https://doi.org/10.1016/j.trc.2019.12.004

A Two-Phase Optimization Model for the Demand-Responsive Customized Bus Network Design

*Di Huang [a], Yu Gu [b], Shuaian Wang [b], Zhiyuan Liu [a,]**

[a] Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, China

[b] Department of Logistics & Maritime Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong

***Corresponding author.** Email: zhiyuanl@seu.edu.cn

Abstract

This paper proposes a new optimization model for the network design problem of the demand-responsive customized bus (CB). The proposed model consists of two phases: insert the passenger requests dynamically in an interactive manner (dynamic phase) and optimize the service network statically based on the overall demand (static phase). To model the network design problem in the dynamic phase, we propose a bi-level programming problem to describe the interactive manner between the operator and the passengers. The upper-level is formulated as a mixed-integer program with the objective of maximizing the operator's revenue, and the lower-level is the passenger's choice problem for a given trip plan provided by the operator. The CB passenger's travel behavior is assumed to follow the stochastic user equilibrium with elastic demand. A dynamic insertion method is developed to address the proposed bi-level model. For the network design problem in the static phase, the service network is re-optimized based on the confirmed passengers with the strict time deviation constraints, which is embedded in the static multi-vehicle pickup and delivery problem. An exact solution method is developed based on the branch-and-bound (B&B) algorithm. Numerical examples are conducted to verify the proposed models and solution algorithms.

Keywords: customized bus, demand-responsive transit, bi-level programming, dynamic insertion, branch-and-bound algorithm

1. Introduction

The demand-responsive transit (DRT) service is an emerging and flexible instrument to enhance the serviceability of urban public transport systems. As shown in Fig. 1(a), the term DRT covers customized (or subscription) bus (CB), shuttle bus, feeder bus, and other on-demand shared mobility services. It is defined as an intermediate form of transit service between the mass transit system, and the highly flexible and personalized services provided by taxis (Mageean & Nelson, 2003; Lyu et al., 2019) (see Fig. 1(b)). The CB system, as an emerging public transportation, aims to provide personalized, flexible and passengers-oriented services to those with similar travel demands in both space and time, and those with specific requirements (e.g., regular commuters, mobility impairment passengers, and passengers living in low-demand areas which are not accessible by conventional transit services) (Qiu et al., 2017; Zhang et al., 2017; Tong et al., 2017; Lyu et al., 2019). It has been identified as an efficient and green alternative to private vehicles and conventional transit services (Ren et al., 2016).

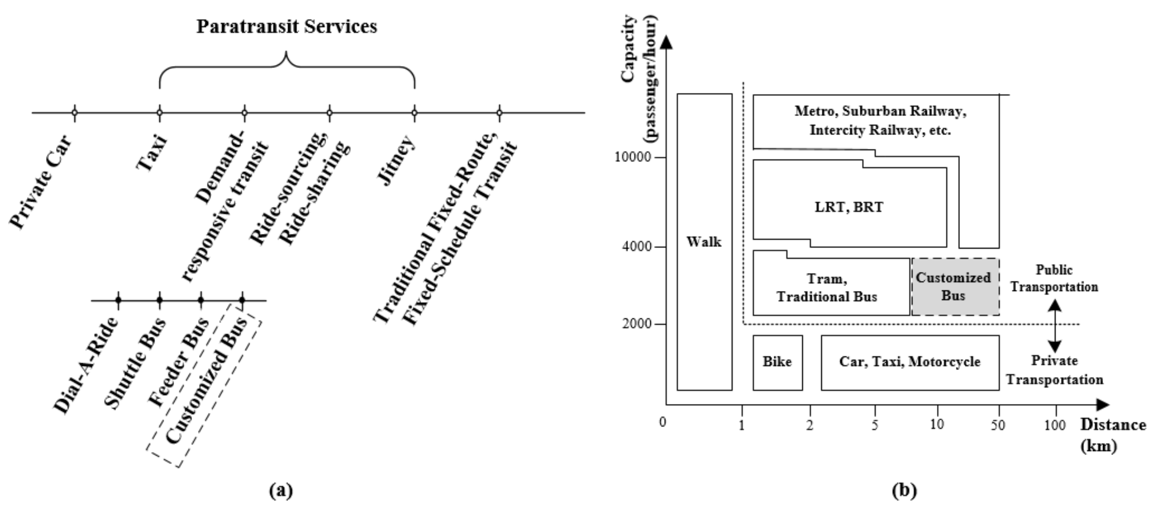


Figure 1. The service characteristic of CB (Source: KFH Group, 2008)

The CB service requires an online communication platform between each passenger and the operator, to determine the vehicle assignment, routing, and scheduling plans (Liu & Ceder, 2015). Such communication is called the ***subscription stage***, and it includes the following processes: (a) Each passenger dynamically submits her/his request online with pickup/delivery locations and times; (b) The operator inputs each request into the CB planning system to modify the existing routes and proposes a trip plan with the estimated pickup/delivery times and ticket price; (c) Once received this proposed trip plan from the operator, the passenger then confirms online to accept this plan or not, based on her/his perceived travel costs. This interactive communication (subscription stage) is dynamically conducted between the operator and each passenger, which is defined as a ***dynamic phase*** for the sake of presentation. Then, prior to the departure of each CB vehicle, the operator will optimize the CB service network/plan with overall confirmed demand, which is defined as the ***static phase*** for the analysis in this paper.

Note that in the dynamic phase, the request of each passenger is occasional and unpredictable. Thus, the process (b) described in the paragraph above remains challenging, which needs to quick respond to the passenger's request within several seconds. Since a fair amount of historical requests are known in advance from the day-to-day operation, they could be used to create initial routes at the beginning of the day. The CB service addressed in this paper is ***passenger-oriented***; namely, once a passenger request pops up, the operator needs to input the request into the existing routes and propose a specific trip plan in the process (b). Therefore, the operator needs to rapidly determine whether such a new request could be inserted to the existing CB routes while fulfilling existing passengers' time windows; or to launch a new route to take this passenger with the risk of deficit in the low demand area. The analytical problems in the dynamic/static phases are of essential importance to the modelling of the CB network design; especially the three interactive communication and decision-making processes in the dynamic phase.

Existing studies on the CB network design usually separate the analysis and objectives of the operator and passenger and do not fully cover the decision-making processes in the dynamic phase. Thus, this paper aims to fill such gap and provide a comprehensive modelling framework for the CB network design problem, where a two-phase optimization approach is proposed to analyze the network design problems in the dynamic phase and static phase, respectively.

For the dynamic phase, a bi-level programming model is proposed, where the CB operator acts as the leader, and each passenger is the follower. In the upper-level, the operator optimizes the CB service routes with the objective of maximizing its profit by inserting the passenger request into the existing routes. In the lower-level, each passenger decides whether to accept the CB service according to perceived travel costs, including in-vehicle travel time, schedule deviations, and transit fare. Herein, a distance-based fare is adopted, which could encourage passengers to make wise routing decisions (Huang et al., 2016). The objective of each passenger is to maximize her/his trip utility, which is *per se* a discrete choice model. The stochastic user equilibrium (SUE) principle is adopted for the passengers' travel behavior. Eventually, with the confirmed demand from the dynamic phase, the static phase is formulated as a static multi-vehicle routing problem with fixed pickup and delivery time window, where the operating efficiency of the CB system could be further improved.

The remainder of this paper is organized as follows. Section 2 reviews the existing studies on modelling of CB network design. The problem statement and network formulation are presented in Section 3. A bi-level programming model for the dynamic phase is proposed in Section 4. Section 5 proposes a static multi-vehicle routing problem with pickup and delivery based on the results obtained from the dynamic phase, which is followed by the illustration of the solution algorithm. Section 6 presents some numerical results to verify the proposed model and algorithm. Finally, Section 7 concludes this paper and outlooks future research.

2. Literature Review

2.1 CB network design

Despite the fact that the CB is a new and innovative transit mode, the practice of the similar on-demand transit service emerged in 1970s, e.g. subscription bus (Chang & Schonfeld, 1991), ADA (Aldaihani et al., 2004), Dail-a-Ride (Ho et al., 2019), PRT (Chebbi & Chaouachi, 2016), etc. The idea of CB is originated from the concept of car-sharing, aiming to serve groups of passengers with similar travel requests (Lyu et al., 2019). Another feature of the CB service is the subscription/pre-pay mechanism for passengers to book seats. It is tailored to meet passengers' preferences of a higher level of service quality compared with conventional bus services (Liu & Ceder, 2015). Potts et al. (2010) conduct a comprehensive review of the existing types of on-demand transit systems and develop a practical guide to service providers. A decision-making framework is proposed which requires more communication and scheduling technologies than the conventional fixed-route bus system. Liu & Ceder (2015) then divide the service design process of the CB system into four steps: travel survey, call for passengers, seats reservation and seats purchase. Chang & Schonfeld (1991) point out that the subscription service is preferable in the on-demand services in that it could efficiently reduce the rejection rate and guarantee a profitable service system.

Generally, the demand pattern during the subscription process can be categorized into two types, which are static and dynamic (see Section 2.2). In the static case, all demands are assumed as known and fixed in advance, which can be obtained from reservations in previous days or subscriptions of regular passengers. In practice, the routing and scheduling design problems need to be solved prior to operations at the beginning of a day, the results of which are not allowed to change afterward. Tong et al. (2017) develop a joint optimization model that addresses two challenging problems in CB practice, which are increasing the passenger ridership and optimizing bus routing and timetabling plans. In this system, passengers can book the recommended lines directly. If there are no feasible lines, passengers' demands are stored in the request pool for future line designs. Guo et al. (2019) develop a mixed integer programming model that determines the bus stop location and route design simultaneously. Lyu et al. (2019) propose a new CB line planning framework called "CB-Planner". By using multi-source data, this framework is capable of determining stop location, bus routes, timetables, and passenger ridership.

2.2 Demand-responsive (passenger-oriented) transit network design

With the help of the advanced real-time data collection and computing technologies, the study of the ad hoc on-demand transit system has become one of the most attractive and challenging topics, which needs to deal with the routing and scheduling modification in service with real-time service requests. Considering the inherent complexity caused by the dynamic decision-making process, the ad hoc on-demand transit system design problem can be addressed by simulation approaches and analytical models. Interested readers could refer to Ronald et al. (2015) for a detailed review of simulation-based approaches to the on-demand transit system. In the dynamic case, the passenger's request is not known in advance (both pickup/delivery times and locations). Two basic solution strategies are widely applied to deal with the dynamics and randomness of passenger demands (Berbeglia et al., 2010): i) solving a static problem each time based on the real-time information (including both new request and cancellation); and ii) solving the static problem only once initially and updating the current solution with heuristic methods, such as insertion heuristics. Horn (2002) develops an incremental insertion method that consists of a set of periodical steepest-descent improvement procedures to minimize additional travel time.

