathcal{V}, p \in \mathcal{P}$
$TW(z) \in \mathcal{T}$	Time interval of the augmented zone $z \in \mathcal{Z}^{\text{aug}}$
Decision variables	
$\bar{t}_{vz}$	Maximum allowable detour time in augmented zone $z \in \mathcal{Z}^{aug}$ for vehicle $v \in \mathcal{V}$
$\mathbf{W}' = \{w_{rv}^{zz'}\}$	The portion of ride request r picked up in zone z and dropped off in zone z', for vehicle v, $v \in \mathcal{V}, z, z' \in \mathcal{Z}^{aug}, r \in \mathcal{R}$
$\bar{\mathbf{W}}' = \{\bar{w}_{rv}^{zz'}\}$	Indicate if space reserved for ride request r from zone z to zone z' for vehicle v, $v \in \mathcal{V}, z, z' \in \mathcal{Z}^{aug}, r \in \mathcal{R}$
$\boldsymbol{\zeta}'_v = \{\zeta_v^{zz'}\}$	The number of passengers travelling from zone z to z' by vehicle $v \in \mathcal{V}$, $z, z' \in \mathcal{Z}^{\text{aug}}$

For P1-VT:

Notation	Description
Sets	
$\mathcal{Z}^{\text{aug}} / \mathcal{Z}_p$	Set of augmented zone/ Set of augmented zones traversed by route $p \in \mathcal{P}$
$\mathcal{P}$	Set of routes
$\mathcal{U}$	Set of vehicle types
$\mathcal{C}$ / $\mathcal{C}_p$	Set of demand categories / Set of demand categories that can be served by route p, $p \in \mathcal{P}$
$\mathcal{O}_{cp}, \mathcal{D}_{cp} \subseteq \mathcal{Z}^{\text{aug}}$	Set of possible origins and destinations for demand category $c \in \mathcal{C}$ by route $p \in \mathcal{P}$
$\mathbb{N}_0$	Set of non-negative integers
Parameters	
$\tau_{\text{I}p}$	Travel time from zone I to next zone in route $p \in \mathcal{P}$, $\text{I} \in \mathcal{Z}$
$\bar{T}$	Duration of time intervals
c_{pu}^{fixed}	The fixed cost for vehicle type u serving route p, $u \in \mathcal{U}, p \in \mathcal{P}$
$\tau_{r+(-)}$	The detour time for picking up (dropping off) any ride requests in zone $r \in \mathcal{Z}$
n_r	Number of passengers for ride request $r \in \mathcal{R}$ or demand category $r \in \mathcal{C}$
$\mathbf{B}_{p+(-)}$	Converting matrix from passenger space reserved to number of passengers boarded (egressed) for vehicles of route $p \in \mathcal{P}$
cap_v	The vehicle capacity of vehicle type $v \in \mathcal{U}$
M	A very large number
m_p	Sum of the number of time intervals and the number of zones visited by route $p \in \mathcal{P}$
$O_r, D_r \in \mathcal{Z}$	Origin and destination of ride request $r \in \mathcal{R}$
Δ_c	Demand volume distribution of demand category $c \in \mathcal{C}$
$\Lambda_{z+(-)}$	Detour time distribution in zone z for access (egress), $z \in \mathcal{Z}$
N_u	Number of vehicles of vehicle type $u \in \mathcal{U}$
$TW(z) \in \mathcal{T}$	Time interval of the augmented zone $z \in \mathcal{Z}^{\text{aug}}$
Decision variables	
$\bar{t}_{uzp}$	Maximum allowable detour time in augmented zone $z \in \mathcal{Z}^{aug}$ for vehicles of type $u \in \mathcal{U}$ and route $p \in \mathcal{P}$
$\mathbf{X}' = \{x_{pu}\}$	Number of vehicles of type u assigned to route p, $u \in \mathcal{U}, p \in \mathcal{P}$
$\boldsymbol{\rho}^{\text{I}} = \{\rho_c^{\text{I}}\}$	Vector of volume reliability measure for demand category $c \in \mathcal{C}$
$\boldsymbol{\rho}^{\text{II}} = \{\rho_{z+}^{\text{II}}, \rho_{z-}^{\text{II}}\}$	Vector of detour time reliability measure, defined for each zone $z \in \mathcal{Z}$

$\mathbf{Y}_{+(-)} = \{y^{z}_{cup+(-)}\}$	The portion of ride requests of category c picked up (dropped off) by vehicles of type u and route p in augmented zone z, $u \in \mathcal{U}, c \in \mathcal{C}, p \in \mathcal{P}, z \in \mathcal{Z}^{\text{aug}}$
$\bar{\mathbf{Y}}_{+(-)} = \{\bar{y}^{z}_{cup+(-)}\}$	The number ride requests of category c that their space reserved (released) in augmented zone z in vehicles of type u and route p, $u \in \mathcal{U}, c \in \mathcal{C}, p \in \mathcal{P}, z \in \mathcal{Z}^{\text{aug}}$
$\boldsymbol{\pi}$	Price factor to be optimized.
δ_c	The number of ride requests of demand category $c \in \mathcal{C}$ required to be served in phase-1