Coslovich et al. (2006) propose a two-phase insertion algorithm based on the concept of route perturbations. In the off-line phase, a feasible neighborhood of the current route is generated, while in the on-line phase, the new request is inserted to minimize the schedule deviations of previous passengers. Pavone et al. (2011) assume that the expected passenger arrival rate follows a certain probability distribution based on historical patterns. van Engelen et al. (2018) then develop an online dynamic insertion method with forecasted demand.

Compared with conventional transit and on-demand transit system design problem, the passengers are involved in the planning process of the service network through the subscription mechanism (Liu et al., 2016). Suhl et al. (2001) first introduce the passenger-oriented dispatching strategy in the railway system. However, due to the limitation of data collection, only simulation data can be used to verify the proposed methodology. With the advent of telecommunication technologies, the information platform through the Internet and smartphones has been constructed that expedites the interactivity between passengers and operators (Kamga, 2013; Liu & Ceder, 2015; Chen et al., 2017; Chen & Nie, 2017). Foth et al. (2013) identify the opportunities of the application of connecting data (e.g., social media, mobile, geospatial information, etc.) in the planning of public transport system. Stelzer et al. (2016) also indicate that an information exchange platform could help to improve the service quality of the transit system. The passenger feedback would play a crucial role in transit operations and managements.

2.3 Vehicle routing problem with pickup and delivery

From the view of operations research, the CB network design problem can be formulated as a vehicle routing problem with pickup and delivery (VRPPD). In literature, minimizing operation cost (Cordeau, 2006), maximizing satisfied demand (Tong et al., 2017), and maximizing the quality of service (Calvo & Colorni, 2007) are three crucial objectives to be optimized separately or simultaneously (Diana & Dessouky, 2004). Generally, the service quality can be measured by the route duration, passenger riding and waiting times, schedule deviation, capacity, etc., some of which can also be formulated as constraints associated with routing and scheduling problems such as coupling, precedence, and time window constraints (Cordeau & Laporte, 2007). Given the objectives and constraints, the VRPPD can be formulated as the mixed-integer programming model with

Table 1-1 Comparison of existing works (CB)

System type	Publication	Subscr-iption	Demand pattern	Objective	Decision variable	Constraints	Solution algorithm
	Chebbi & Chaouachi (2016)	No	Static	Min. of empty movement and fleet size	Route	PA	Heuristic algorithm
	Cao & Wang (2017)	No	Static	Min. of system cost	# of passengers choosing CB	FL; FC; RL	Exact algorithm
	Ma et al. (2017)	No	Static	Min. of system cost	# of passengers choosing CB	RL; FC	Exact algorithm
CB	Tong et al. (2017)	Yes	Static	Max. of served pax.	Stop location; Route; Schedule	LF; PD; VC; TW	Lagrangian decomposition
	Guo et al. (2019)	Yes	Static	Min. of total system cost	Route; Passenger assignment	LF; VC; RL; FL; PA	Heuristic and exact algorithms
	Lyu et al. (2019)	No	Static	Max. of profit	Route; Passenger assignment	FC; PA; VC	Heuristic algorithm
	This paper	Yes	Both	Max. of profit; Min. of pax. cost	Route; Schedule; Passenger assignment	VC; TW; PA; PD	Heuristic insertion & Exact algorithm

Table 1-2 Comparison of existing works (other types of DRT)

System type	Publication	Subscr-iption	Demand pattern	Objective	Decision variable	Constraints	Solution algorithm
Other types of DRT	Horn (2002)	Yes	Dynamic	Min. of total travel time & ridership	Route	TW	Heuristic insertion
	Diana & Dessouky (2004)	No	Static	Min. of total distance, excess ride time & idle time	Route	TW	Heuristic insertion
	Cordeau (2006)	No	Static	Min. of routing cost	Route	VC; RL; TW; PD; PA	Exact algorithm
	Coslovich et al. (2006)	No	Dynamic	Min. of dissatisfaction	Route	TW	Heuristic insertion
	Calvo & Colorni (2007)	No	Static	Max. of service quality	Route	PA; LF; VC; TW	Heuristic algorithm
	Dondo & Cerdá (2007)	No	Static	Min. of system cost	Route; Passenger assignment	PA; RL; TW; VC	Heuristic algorithm
	Ilani et al. (2014)	No	Static	Min. # of routes and wasted time	Route; Schedule	VC; TW	Heuristic algorithm

Note: LF: Load factor; VC: Vehicle capacity; TW: Time window; RL: Route length; FL: Fleet size; PA: Passenger assignment: PD: Pickup and delivery; FC: Flow conservation routing and scheduling variables. Cordeau (2006) proposes a model defined on a set of binary three-index variables, which considers both the routing and vehicle assignment. To compact the model, Ropke et al. (2007) define the binary routing variables in two indexes, where the vehicle index and corresponding pairing and precedence constraints are simplified, and it is capable of solving larger instances.

The VRPPD is NP-hard since it is the generalization of VRP (Berbeglia et al., 2007). Only instances with a small number of requests can be solved efficiently by exact algorithms, e.g., branch-and-bound (B&B) (Qiu et al., 2017), branch-and-cut (Cordeau, 2006; Ropke et al., 2007), and branch-and-price (Gutiérrez-Jarpa et al., 2010) algorithms. Most of the exact algorithms are developed based on the B&B framework by adding cutting planes or applying column generation techniques. Other techniques concerning the reduction of the problem scale are also widely applied, such as the Bender's decomposition (Codato & Fischetti, 2006) and the reduction approach (Ilani et al., 2014). Due to the intrinsic complexity of the VRPPD, most of the existing models are solved by heuristics or metaheuristics approaches for dealing with large-scale problems in real-life practices (Dondo & Cerdá, 2007). The two-phase phase heuristics are widely adopted to deal with the instance with a large number of passenger requests, including clustering phase (dividing passengers that have similar trip requests into subsets, each of which is corresponding to a route/vehicle) and routing phase (determining the visiting sequence of each route). Accordingly, two different strategies can be conducted, that is, the cluster-first-route-second approach (Berbeglia et al., 2007; Dondo & Cerdá, 2007; Laporte, 2009), and the route-first-cluster-second approach, which has been verified with poor performance (Cordeau et al., 2007).

2.4 Objectives and contributions

According to the literature, the distinction between static and dynamic CB problems is blurred in practice, especially for the demand uncertainty and the passenger's preference (Cordeau & Laporte, 2007). For instance, in the static case, both request introduction and cancellation happen during the operation. Meanwhile, the dynamic CB problem may contain a number of known requests before the operation. Hence, in an on-demand transit system, the demand pattern is not limited to be static or dynamic. For instance, it is unnecessary to satisfy all requests from the operator's point of view, while the passengers can decide whether to accept or refuse the provided service. There has been some work in formulating the selective dial-a-ride problem based on the principle that the vehicle would visit the request only if it is profitable to serve (Qiu et al., 2017). Though an increased interest in passenger-oriented CB network design problem can be observed recently, the existing scientific literature related to optimization problems in such systems is still relatively scarce. Meanwhile, the emphasis of CB service has turned to passenger satisfaction and the reduction of passenger inconvenience. There are few studies on the modelling of the impacts of passenger's decisions on the routing and scheduling problems. Thus, the design and modelling of transit services should take into consideration the requirements and benefits of both passengers and operators.

In this sense, this problem could be modeled by the bi-level programming problem on the basis of the Stackelberg (leader-follower) game. Nair & Miller-Hooks (2014) developed a bi-level programming model of flexible public transit configuration optimization based on network balance. At the upper level, the operator determines the optimal system configuration, while at the lower level, the passengers optimize their own travel plans. Yu et al. (2015) optimized the route networks of shuttle bus by a bi-level nonlinear mixed integer programming model. The upper level problem optimizes the routing and stopping decisions by minimizing the total system cost, including both operators and passengers. Passengers minimize their walk trips at the lower level.

Given the increasing importance placed on the interaction between passengers and operators, it becomes salient to develop a new framework that considers the passengers' and operators' decisions integratedly. In the passenger-oriented transit service design problem, the decision-making process of the passenger and the operator is hierarchical, the objectives of which are conflictual. In the proposed CB problem, passengers and operators can dynamically exchange information of preservation and vehicle routing and scheduling information on the subscription platform. Nonetheless, existing studies mainly concentrated on the design and optimization of the service network with known demands, which cannot adequately and fully take the advantages of the advanced on-demand service platform. Consequently, the dynamic interaction process between passengers and operators could not be embodied and give rise to various problems such as information lag and too much delay in practice. Despite its practical significance, the modelling of the passenger-oriented transit service, as well as the construction of the information platform is still an open question, since few of the existing studies of the DRT problem have considered the interactive mechanism between the passenger and the operator.

Hence, the contributions of this study are threefold. First, an integrated decision-making framework for the demand-responsive CB network design problem is proposed. Both the objectives of the operator and passenger are considered comprehensively in the CB route design process. Second, a two-phase optimization model is proposed to separate the trip request processing (dynamic phase) and vehicle routing (static phase) problems. Third, the interactive mechanism between the passenger and the operator in the dynamic phase (subscription stage) is modeled by a bi-level programming model. At the upper level, the CB operator optimize the service network; at the lower level, the passengers make mode choices.

3. Problem Statement

As mentioned in the Introduction, in the dynamic phase (subscription stage), each passenger occasionally proposes her/his trip request online with the pickup/destination locations and desired times, based on which the operator can make network design decisions. Potts et al. (2010) indicate that the following two issues should be addressed in the dynamic phase: i) how long in advance the passenger should make the request (i.e., the buffer time); ii) how does the operator negotiate with the passenger for desired pickup/delivery times and locations? For these two issues, a subscription and routing mechanism is proposed as follows for the operator and passengers in the timeline, respectively.

As shown in Fig. 2, the timeline of the CB subscription process considers both the operator and the passengers. For any passenger r, let t_r^P denote her/his desired pickup time. For the operator, some buffer time (e.g., 1 hour) is needed to gather the demand and also optimize the routing/scheduling, and we use BT to denote this buffer time. Let t_r^R denote the time that passenger r submits a request. Hence, $t_r^R < \overline{t}_r^R$, where $\overline{t}_r^R = t_r^R + BT$ is the end-time for request. After submitting the trip request, the passenger would shortly receive the feedback including the offered pickup/delivery times and trip fare, and then decide whether to accept this trip plan. Note that the processing of the passenger request follows the first-come-first-serve principle. The time gap between a passenger receiving the feedback and making the decision is neglected.

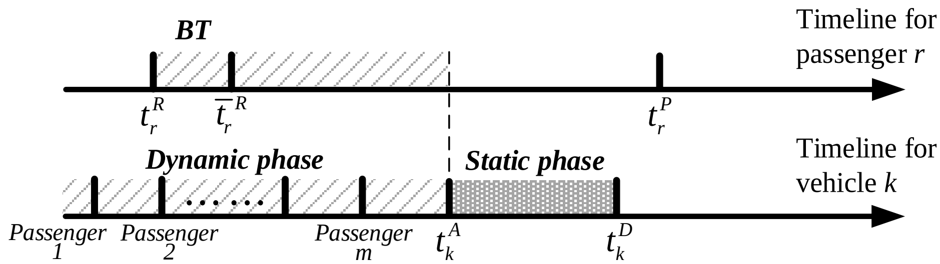


Figure 2. Timelines of the CB network design process

As to the operator, the current routing plan needs to be modified dynamically when new requests occur. As shown in Fig. 2, assume that passenger r is assigned to vehicle k, let t_k^D and t_k^A denote the vehicle's scheduled departure time from the depot and the end-time of receiving requests. The previous passengers that have confirmed their subscription are considered as fixed and the routing plan is not allowed to change at time t_k^D. To sum up, the CB network design problem addressed in this paper can be described as a two-phase procedure: i) in the dynamic phase (i.e., $t \le t_k^A$), the passengers occasionally subscribe services; and the operator needs to estimate the trip costs, offer a service plan and price, and then communicate with each passenger to confirm the trip. To provide a profitable service network, a model is needed for the operator to dynamically insert each trip request into the current service network; ii) in the static phase (i.e., $t_k^A < t \le t_k^D$), the subscription process is terminated, and the bus service network is holistically optimized based on the confirmed passenger demand.

4. The two-phase optimization model of CB system

4.1 Network formulation

Consider a graph $G = (V, A)$, where $V = \{v_0, v_1, ..., v_n\}$ is the set of vertices and $A = \{(v_i, v_j) : v_i, v_j \in V, i \ne j\}$ is the set of links. The vertex set V comprises three subsets: pickup vertex set V_p, delivery vertex set V_d, and depot v_0. Additionally, let R denote the set of new arising requests, and V_r the vertex set containing the spatial information of request $r \in R$, i.e., $V_r = \{v_i^{r,p}, v_j^{r,d}\}$, and $v_i^{r,p} \in V_p$, $v_j^{r,d} \in V_d$. Each request is also associated with a desired pickup time t_r^p and delivery time t_r^d. The cumulative number of requests between an origin-destination (OD) pair (v_i, v_j) is denoted by q_{ij}. The fleet of homogeneous vehicles is denoted as K; all vehicles have the same capacity cap. Let J_k denote the route served by vehicle $k \in K$, which can be represented by a set of vertices, V_k, and $V_k = \{v_i \,|\, (v_i, v_j) \in J_k, v_j \in V\}$. Table 2 lists the sets, indices, and parameters used in the following sections.

Table 2. List of notations

Notation	Description
Sets	
A	set of links
K	set of vehicles
R	set of requests
V	set of vertices
V_p	set of pickup vertices
V_d	set of delivery vertices
V_k	set of vertices on route J_k, $k \in K$
Parameters	
cap	capacity of the vehicle
c_{ij}^0	free flow travel time between OD pair (v_i, v_j)
d_{ij}	distance between OD pair (v_i, v_j)
$q_{\min}$	minimum load factor
T	total revenue
$\tilde{T}$	expected total revenue
$t_{\max}$	time deviation threshold
t_r^p, t_r^d	desired pickup/delivery times of request r at the pickup vertex v_i and delivery vertex v_j
t_{ij}	travel time between vertex v_i and v_j
$v_i^{r,p}$, $v_j^{r,d}$	desired pickup/delivery vertices of request r which are located at vertices v_i and v_j

α	operating cost per unit travel distance
β	dispatching fee of a vehicle
γ	variance parameter
λ	passenger's value of time
μ	monetary penalty on the time deviation
τ_{ij}^{r}	transit fare of request r between vertices v_i and v_j
τ_{ij}'	regular distance-based fare between vertices v_i and v_j
$\bar{\tau}$	distance-based fare rate
τ_0	fare of dispatching an additional vehicle
Variables	
N_{ij}^{k}	number of passengers assigned to vehicle k between OD pair (v_i, v_j)
$t_r^{p'}$, $t_r^{d'}$	actual pickup/delivery times of request r at the pickup vertex v_i and delivery vertex v_j
$t_i^{k,A}$, $t_i^{k,D}$	offered arrival and departure times of vehicle k at vertex v_i
x_r^{k}	request-to-vehicle variable (equals to 1, if request r is assigned to vehicle k, and 0, otherwise)
y_{ij}^{k}	routing variable (equals to 1, if route segment (v_i, v_j) is traveled by vehicle k, and 0, otherwise)
δ_k	vehicle dispatching variable (equals to 1, if vehicle k is dispatched, and 0, otherwise)

4.2 The Dynamic Phase

In the dynamic phase, new requests arrive occasionally. As mentioned in the Introduction, the dynamic phase includes two decision-making problems: the operator plans the service system to maximize its profit; while passengers, based on their perceived travel costs, decide whether to accept or reject the offered trip plans to maximize their trip utilities. In the following sections, a bi-level programming model is proposed to cope with these two decision-making problems systematically: in the upper-level, the operator acts as the leader who designs the service network; the lower-level formulates the follower's decision-making problem based on the offered services. The passengers' travel behavior is analyzed through the discrete choice model giving the SUE principle. Mathematical formulations of the bi-level model are provided as follows.

4.2.1 The upper-level problem

The upper-level problem is to design the CB service network concerning the real-time requests by inserting them into the existing CB network or launching a new route for these new passengers. Each request r includes four values: the desired pickup/delivery vertices $v_i^{r,p}$ and $v_j^{r,d}$, and associated times t_r^p and t_r^d. A penalty incurs when the desired pickup/delivery time is violated. To minimize the time deviation and increase the system serviceability, we use a time deviation threshold $t_{\max}$ for each request. For the request r, the offered arrival time $t_i^{k,A}$ of the assigned vehicle k at vertex v_i should follow the time interval $[t_r^p - t_{\max}, t_r^p + t_{\max}]$. The boarding/alighting time of a passenger is assumed as zero. The overlap of multiple arrival intervals occurs when several passenger requests appear at the same pickup vertex in a short time.

Assume that n passengers are assigned to vehicle k at the vertex v_i. Let t_1^P and t_n^P denote the desired pickup times of the first and last passengers by sorting them with their desired pickup times, respectively. As shown in Fig. 3, to satisfy the feasible arrival time intervals for all passengers, the vehicle should arrive no later than the latest pickup time of the first passenger ($t_1^p + t_{\max}$). And the departure time should be no earlier than the earliest pickup time of the last passenger ($t_n^p - t_{\max}$). Thus, the arrival and departure times of vehicle k at vertex v_i should satisfy

$$t_i^{k,A} \le \min_{r \in R}\{t_r^p\} + t_{\max}, \quad \forall v_i \in V_p \cap V_r, \tag{1}$$

$$t_i^{k,D} \ge \max_{r \in R}\{t_r^p\} - t_{\max}, \quad \forall v_i \in V_p \cap V_r. \tag{2}$$

At a delivery vertex, a lateness penalty is considered. For a delivery vertex v_j, the arrival time of vehicle k should not be later than the latest delivery time of the earliest passenger. Hence,

$$t_j^{k,A} \le \min_{r \in R}\{t_r^d\} + t_{\max}, \quad \forall v_j \in V_d \cap V_r. \tag{3}$$

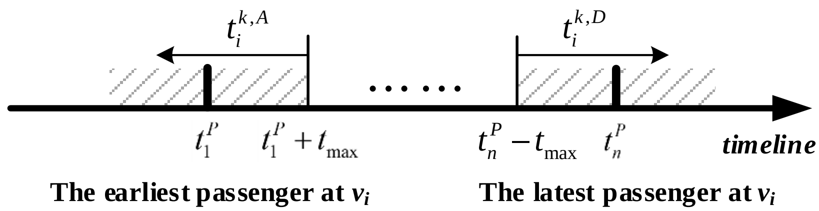


Figure 3. Illustration of the constraints on arrival and departure times

After determining the vehicle departure and arrival times at each vertex, the actual pickup and delivery times of request r, $t_r^{p'}$ and $t_r^{d'}$, can be obtained as follows:

- for a pickup vertex $v_i \in V_p$:

$$t_r^{p'} = \begin{cases} t_r^p, & t_r^p \in [t_i^{k,A}, t_i^{k,D}] \\ t_i^{k,A}, & t_r^p < t_i^{k,A} \\ t_i^{k,D}, & t_r^p > t_i^{k,D} \end{cases}, \tag{4}$$

- for a delivery vertex, $v_j \in V_d$:

$$t_r^{d'} = \begin{cases} t_j^{k,A}, & t_r^d \le t_j^{k,A} \\ t_r^d, & otherwise \end{cases}. \tag{5}$$

As aforementioned, serving the passengers in low-demand areas has a risk of the deficit. To ensure a profitable CB system, an additional fare τ_0, is charged when the number of confirmed passengers does not meet the requirement of minimum load factor $q_{\min}$. Moreover, the distance-based fare scheme is adopted to encourage the wise routing decisions of the passenger. Thus, the CB fare of passenger request r between OD pair (v_i, v_j), denoted by τ_{ij}^r, can be defined as follows:

$$\tau_{ij}^r = \begin{cases} \tau'_{ij} + \tau_0, & N_{ij}^k \le q_{\min} \\ \tau'_{ij}, & N_{ij}^k > q_{\min} \end{cases}, \forall r \in R, v_i \in V_p, v_j \in V_d, \tag{6}$$

where τ'_{ij} is a regular distance-based fare between vertices v_i and v_j, and $\tau'_{ij} = \bar{\tau} \cdot d_{ij}$, where $\bar{\tau}$ is the fare rate and d_{ij} is the distance between v_i and v_j. N_{ij}^k is the number of passengers assigned to vehicle k between OD pair (v_i, v_j). As a result, the total revenue (T) collected from the CB system is $T = \sum_{v_i \in V_p} \sum_{v_j \in V_d} \sum_{r \in R} \tau_{ij}^r$.

The upper level concerns the operator's decision-making problem on the network design. According to the general scheme of the CB network design problem, the variables are partitioned into two subsets. The first set of variables are related to the design of CB routes to serve passengers. The binary variables x_r^k takes the value of 1 if request r is assigned to vehicle k, and 0 otherwise. The binary variables y_{ij}^k assume a value of 1 if route segment (v_i, v_j) is traveled by vehicle k, and 0 otherwise. The binary variables δ_k indicate whether vehicle k is dispatched or not.