For P2-R:

Notation	Description
Sets	
$\mathcal{Z}^{\text{aug}} / \mathcal{Z}_p$	Set of augmented zone/ Set of augmented zones traversed by route $p \in \mathcal{P}$
$\mathcal{P}$	Set of routes
$\mathcal{R}^{\kappa}$	Set of ride requests under scenario $\kappa \in \mathcal{K}$
$\mathcal{O}_{rp}, \mathcal{D}_{rp} \subseteq \mathcal{Z}^{\text{aug}}$	Set of possible origins and destinations for ride request $r \in \mathcal{R}$ or demand category $r \in \mathcal{C}$ by route $p \in \mathcal{P}$
$\mathbb{N}_0$	Set of non-negative integers
Parameters	
$\tau_{\text{I}p}$	Travel time from zone I to next zone in route $p \in \mathcal{P}$, $\text{I} \in \mathcal{Z}$
$\bar{T}$	Duration of time intervals
c_r^{adhoc}	The ad hoc service cost for ride request $r \in \mathcal{R}$
$\tau_{r+(-)}$	The detour time for picking up (dropping off) ride request $r \in \mathcal{R}$ or for any ride requests in zone $r \in \mathcal{Z}$
n_r	Number of passengers for ride request $r \in \mathcal{R}$ or demand category $r \in \mathcal{C}$
$\mathbf{B}_p$	A converting matrix from passenger flow to the number of in-vehicle passengers for vehicles of route $p \in \mathcal{P}$
cap_v	The vehicle capacity of vehicle $v \in \mathcal{V}$ or vehicle type $v \in \mathcal{U}$
m_p	Sum of the number of time intervals and the number of zones visited by route $p \in \mathcal{P}$
N_u	Number of vehicles of vehicle type $u \in \mathcal{U}$
$\mathbf{X} = \{x_{pv}\}$	A matrix indicates if vehicle v is serving route p, $v \in \mathcal{V}, p \in \mathcal{P}$
$TW(z) \in \mathcal{T}$	Time interval of the augmented zone $z \in \mathcal{Z}^{\text{aug}}$
Decision variables	
$\bar{t}_{pz}$	Maximum allowable detour time in augmented zone $z \in \mathcal{Z}^{aug}$ for route $p \in \mathcal{P}$
$\mathbf{W}'' = \{w^{zz'}_{rp}\}$	The portion of ride request r picked up in zone z and dropped off in zone z', for vehicles of route p, $p \in \mathcal{P}, z, z' \in \mathcal{Z}^{aug}, r \in \mathcal{R}$
$\boldsymbol{\zeta}'_p = \{\zeta^{zz'}_p\}$	The number of passengers travelling from zone z to z' by vehicles of route $p \in \mathcal{P}$, $\forall z, z' \in \mathcal{Z}^{\text{aug}}$

References

An, K., Lo, H.K., 2016. Two-phase stochastic program for transit network design under demand uncertainty. *Transportation Research Part B: Methodological* 84, 157–181.

An, K., Lo, H.K., 2014. Ferry service network design with stochastic demand under user equilibrium flows. *Transportation Research Part B: Methodological* 66, 70–89.

Bertsimas, D., Jaillet, P., Martin, S., 2019. Online vehicle routing: The edge of optimization in large-scale applications. *Operations Research* 67, 143–162.

Cascetta, E., Coppola, P., 2016. Assessment of schedule-based and frequency-based assignment models for strategic and operational planning of high-speed rail services. *Transportation Research Part A: Policy and Practice* 84, 93–108.

Cordeau, J.F., Laporte, G., 2007. The dial-a-ride problem: Models and algorithms. *Annals of Operations Research* 153, 29–46.

Errico, F., Crainic, T.G., Malucelli, F., Nonato, M., 2013. A survey on planning semi-flexible transit systems: Methodological issues and a unifying framework. *Transportation Research Part C: Emerging Technologies* 36, 324–338.

Fu, Q., Liu, R., Hess, S., 2012. A Review on Transit Assignment Modelling Approaches to Congested Networks: A New Perspective. *Procedia - Social and Behavioral Sciences* 54, 1145–1155.

Gong, M., Hu, Y., Chen, Z., Li, X., 2021. Transfer-based customized modular bus system design with passenger-route assignment optimization. *Transportation Research Part E: Logistics and Transportation Review* 153, 102422.

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

Hamdouch, Y., Szeto, W.Y., Jiang, Y., 2014. A new schedule-based transit assignment model with travel strategies and supply uncertainties. *Transportation Research Part B: Methodological* 67, 35–67.

Han, Chen, Li, Zhang, Sun, 2019. Customized Bus Network Design Based on Individual Reservation Demands. *Sustainability* 11, 5535.

Ho, S.C., Haugland, D., 2011. Local search heuristics for the probabilistic dial-a-ride problem. *OR Spectrum* 33, 961–988.

Huang, W., Li, L., Lo, H.K., 2018. Adaptive traffic signal control with equilibrium constraints under stochastic demand. *Transportation Research Part C: Emerging Technologies* 95, 394–413.

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

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. *AIP Conference Proceedings* 1631, 58–62.

Lee, E., Cen, X., Lo, H.K., 2021a. Zonal-based flexible bus service under elastic stochastic demand. *Transportation Research Part E: Logistics and Transportation Review* 152, 102367.

Lee, E., Cen, X., Lo, H.K., Ng, K.F., 2021b. Designing Zonal-Based Flexible Bus Services Under Stochastic Demand. *Transportation Science* 55, 1280–1299.

Li, L., Huang, W., Lo, H.K., 2018. Adaptive coordinated traffic control for stochastic demand. *Transportation Research Part C: Emerging Technologies* 88, 31–51.

Liang, S., Zhao, S., Lu, C., Ma, M., 2016. A self-adaptive method to equalize headways: Numerical analysis and comparison. *Transportation Research Part B: Methodological* 87, 33–43.

Liu, T., Ceder, A., 2015. Analysis of a new public-transport-service concept: Customized bus in China. *Transport Policy* 39, 63–76.

Liu, X., Qu, X., Ma, X., 2021. Improving flex-route transit services with modular autonomous vehicles. *Transportation Research Part E: Logistics and Transportation Review* 149, 102331.

Lo, H.K., An, K., Lin, W., 2013. Ferry service network design under demand uncertainty. *Transportation Research Part E: Logistics and Transportation Review* 59, 48–70.

Ma, J., Yang, Y., Guan, W., Wang, F., Liu, T., Tu, W., Song, C., 2017. Large-scale demand driven design of a customized bus network: A methodological framework and beijing case study. *Journal of Advanced Transportation* 2017.

Mahmoudi, M., Zhou, X., 2016. Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows: A dynamic programming approach based on state-space-time network representations. *Transportation Research Part B: Methodological* 89, 19–42.

Mokhtar, H., Krishnamoorthy, M., Dayama, N.R., Kumar, P.N.R., 2020. New approaches for solving the convoy movement problem. *Transportation Research Part E: Logistics and Transportation Review* 133, 101802.

Mourad, A., Puchinger, J., Chu, C., 2019. A survey of models and algorithms for optimizing shared mobility. *Transportation Research Part B: Methodological* 123, 323–346.

Ritzinger, U., Puchinger, J., Hartl, R.F., 2015. A survey on dynamic and stochastic vehicle routing problems. *International Journal of Production Research* 7543, 1–17.
Ropke, S., Cordeau, J.-F., Laporte, G., 2007. Models and branch-and-cut algorithms for pickup and delivery problems with time windows. *Networks* 49, 258–272.
Sánchez-Martínez, G.E., Koutsopoulos, H.N., Wilson, N.H.M., 2016. Real-time holding control for high-frequency transit with dynamics. *Transportation Research Part B: Methodological* 83, 1–19.
Schilde, M., Doerner, K.F., Hartl, R.F., 2014. Integrating stochastic time-dependent travel speed in solution methods for the dynamic dial-A-ride problem. *European Journal of Operational Research* 238, 18–30.
Steinzen, I., Gintner, V., Suhl, L., Kliewer, N., 2010. A time-space network approach for the integrated vehicle-and crew-scheduling problem with multiple depots. *Transportation Science* 44, 367–382.
Sun, L., Teunter, R.H., Hua, G., Wu, T., 2020. Taxi-hailing platforms: Inform or Assign drivers? *Transportation Research Part B: Methodological* 142, 197–212.
Szeto, W.Y., Jiang, Y., 2014. Transit assignment: Approach-based formulation, extragradient method, and paradox. *Transportation Research Part B: Methodological* 62, 51–76.
Tong, C.O., Wong, S.C., 1999. A stochastic transit assignment model using a dynamic schedule-based network. *Transportation Research Part B: Methodological* 33, 107–121.
Tong, L. (Carol), Zhou, L., Liu, J., Zhou, X., 2017. Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing. *Transportation Research Part C: Emerging Technologies* 85, 451–475.
Ulmer, M.W., Goodson, J.C., Mattfeld, D.C., Hennig, M., 2019. Offline–Online Approximate Dynamic Programming for Dynamic Vehicle Routing with Stochastic Requests. *Transportation Science* 53, 185–202.
Ulmer, M.W., Mattfeld, D.C., Köster, F., 2018. Budgeting time for dynamic vehicle routing with stochastic customer requests. *Transportation Science* 52, 20–37.
Zhang, J., Wang, D.Z.W., Meng, M., 2017. Analyzing Customized Bus Service on a Multimodal Travel Corridor: An Analytical Modeling Approach. *Journal of Transportation Engineering Part a-Systems* 143, 1–12.
Zhang, Y., Lam, W.H.K., Sumalee, A., Lo, H.K., Tong, C.O., 2010. The multi-class schedule-based transit assignment model under network uncertainties. *Public Transport* 2, 69–86.
Zhao, M., Li, X., Yin, J., Cui, J., Yang, L., An, S., 2018. An integrated framework for electric vehicle rebalancing and staff relocation in one-way carsharing systems: Model formulation and Lagrangian relaxation-based solution approach. *Transportation Research Part B: Methodological* 117, 542–572.