The second set of variables is related to vehicle scheduling including the arrival/departure times at each vertex, which has been defined in Eqs. (4) and (5). For simplicity, we adopt $\mathbf{N}$ as the vector of decision variables in the upper-level, and $\mathbf{N} = \{x_r^k, y_{ij}^k, \delta_k, t_i^{k,A}, t_i^{k,D}, t_j^{k,A}\}$. Considering the stochasticity in the passenger's choice of accepting the offered CB service, the revenue is rationally assumed to be a random variable. Let $\mathbf{P}$ denote the vector of the probabilities that passengers accept the offered CB trip plan. Evidently, the expected revenue ($\tilde{T}$) with respect to $\mathbf{P}$ giving specific CB network design decisions $\mathbf{N}$ is $\tilde{T} = E\left[\mathbf{N}, \mathbf{P}\left(\mathbf{N}\right)\right]$, where $E(\cdot)$ is the expectation operator. In sum, the objective function of the upper-level can be formulated as maximization of the operator's profit equal to the expected total revenue minus the operating cost.

The upper-level is formulated as a mixed nonlinear integer program:

$$\max z_1(\mathbf{N}) = E\left[\mathbf{N}, \mathbf{P}\left(\mathbf{N}\right)\right] - \alpha \cdot y_{ij}^k \cdot d_{ij} - \beta \cdot \sum_{k \in K} \delta_k \tag{7}$$

subject to (1)-(3),

$$\sum_{r \in R} x_r^k \le cap, \ \forall k \in K, \tag{8}$$

$$\sum_{k \in K} x_r^k = 1, \ \forall r \in R, \tag{9}$$

$$\sum_{v_j \in V} y_{0j}^k = \sum_{v_i \in V} y_{i,2n+1}^k = 1, \ \forall k \in K, \tag{10}$$

$$x_r^k \in \{0,1\}, \ \forall r \in R, \ k \in K, \tag{11}$$

$$y_{ij}^k \in \{0,1\}, \ \forall (v_i, v_j) \in A, \ k \in K, \tag{12}$$

$$\delta_k \in \{0,1\}, \ \forall k \in K, \tag{13}$$

where α and β are the operating cost per unit travel distance and the fixed cost of dispatching an additional vehicle.

Eqs. (1)-(3) define the arrival and departure times at pickup/delivery vertices satisfying maximum time deviation constraints. Eq. (8) refers to the capacity constraint; i.e., the number of passengers in a vehicle should not exceed its capacity. Eq. (9) ensures that each request is served by exactly one vehicle. Eq. (10) guarantees that each route starts and ends at the depots. Eqs. (11)-(13) define the binary variables.

4.2.2 The lower-level problem

The lower level intends to describe the passenger's travel behavior, which gives the variables needed in objective function at the upper level. Given the offered trip plan with pickup/delivery times and trip fare, each passenger needs to choose whether to accept it or reject it and shift to other travel modes. Hence, the passenger's choice behavior should be characterized by the discrete choice model, which follows the SUE principle with elastic demand (Sheffi, 1985; Meng & Liu, 2011). Given a specific trip plan based on the current CB network $\mathbf{N}$ from the operator, passengers have a perception error on the travel cost when making decisions. The passenger's perceived travel cost of accepting the offered CB service, denoted by $\tilde{C}_{ij}(\mathbf{N})$ is therefore a random variable following a certain distribution. The travel utility between an OD pair (v_i, v_j), denoted by $\tilde{U}_{ij}(\mathbf{N})$, equals

$$\tilde{U}_{ij}(\mathbf{N}) = \bar{U}_{ij} - \tilde{C}_{ij}(\mathbf{N}), \tag{14}$$

where $\bar{U}_{ij}$ is a constant representing the maximum benefit that passengers can gain from the CB trip. Hence, for a certain passenger, if her/his perceived travel utility of accepting the offered CB service is negative, s/he will reject the trip plan and shift to other travel modes.

With the assumption of the SUE principle for passenger's choice behavior, given the offered trip plan and CB network $\mathbf{N}$, the passenger's perceived travel cost between an OD pair (v_i, v_j) is modeled as the summation of a systematic term $c_{ij}(\mathbf{N})$ and an error term ξ_{ij},

$$\begin{aligned} &\tilde{C}_{ij}(\mathbf{N}) = c_{ij}(\mathbf{N}) + \xi_{ij}, \\ &\xi_{ij} \sim N\left(0, \gamma c_{ij}^0\right), \end{aligned} \tag{15}$$

where error term ξ_{ij} is a normally distributed random variable with zero mean and distance-independent variance equal to γc_{ij}^0. Hence, c_{ij}^0 is the free flow travel time between OD pair (v_i, v_j) and constant γ is termed as variance parameter.

The systematic travel cost is composed of three components: in-vehicle travel cost, penalties on time deviations at origin and destination, and trip fare. Given the offered pickup and delivery times from the upper-level, t_i^p and t_j^d at origin v_i to destination v_j respectively, the systematic travel cost can be obtained as follows,

$$c_{ij}(\mathbf{N}) = \lambda \cdot t_{ij} + \mu \cdot (TV_i^p + TV_j^d) + \tau_{ij}, \tag{16}$$

where λ is the value of travel time, t_{ij} is the travel time between an OD pair (v_i, v_j), which is associated with the proposed trip plan and network design decisions from the operator. TV_i^p and TV_j^d are time deviations, which are calculated by the difference between actual and desired pickup and delivery times. μ denotes the monetary penalty on time deviations.

As aforementioned, the passenger will accept the proposed trip plan when the perceived travel cost is lower than the maximum benefit, which gives rise to the elasticity of passenger demand. The probability that passengers choose the trip plan P^{CB} is the probability that the travel utility is larger than zero, namely,

$$P^{CB} = \Pr\left[\tilde{C}_{ij}(\mathbf{N}) < \bar{U}_{ij}\right], \ \forall v_i, v_j \in V, \tag{17}$$

where the passengers' travel utility $\tilde{U}_{ij}(\mathbf{N})$ is defined on her/his perceived travel costs based on the deterministic travel cost $c_{ij}(\mathbf{N})$, which is assumed as a random variable covering the population variation.

4.2.3 Solution algorithm for the bi-level programming model

In the proposed model of the dynamic phase, the operator should rapidly respond to the request by inserting new requests into the existing network. In view of the occasional arrivals of new requests, two approaches are widely applied to address the dynamic vehicle routing problem: (1) solving the static problem each time when the new request is proposed, and (2) solving the static vehicle routing problem only once in the initial stage and then updating the current solution when new request is proposed by heuristic methods (e.g., insertion, deletion and interchange heuristics) (Berbeglia et al., 2010). Considering the interactive decision-making manner between the operator and passengers, this paper develops a dynamic insertion approach to address the proposed bi-level model. Specifically, assume that there is a set of routes generated based on historical demands. When a new request is proposed, the operator scans the historical route set to find feasible insertion options. If no feasible insertion can be conducted based on historical routes, a new route should be generated specifically for this new request. Note that the new insertion should not incur any violation on the passengers that have confirmed their services.

In sum, an insertion checking algorithm is developed to find feasible insertion schemes on the current CB network.

Algorithm 1. The insertion checking algorithm

Input: A new request $r=\left\{v_i^{r,p},v_j^{r,d},t_r^p,t_r^d \middle| v_i^{r,p}\in V_p, v_j^{r,d}\in V_d\right\}$

Step 1: If $v_i^{r,p}$ and $v_j^{r,d}$ already exist in an existing route J_k, and the current arrival and departure time at $v_i^{r,p}$ and $v_j^{r,d}$ of route J_k, $t_i^{k,A}$, $t_i^{k,D}$, and $t_j^{k,A}$ are within the acceptable time intervals, then request r can be directly insert into route J_k. If so, record it as a feasible insertion scheme.

Step 2: If the delivery vertex of request r, $v_j^{r,d}$, already exists in route J_k, but the pickup vertex $v_i^{r,p}$ is not in route J_k, then apply the checking process in ***Step 2.1*** for $v_j^{r,d}$ and scan all existing vertices $v_m\in V_k$ in route J_k for inserting $v_i^{r,p}$:

Step 2.1: If there is no passenger at v_m currently, then v_m can be replaced by $v_i^{r,p}$: Remove v_m from J_k and add $v_i^{r,p}$ to J_k, hence the time interval for serving vertex $v_i^{r,p}$ can be expressed as: $[t_{m-1}^{k,D}+t_{m-1,i}, t_{m+1}^{k,A}-t_{i,m+1}]$, where . Check whether this time interval intersects the acceptable time interval $[t_r^p-t_{\max}, t_r^p+t_{\max}]$. If so, record it as a feasible insertion scheme.

Step 2.2: If there are passengers at v_m currently, then v_m cannot be removed. Check if $v_i^{r,p}$ can be inserted into the place after v_m. The time interval for the serving vertex $v_i^{r,p}$ can be expressed as: $[t_m^{k,D}+t_{m,i}, t_{m+1}^{k,A}-t_{i,m+1}]$. Check whether this time interval intersects the acceptable time interval $[t_r^p - t_{\max}, t_r^p + t_{\max}]$. If so, record it as a feasible insertion scheme.

Step 3: If the pickup vertex of the request r, $v_i^{r,p}$, already exists in route V_k, but the delivery vertex $v_j^{r,d}$ is not in route J_k, then apply the checking process in ***Step 3.1*** for $v_i^{r,p}$ and scan all existing vertices $v_m \in V_k$ in route J_k for inserting $v_j^{r,d}$:

Step 3.1: If there is no demand at v_m currently, than v_m can be replaced by $v_j^{r,d}$: Remove v_m from J_k and add $v_j^{r,d}$ to J_k, hence the time interval for the serving vertex $v_j^{r,d}$ can be expressed as: $[t_{m-1}^{k,D} + t_{m-1,i}, t_{m+1}^{k,A} - t_{i,m+1}]$. Check whether this time interval intersects the acceptable time interval $[t_r^d + t_{\max}, +\infty)$. If so, record it as a feasible insertion scheme.

Step 3.2: If there are passengers at v_m currently, then v_m cannot be removed. Check if $v_j^{r,d}$ can be inserted into the place after v_m. The time interval for the serving vertex $v_j^{r,d}$ can be expressed as: $[t_m^{k,D} + t_{m,i}, t_{m+1}^{k,A} - t_{i,m+1}]$. Check whether this time interval intersects the acceptable time interval $[t_r^d + t_{\max}, +\infty)$. If so, record it as a feasible insertion scheme.

Given the above insertion checking subroutine of a new request r, a dynamic insertion algorithm is proposed as follows:

Algorithm 2. The dynamic insertion algorithm

Step 1: Initialization.

Input the set of existing routes J_k with the list of visiting vertices V_k for each k and the related arrival and departure times, $t_i^{k,A}$ and $t_i^{k,D}$, at each vertex $v_i \in V_k$. Input the newly received request r with its pickup/delivery vertices $v_i^{r,p}$ and $v_j^{r,d}$, and desired pickup/delivery times t_r^p and t_r^d.

Step 2: Searching for feasible insertion schemes

Step 2.1: For each historical route J_k, apply the Algorithm 1, record every feasible insertion scheme;

Step 2.2: If no feasible insertion scheme can be found in the insertion checking process, generate a new route for request r from depot v_0 to $v_i^{r,p}$ and $v_j^{r,d}$.

Step 3: Evaluation of feasible insertion schemes

Step 3.1: For each feasible insertion scheme obtained in ***Step 2***, calculate the profit of operators and the general travel cost of the passenger, calculate the probability of the passenger to choose this scheme, then obtain the expected profit of the scheme;

Step 3.2: Store the insertion scheme with the highest expected profit, update historical routes J_k, the set of visiting vertices V_k for each k, and the arrival and departure times, $t_i^{k,A}$ and $t_i^{k,D}$, at each vertex $v_i \in V_k$.

Step 4: If there is a new request submitted to operators, go to ***Step 1***; otherwise, end.

By applying this algorithm to the new requests, a set of new CB routes can be designed at the end of the dynamic phase. Meanwhile, the passengers receive service information and decide whether to confirm the provided CB services. The confirmed passengers are then considered in the re-optimization of the static phase.

4.3 The Static Phase

In view of the demand elasticity in the dynamic phase, the operator cannot obtain the actual demands when designing the CB network. Evidently, the solution of the network design problem that has obtained so far is suboptimal. Whereas in the static phase, no new requests are allowed to input into the current CB network (see Fig. 2), and the passenger demand is considered known and fixed. In this regard, it is necessary for the operator to re-optimize the CB network. Such network design, taking into consideration the pickup/delivery times that have been confirmed jointly by the operator and passenger, is formulated as a static VRPPD with hard time constraints. Besides, the transit fare that has also been confirmed by the passenger would not change, while the revenue of the operator is taken as a constant. The objective function can be simplified to the minimization of the operation cost. To obtain the exact solution of the model in a reasonable time, this section develops a B&B algorithm based on the graph search strategy.

4.3.1 Static CB network design problem

As aforementioned, the total revenue in Eq. (7) is known and constant. The objective function of this problem can be rewritten as follows:

$$\min z_2 = \alpha \cdot \sum_{k\in K} \sum_{v_i, v_j \in V_k} y_{ij}^k \cdot d_{ij} + \beta \cdot \sum_{k \in K} \delta_k \ . \tag{18}$$

Besides, the time deviations in the static phase are determined which do not allow any lateness. Therefore, in the model of the static phase, the time deviation constraints (3)-(4) should be rewritten as:

$$x_r^k \cdot t_i^{k,D} \le \min_{r\in R}\{t_r^{p'}\},\ \forall i \in V_p \bigcap V_r\,, \tag{19}$$

$$x_r^k \cdot t_j^{k,A} \le \min_{r\in R}\{t_r^{d'}\},\ \forall j \in V_d \bigcap V_r\,, \tag{20}$$

where $t_r^{p'}$ and $t_r^{d'}$ are the pickup and delivery times provided by operators in the dynamic phase. The other constraints are the same as Eqs. (8)-(13) described in Section 3.2.1.

The problem we introduced here has clear relationships with other families of routing problems, such as the traveling salesman problem and the dial-a-ride problem. The original dial-a-ride problem has been shown to be NP-hard (Baugh et al., 1998). Therefore, it is important to address the computational complexity of the proposed problem which can be seen as a modified scenario of the original dial-a-ride problem. The following proposition is provided, the proof of which is in the **Appendix A**.

Proposition 1. The static CB network design problem is NP-hard.

4.3.2 Solution algorithm in the static phase

The B&B algorithm is one of the most successful exact approaches to solve the combinatorial optimization problem. It intends to find the optimal solution by reducing the search space dynamically based on the tree searching strategy. In general, the B&B algorithm is composed of three main aspects: i) the branching strategy: splitting the search space into smaller spaces recursively; ii) the lower bound: evaluating each node of the tree; iii) the exploration strategy: after each node evaluation, specifying the node to be processed for the next branching.

In view of the NP-hardness of the proposed optimization model, the efficiency in handling the network design problem is highly desirable, especially in the practical implementations. Here we follow the "cluster-first-route-second" scheme to reduce the searching space for solving large-scale problems by classifying passengers with similar temporal and spatial requirements (Tong et al., 2017). Note that in the dynamic phase, requests with compatible pickup/delivery times at each pickup vertex have been assigned to the same vehicle. Hence, the results of the request assignment can be treated as passenger groups inputting to the B&B algorithm.

In this algorithm, a route is represented by an integer sequence of pickup and delivery vertices. we define $\tilde{P}$, $\tilde{D}$ and $\tilde{C}$ as the sets of unvisited pickups, undelivered passengers, and vertex list of the current route. At each vertex v_i, three possible operations are modeled as different operations, which are:

(1) Pick up a new request: add v_i into $\tilde{D}$ and $\tilde{C}$, remove v_i from $\tilde{P}$;

(2) Deliver a request: add v_i into $\tilde{C}$, remove v_i from $\tilde{D}$;

(3) Dispatch a new vehicle: if $\tilde{D}$ is empty, add v_0 to $\tilde{C}$.

In the proposed B&B algorithm, the above three possible operations can be translated into branches of the search tree (Qiu et al., 2017). A depth-first search strategy is applied to generate routes, which is bounded by the time deviation constraints (19) and (20), and the capacity constraint (8). To reduce computational burden, the current solution is compared to the optimal solution and is abandoned if its theoretical lowest cost is higher than the cost of the optimal solution. Given a predetermined lowest cost of serving a request, the theoretical lowest cost is obtained by summing up the cost of the current solution and the lowest cost of serving remaining requests. In the multi-vehicle case, a vehicle counter k is applied to record the number of vehicles that have been processed, while an additional branch is added when a vehicle finishes a trip. Accordingly, the graph search algorithm can be designed recursively if a new vehicle is needed. The detailed solution algorithm for the CB service network optimization in the static phase is as follows:

Algorithm 3. The graph search algorithm

Step 1: Initialization.

Set $\tilde{P} = \{v_1, ..., v_n\}$, $\tilde{D} = \varnothing$, and generate the current route $\tilde{C} = \{v_0\}$. Set the current lowest cost $c_{\min} = infinity$ and the current optimal solution $R_{opt} = \varnothing$.

Step 2: Graph searching:

Step 2.1: Check the feasibility of the generated route:

- If the time deviation constraints (19) and (20) or the capacity constraint (8) is violated, return.
- If the current theoretical lowest cost is higher than the cost of the optimal solution, return.

Step 2.2: Check whether there exist remaining requests:

- If $\tilde{D}$ and $\tilde{P}$ are empty, calculate the total cost using Eq. (18). If the current cost is lower than the current lowest cost, update the current lowest cost $c_{\min} = c_{current}$ and the current optimal solution $R_{opt} = R_{current}$. Return.

Step 2.3: Generate all possible combinations of routes. For each vertex $v_i \in \tilde{P} \cup \tilde{D}$:

- If $v_i \in \tilde{P}$, pick up this new request, call the **Algorithm 3**, remove v_i from $\tilde{D}$ and $\tilde{C}$, add v_i to $\tilde{P}$;
- If $v_i \in \tilde{D}$, call the **Algorithm 3**, remove v_i from $\tilde{C}$ and add v_i to $\tilde{D}$;
- If $\tilde{D}$ is empty, update the current solution $R_{current}$ and cost $c_{current}$, add v_0 to $\tilde{C}$, call the **Algorithm 3**, remove v_0 from $\tilde{C}$.

5. Numerical Examples

Two numerical examples are conducted to illustrate the properties of the proposed network design method and the effectiveness of the algorithm applied. The algorithms were coded in the Visual C++ language and executed on a personal computer (Intel Core i7 CPU @ 2.2GHz).

5.1 Numerical test

The well-known Sioux Falls network is first utilized to verify the effectiveness of the proposed model and algorithms. As shown in Fig. 4, the network has 24 vertices and 38 bidirectional links. The number on links denote the link travel time. The planning horizon starts from 6 to 9 a.m., which is equally discretized into six-minute time intervals. Two historical routes are depicted using blue arrows. The visiting sequence and corresponding arrival and departure times of the two historical routes are shown in Table 3. The area highlighted by the green ellipse is the CBD district, the vertices in which are passengers' possible delivery locations, namely vertices 10, 11, 14, 15. All other vertices are possible pickup locations. The vehicle capacity cap is 10. The minimum load factor $q_{\min}$ is 5. The other parameters used are: $\alpha = \$10$, $\beta = \$20$, $\lambda = \$2$, $\mu = \$4$.

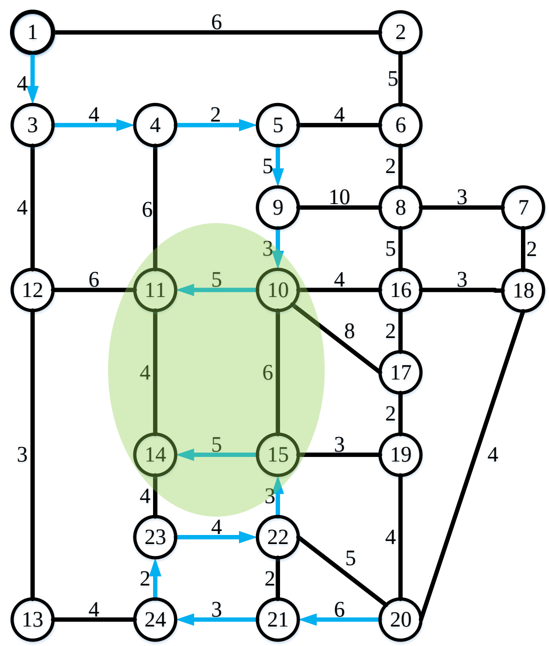


Figure 4. The example of Sioux Falls network

Table 3. Historical CB route information

Route No.	Visiting vertex ID	Arrival time	Departure time
	1	-	2
	4	10	12
1	5	14	16
	10	24	24
	11	29	-
	20	-	0
	21	6	7
2	23	12	16
	15	23	23
	14	28	-

5.1.1 Experimental setup

We test our proposed model and solution algorithm with random instances of requests, where the passengers' requests are sampled as follows. The passenger's pickup/delivery locations and times are generated following the uniform distribution. We generate 36 instances for each number of requests and consider different combinations of the passenger's heterogeneities: (a) regular and (b) irregular CB passengers, (c) higher value of time and (d) higher penalty on time deviation. Herein, we assume that the passenger's perceived travel cost follows the normal distribution, where the regular CB passengers have smaller variance. The relationship between the passenger's value of time and the penalty on time deviation reflects the passenger's willingness to accept greater time deviations to reduce in-vehicle travel time. To shorten the algebra, we define the parameter $\eta = \lambda / \mu$ as the heterogeneity in the ratio of the value of time over the penalty on time violation. Hence, the passenger's perceived travel cost is generated based on Eq. (15) in the form $\tilde{C}_{ij} = \eta \cdot t_{ij} + (TV_i^p + TV_j^d) + \tau_{ij} + \xi_{ij}$, with ξ_{ij} following the normal distribution with zero mean and variance σ^2.

5.1.2 Optimal results

Table 4 reports the computational results with the number of requests goes from 5 to 30. The number of CB routes needed in the dynamic and static phases are reported in column 3 and 4. It reveals that the number of routes that needed to serve all the confirmed request can be largely reduced in the static phase. As the number of requests increases, the number of CB routes increases, reaching a maximum value of 5 when most of the vertices in the Sioux Falls network is covered. From the economic point of view, the operating cost can be reduced by 22.8% on average during the re-optimization of the service network. Evidently, the average transit fare decreases as the number of requests rises. In brief, this reveals that the CB service preserves the characteristic of the shared mobility services while the ridesharing fare decreases with more participants.

Table 3. Historical CB route information

Route No.	Visiting vertex ID	Arrival time	Departure time
1	1	-	2
	4	10	12
	5	14	16
	10	24	24
	11	29	-
2	20	-	0
	21	6	7
	23	12	16
	15	23	23
	14	28	-

Table 4. Optimal results of the numerical example

# of requests	# of confirmed requests	# of CB routes		Revenue ($)	Ave. fare ($)	Operating cost ($)		CPU time (s)
		Phase D[a]	Phase S[a]			Phase D	Phase S	
5	5	3	1	103	20.6	106	67	0.15
10	9	4	1	209	20.9	120	81	3.44
15	13	7	2	300	20	182	146	21.57
20	17	9	5	389	19.5	219	178	95.11
25	19	12	5	445	17.8	216	192	199.6
30	21	15	5	503	16.8	214	175	320.7

Note: [a]Phases D and S refer to the dynamic and static phases, respectively.

To better reflect the passengers' choices, the detailed passengers' trips information of an instance with 30 requests is presented in Table 5. It is shown that passengers are more likely to reject the offered trip plan with pickup lateness. Otherwise, if the offered pickup time is earlier than the desired pickup time, passengers intend to accept the trip plan to board on the bus earlier than they desired. It also reveals a plausible result that the optimization process of static phase can efficiently improve the CB system's level of service by guaranteeing the on-schedule pickups and earlier deliveries than they desired, which is essential for morning commutes.

Table 5. Results of the passenger's travel information in the instance of 30 requests

Passenger ID	Passenger choices	Travel time		Time violation			
		Phase D	Phase S	Phase D		Phase S	
				Pickup	Delivery	Pickup	Delivery
1	Accept	25	19	0	0	0	6
2	Accept	25	19	0	-1	0	6
3	Accept	25	19	0	-2	0	6
4	Accept	22	19	0	0	0	6
5	Accept	11	8	0	1	1	4
6	Accept	8	8	0	5	1	4
7	Accept	8	8	0	3	1	4
8	Accept	25	12	0	3	1	13
9	Accept	25	12	0	4	1	13
10	**Reject**	**15**	**-**	**0**	**0**	**-**	**-**
11	Accept	23	19	0	0	0	4
12	**Reject**	**11**	**-**	**0**	**0**	**-**	**-**
13	Accept	14	10	0	2	1	6
14	**Reject**	**17**	**-**	**0**	**0**	**-**	**-**
15	**Reject**	**27**	**-**	**0**	**0**	**-**	**-**
16	Accept	26	12	0	0	2	14
17	**Reject**	**14**	**-**	**0**	**0**	**-**	**-**
18	Accept	19	19	0	3	0	3
19	Accept	19	19	0	1	0	3
20	Accept	17	17	0	5	0	5
21	Accept	19	12	-3[a]	4	2	14
22	**Reject**	**16**	**-**	**0**	**0**	**-**	**-**
23	**Reject**	**15**	**-**	**0**	**0**	**-**	**-**
24	Accept	13	8	0	0	0	5
25	**Reject**	**13**	**-**	**-3**	**0**	**-**	**-**
26	Accept	25	10	5	-4	1	6
27	Accept	25	10	5	-4	0	6
28	Accept	25	19	0	-2	0	4
29	**Reject**	**13**	**-**	**-8**	**2**	**-**	**-**
30	Accept	13	12	-2	2	0	4

Note: [a]The positive and negative values denote the amount of earliness and lateness, respectively.

Table 6 compares the different combinations of weights on passengers' in-vehicle travel time and deviations in pick/delivery times concerning passengers' choice behavior. The regular passengers who have an accurate estimation of the travel cost are unaffected by the change of travel time or schedule deviations in their total travel costs. Meanwhile, the irregular commuters, who are new to the CB system or do not take the CB service in daily commutes, are more sensitive to the time deviations on their desired pickup/delivery times.

Table 6. Comparisons between the regular and irregular passengers.

	(a) Regular passengers $\sigma^2 = 0.2$								
	$\eta = 0.5$			$\eta = 1.0$			$\eta = 1.5$		
# of Requests	# of confirmed requests	Ave. fare ($)	Ave. travel cost ($)	# of confirmed requests	Ave. fare ($)	Ave. travel cost ($)	# of confirmed requests	Ave. fare ($)	Ave. travel cost ($)
5	5	20.6	64.6	5	20.6	64	5	20.6	64.3
10	9	18.8	55.2	10	20.9	55.2	9	20.9	55.35
15	14	18.8	51.2	13	20.9	55.2	10	18.8	55.35
20	18	23.8	50.3	18	23.8	59.7	18	23.8	69.0
25	24	26.8	52.8	23	26.3	62.2	22	25.7	71.68
30	28	26.7	50.2	28	26.7	59.5	27	25.1	68.7

	(b) Irregular passengers $\sigma^2 = 0.5$								
	$\eta = 0.5$			$\eta = 1.0$			$\eta = 1.5$		
# of Requests	# of confirmed requests	Ave. fare ($)	Ave. travel cost ($)	# of confirmed requests	Ave. fare ($)	Ave. travel cost ($)	# of confirmed requests	Ave. fare ($)	Ave. travel cost ($)
5	5	20.6	53.2	5	20.6	53.2	5	20.6	53.5
10	9	16.9	45.75	10	20.9	55.2	9	16.9	55.35
15	13	23.07	51.2	14	24.3	60.5	11	20	60.6
20	16	21.8	50.3	16	21.8	50.2	19	24.7	69.0
25	21	25	52.8	19	23.4	62.2	23	26.3	71.7
30	25	26.2	50.2	28	25.2	59.3	27	25.1	68.5

5.1.3 Sensitivity analysis

Finally, we perform a sensitivity analysis aiming to identify the relationship between the time deviation threshold and other operating decisions, e.g. the total number of CB routes. As discussed in Section 4.1, the time deviation threshold $t_{\max}$ is adopted to guarantee the system serviceability by limiting the time deviation from the passenger's desired pickup/delivery time. Fig. 5 illustrates the impact of $t_{\max}$ on the number of CB routes and the average choice probability of 20 passenger requests. It is shown that the number of CB routes decreases with the increase of time deviation threshold when $t_{\max} \leq 7$, which reflects that a higher permitted time violation makes more requests able to be served by historical routes. When $t_{\max} > 7$, the requests submitted early can all be served by historical routes, making historical routes unreasonable and incapable to serve requests submitted relatively late. The average passenger's choice probability of CB is also decreased sharply with the increase of $t_{\max}$, implying that allowance of larger time deviation would result in an unreliable service network that passengers are unlikely to accept the offered services.

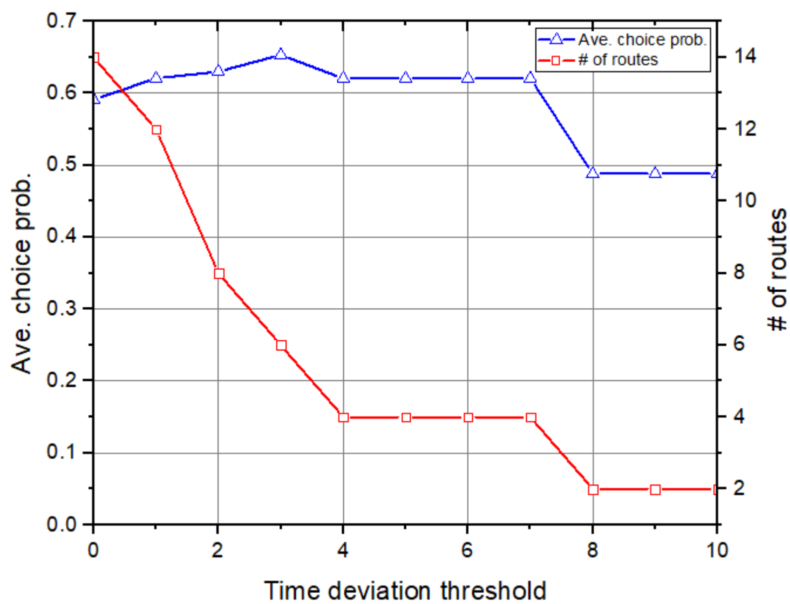
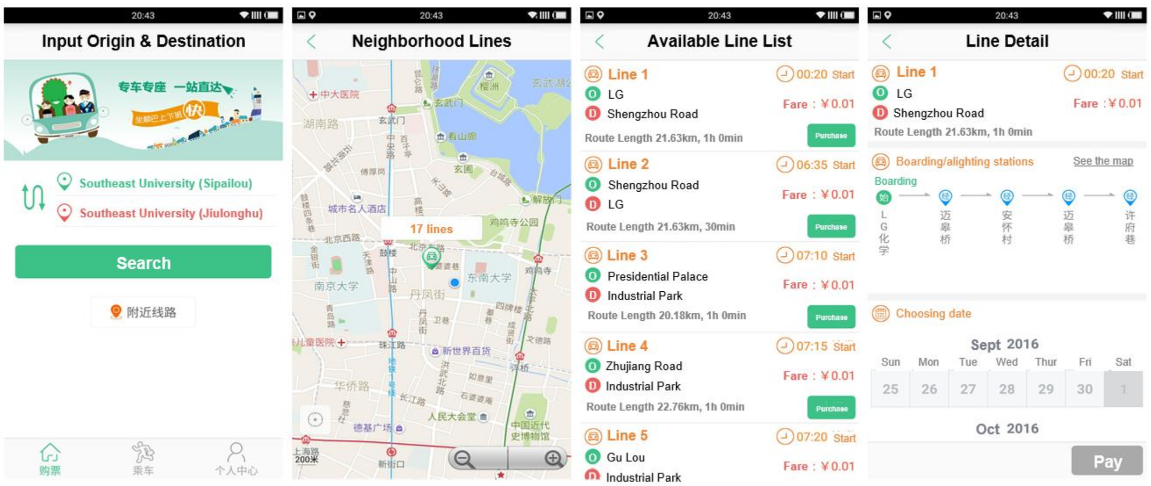


Figure 5. Impacts of time deviation on the number of CB routes and passengers' choice probabilities

5.2 Case study

5.2.1 Overview

In this section, a case study for real-world CB is presented. Empirically observed data from a CB company in Nanjing are applied. It provides both commuter and on-demand services. The commuter service is provided from suburban areas (i.e., large communities) to the central business district or industrial park during morning/evening peak hours, which is similar to the one-to-one transportation service for regular commuters with a weekly or monthly subscription. The on-demand services are designed according to the real-time demand. Fig. 6 shows the interface of the mobile application of the CB service. The connection and interactive decisions between passengers and the CB operator are completed through this on-demand service platform.



(a) (b) (c) (d)

Figure 6. The interface of the mobile application of the CB service

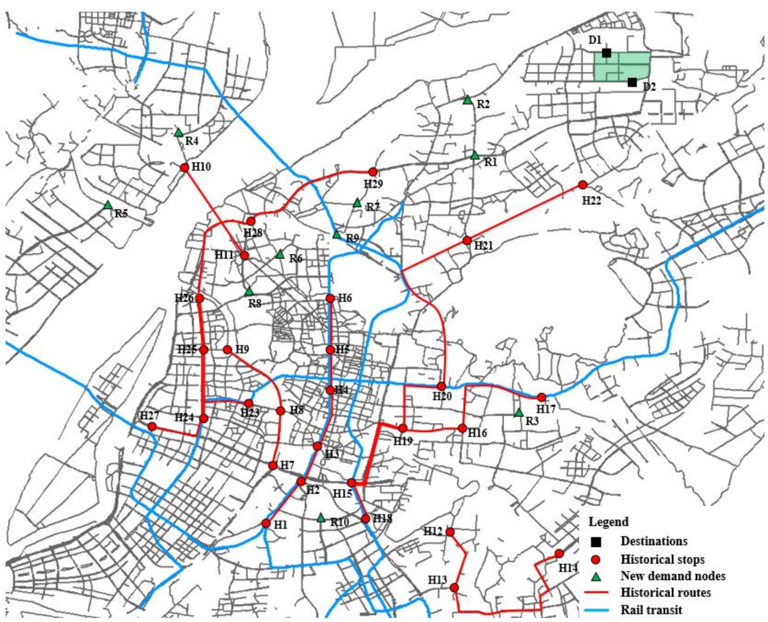


Figure 7. The distribution of historical stops and real-time requests.

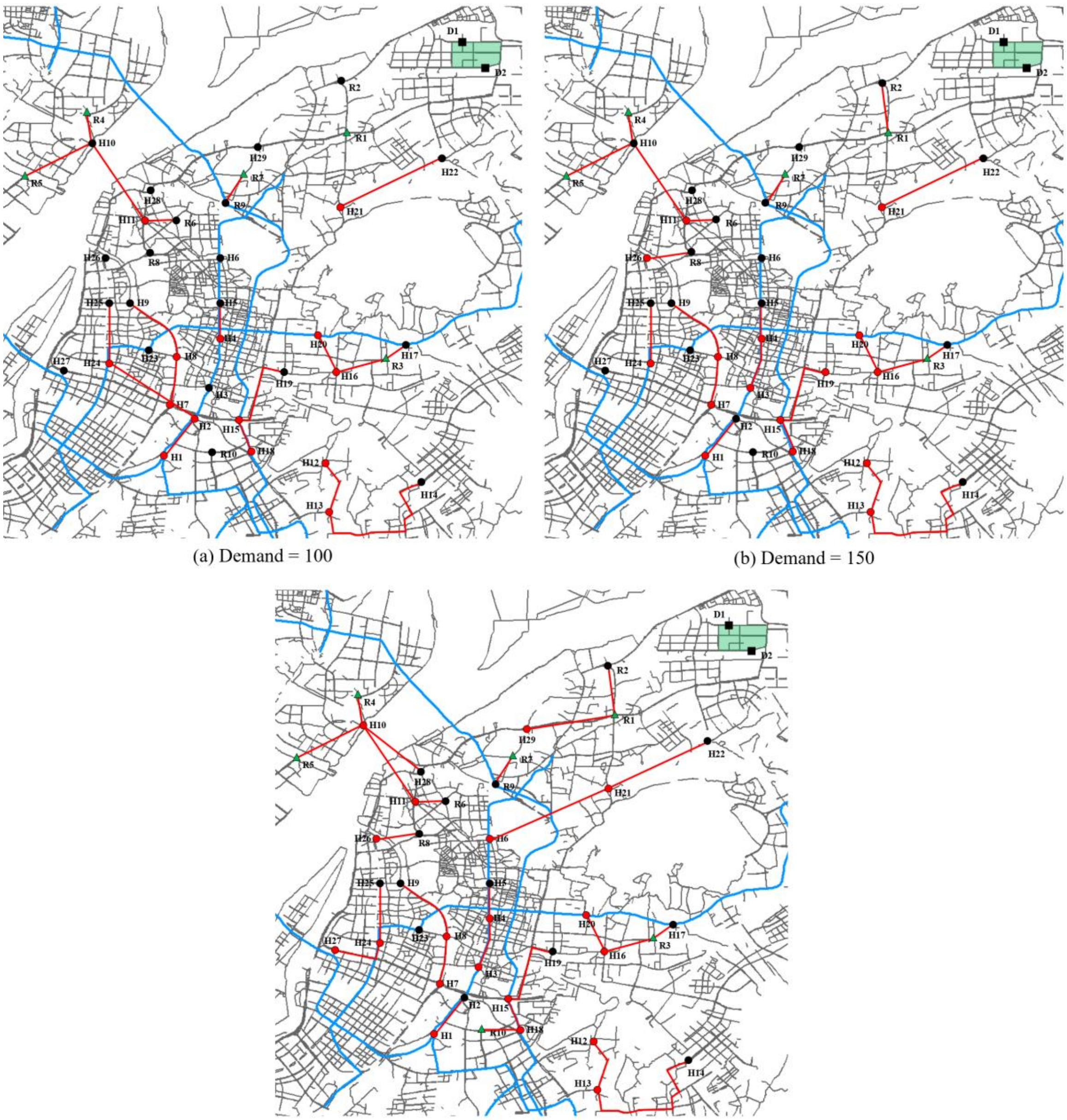
To conduct the experiments in such a real-world circumstance, eight existing commuter lines are selected as the historical routes (see Fig. 7). The destinations are located at the industrial park (the shaded area in Fig. 7) in the suburban area. In addition to regular commuters, ad hoc passengers are also allowed to make subscription through the on-line platform. We consider a fleet of minibuses with a 15-seat capacity. The total fleet size is 30. 100 home-to-work requests are generated based on the real-world data set from 6:30 AM to 9:00 AM.

5.2.2 Computational results

The computational results for this case study are reported in Table 7. In order to verify the robustness of the proposed methodology as well as the CB system, two scenarios more are presented, where additional 150 and 200 requests are randomly generated. It is shown that the acceptance rates of the Scenarios 1 and 2 are almost the same, and the acceptance of more passengers can efficiently reduce the average fare, which is in line with the "shared-mobility" of the CB system. In Scenario 1, each route is served by one vehicle, where the load factor is 11.1. Scenario 2 requires a smaller number of CB routes but more vehicles than that of Scenario 1. It can be found that some routes which only have one pickup stop are merged into other routes in Scenario 2 but served by larger fleet size (see Fig. 8, where the black dot represents the last stop of each route.). It indicates that when the demand level of an area is lower, it is more efficient to set a direct line between the origin and destination (i.e., one-to-one) rather than many-to-one or many-to-many. That is because the CB operator intends to dispatch more vehicles instead of sacrificing the passengers' waiting time. This conclusion is also confirmed in Scenario 3 when the demand level is getting large. There is no one-to-one type of line in this scenario and the passengers are served with only 12 CB routes but with a higher load factor.

Table 7 Results for the case study in Nanjing

Scenario	# of requests	Accept rate	Profit ($)	Ave. fare ($)	# of routes	# of vehicles	Load factor
1	200	0.91	10,789	40.35	18	18	11.1
2	250	0.92	11,812	38.87	15	20	12.5
3	300	0.87	13,780	34.48	12	23	13



(a) Demand = 100

(b) Demand = 150

(c) Demand = 200

Figure 8. Optimal CB service network under different demand patterns.

6 Conclusions

In this paper, an integrated decision-making framework for the demand-responsive CB system has been proposed. As an essential supplement to the multimodal transit system, the on-demand transit service provides a flexible travel pattern that considers the passenger's requests sufficiently.

To model this decision-making that involves both the CB operator and passengers, a two-phase optimization model was proposed. In the dynamic phase, the network design decisions of the CB network are determined by the operator and passenger sequentially, while the passenger occurs dynamically by specifying their desired pickup/delivery times and locations and then decides whether to take the CB service based on the operator's network design decisions. In the static phase, the CB services are re-optimized based on the confirmed demands to further optimize the service network to maximize its profit. Through the proposed framework, the passenger's mode choice activity is considered implicitly in the service network optimization process, which is usually conducted by the operator alone. During this process, the passenger's model choice behavior is described by a binary choice model, where the passenger's perceived travel cost is decided based on the operator's network design decisions.

Several potential enhancements could be considered in future works: (1) integrate the process of predicting future requests into the dynamic phase to generate more reasonable service designs; (2) introduce a nonlinear price scheme for requests with different ODs and Acknowledgments pickup/delivery time deviations; (3) take into account the role of government in CB service design and investigate the game between government, CB operator, and passengers.

Acknowledgments

This study is supported by the Youth Projects (No. 71501038), General Projects (No. 71771050) of the National Natural Science Foundation of China and the Scientific Research Foundation of Graduate School of Southeast University (No. 3221009758).

Appendix A. NP-hardness of the CB network design problem

Proposition 1. The static CB network design problem is NP-hard.

Proof. To prove that the CB network design problem is NP-hard, we first introduced other NP-hard problems, such as the traveling salesman problem with time windows (TSPTW) and the dial-a-ride problem (DARP). Baugh et al. (1998) have proved that the TSPTW can be reduced to the DARP by extending a vertex in the graph of TSPTW to a pair of pickup/delivery vertices. Compared with the DARP, a vertex in a CB system can be visited more than once by different vehicles to serve various groups of passengers with close time windows and destinations. The decision version of the CB network design problem is presented firstly. Then we further prove that an instance of the DARP for graph $G=(V,A)$ can be reduced to an instance of the CB network design problem. The decision version of the CB network design problem can be stated as follows.

Given a weighted graph $G'=(V',A')$ consisting of pickup/delivery vertices, and a depot. The passengers at each vertex of G' are divided into groups with similar pickup/delivery time windows and destinations. The decision version of the CB network design problem is whether it is possible to visit each group exactly once within its time window in k cycles at a cost not exceeding C. A feasible cycle should satisfy the following constraints defined in Section 5.1: the arrival time at a vertex is earlier than the departure time, the pickup/delivery pairs are in the same cycle, the pickup vertex is visited before the corresponding delivery vertex, and each cycle starts and ends at the depot. The decision version of the CB network design problem is shown to be NP-complete by the following statements:

i) The CB network design problem is NP.

It can be checked that each cycle is feasible, and the summation of cycle costs does not exceed C. This checking process can be done in polynomial time.

ii) The DARP can be reduced to the CB network design problem

Let graph $G=(V,A)$ be the input of the DARP, $V=\{v_0\}\cup\{v_i^p, v_i^d \mid 1\le i\le n\}$, where v_i^p and v_i^d denote the vertex at which passenger i is picked up and delivered, respectively, and v_0 denotes the depot. In the DARP, a cycle can be found with minimal weights through $2n+1$ vertices satisfying that every passenger is picked up before s/he is delivered. It is shown that the graph G can be transformed into a graph G' that can be considered as the input to the CB network design problem using the following graph constructing rules:

- Given a dial-a-ride cycle on $2n$ vertices, $V = \{v_i^p, v_i^d \mid 1 \le i \le n\}$, for every pickup vertex, add a corresponding vertex in G' and create n dummy pickup vertices, $V_P = \{p_j^i \mid 1 \le i, j \le n\}$, with the same time window; for every delivery vertex, add a corresponding vertex in G' and create n dummy delivery vertices, $V_D = \{d_j^i \mid 1 \le i, j \le n\}$, with the same window (see Fig. A1).
- Construct a CB network with vertices, $V' = \{v_0\} \bigcup V \bigcup V_P \bigcup V_D$, by pairing off p_j^i and d_j^i, and let p_j^i be a pickup vertex and d_j^i be a delivery vertex for dummy passenger j between OD pair (v_i^p, v_i^d). For all vertices v_i in V' and all j, let $d(v_0, v_i) = 0$, $d(v_i^p, p_j^i) = 0$, $d(v_i^d, d_j^i) = 0$, $d(p_j^i, d_j^i) = 0$, $d(p_j^i, v_i) = d(v_i^p, v_i)$ and $d(d_j^i, v_i) = d(v_i^d, v_i)$, where is the weight between vertices x and y.

G' could be constructed from G in polynomial time. Given an optimal solution of the CB network design problem consisting of n cycles, it is easy to see that removing the dummy pickup/delivery vertices yields a valid dial-a-ride cycle with n set to 1. Thus, if G' has a solution of the CB network design problem, then G has a DARP solution, and vice versa. The decision version of CB network design problem is NP-complete.□

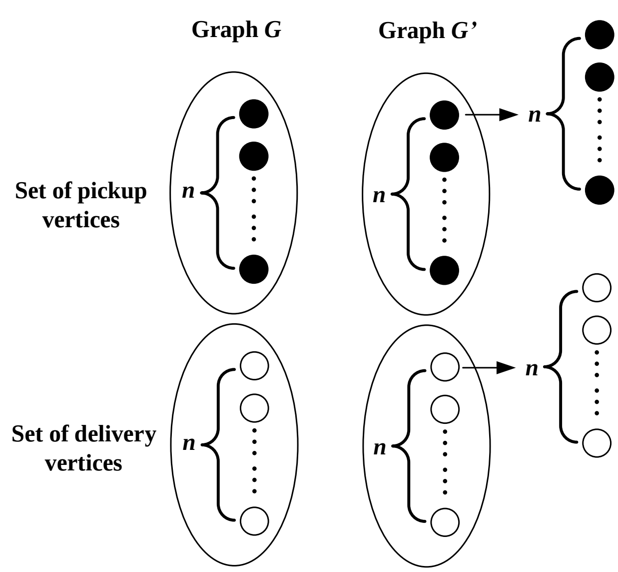


Figure A1. The construction of G' from G

Reference

Aldaihani, M. M., Quadrifoglio, L., Dessouky, M. M., & Hall, R. (2004). Network design for a grid hybrid transit service. *Transportation Research Part A, 38*(7), 511-530.

Baugh, J. W., Kakivaya, G. K. R., & Stone, J. R. (1998). Intractability of the dial-a-ride problem and a multiobjective solution using simulated annealing. *Engineering Optimization, 30*(2), 91-123.

Berbeglia, G., Cordeau, J. F., Gribkovskaia, I., & Laporte, G. (2007). Static pickup and delivery problems: a classification scheme and survey. *TOP, 15*(1), 37-40.

Berbeglia, G., Cordeau, J. F., & Laporte, G. (2010). Dynamic pickup and delivery problems. *European Journal of Operational Research, 202*(1), 8-15.

Calvo, R. W., & Colorni, A. (2007). An effective and fast heuristic for the dial-a-ride problem. *4OR, 5*(1), 61-73.

Cao, Y., & Wang, J. (2017). An optimization method of passenger assignment for customized bus. *Mathematical Problems in Engineering, 2017*, 1-9.

Chang, S. K., & Schonfeld, P. M. (1991). Optimization models for comparing conventional and subscription bus feeder services. *Transportation Science, 25*(4), 281-298.

Chebbi, O., & Chaouachi, J. (2016). Reducing the wasted transportation capacity of personal rapid transit systems: an integrated model and multi-objective optimization approach. *Transportation research part E, 89*, 236-258.

Chen, P. W., & Nie, Y. M. (2017). Connecting e-hailing to mass transit platform: analysis of relative spatial position. *Transportation Research Part C, 77*, 444-461.

Chen, X. M., Zahiri, M., & Zhang, S. (2017). Understanding ridesplitting behavior of on-demand ride services: an ensemble learning approach. *Transportation Research Part C, 76*, 51-70.

Codato, G., & Fischetti, M. (2006). Combinatorial Benders' cuts for mixed-integer linear programming. *Operations Research, 54*(4), 756-766.

Cordeau, J. F. (2006). A branch-and-cut algorithm for the dial-a-ride problem. *Operations Research, 54*(3), 573-586.

Cordeau, J. F., & Laporte, G. (2007). The dial-a-ride problem: models and algorithms. *Annals of Operations Research, 153*(1), 29-46.

Cordeau, J. F., Laporte, G., Savelsbergh, M. W. P., & Vigo, D. (2007). *Chapter 6 Vehicle*

Routing: Elsevier B.V.

Coslovich, L., Pesenti, R., & Ukovich, W. (2006). A two-phase insertion technique of unexpected customers for a dynamic dial-a-ride problem. *European Journal of Operational Research, 175*(3), 1605-1615.

Diana, M., & Dessouky, M. M. (2004). A new regret insertion heuristic for solving large-scale dial-a-ride problems with time windows. *Transportation Research Part B, 38*(6), 539-557.

Dondo, R., & Cerdá, J. (2007). A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows. *European Journal of Operational Research, 176*(3), 1478-1507.

Foth, M., Schroeter, R., & Ti, J. (2013). Opportunities of public transport experience enhancements with mobile services and urban screens. *International Journal of Ambient Computing and Intelligence, 5*(1), 1-18.

Gutiérrez-Jarpa, G., Desaulniers, G., Laporte, G., & Marianov, V. (2010). A branch-and-price algorithm for the vehicle routing problem with deliveries, selective pickups and time windows. *European Journal of Operational Research, 206*(2), 341-349.

Guo, R., W., & Zhang, W. (2018). Route design problem of customized buses: mixed iInteger programming model and case study. *Journal of Transportation Engineering Part A, 144*(11), 1-14.

Ho, S. C., Szeto, W. Y., Kuo, Y. H., Leung, J. M., Petering, M., & Tou, T. W. (2018). A survey of dial-a-ride problems: literature review and recent developments. *Transportation Research Part B, 111*, 395-421.

Horn, M. E. (2002). Fleet scheduling and dispatching for demand-responsive passenger services. *Transportation Research Part C, 10*(1), 35-63.

Huang, D., Liu, Z., Liu, P., & Chen, J. (2016). Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure. *Transportation Research Part E, 96*, 1-19.

Ilani, H., Shufan, E., Grinshpoun, T., Belulu, A., & Fainberg, A. (2014). A reduction approach to the two-campus transport problem. *Journal of Scheduling, 17*(6), 587-599.

Kamga, C., Yazici, M. A., & Singhal, A. (2013). Implementation of interactive transit

information kiosks at New York City transit facilities: analysis of user utilization and lessons learned. *Transportation Research Part C, 35*, 218-231.
KFH Group (2008). *Guidebook for measuring, assessing, and improving performance of demand-response transportation* (Report No. 124). Washington DC: Transportation Research Board.
Laporte, G. (2009). Fifty years of vehicle routing. *Transportation Science, 43*(4), 408-416.
Liu, T., & Ceder, A. (2015). Analysis of a new public-transport-service concept: Customized bus in China. *Transport Policy, 39*, 63-76.
Liu, T., Ceder, A., Bologna, R., & Cabantous, B. (2016). Commuting by customized bus: A comparative analysis with private car and conventional public transport in two cities. *Journal of Public Transportation, 19*(2), 55-74.
Lyu, Y., Chow, C. Y., Lee, V. C., Ng, J. K., Li, Y., & Zeng, J. (2019). CB-Planner: A bus line planning framework for customized bus systems. *Transportation Research Part C, 101*, 233-253.
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*, 1-14.
Mageean, J., & Nelson, J. D. (2003). The evaluation of demand responsive transport services in Europe. *Journal of Transport Geography, 11*(4), 255-270.
Meng, Q., & Liu, Z. (2011). Trial-and-error method for congestion pricing scheme under side-constrained probit-based stochastic user equilibrium conditions. *Transportation, 38*(5), 819-843.
Nair, R., & Miller-Hooks, E. (2014). Equilibrium network design of shared-vehicle systems. *European Journal of Operational Research, 235*(1), 47-61.
Pavone, M., Frazzoli, E., & Bullo, F. (2010). Adaptive and distributed algorithms for vehicle routing in a stochastic and dynamic environment. *IEEE Transactions on automatic control, 56*(6), 1259-1274.
Potts, J. F., Marshall, M. A., Crockett, E. C., & Washington, J. (2010). *A Guide for Planning and Operating Glexible Public Transportation Services* (TCRP Report 140). Washington, D.C.: Transportation Research Board.
Qiu, G., Song, R., He, S., Xu, W., & Jiang, M. (2018). Clustering passenger trip data for

the potential passenger investigation and line design of customized commuter bus. *IEEE Transactions on Intelligent Transportation Systems, 2018*, 1-10.

Qiu, X., Feuerriegel, S., & Neumann, D. (2017). Making the most of fleets: A profit-maximizing multi-vehicle pickup and delivery selection problem. *European Journal of Operational Research, 259*(1), 155-168.

Ren, Y., Chen, G., Han, Y., & Zheng, H. (2016). Extracting potential bus lines of customized city bus service based on public transport big data. *IOP Conference Series: Earth and Environmental Science, 46*(2016), 1-8.

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

Ronald, N., Thompson, R., & Winter, S. (2015). Simulating demand-responsive transportation: a review of agent-based approaches. *Transport Reviews, 35*(4), 404-421.

Sheffi, Y. (1985). *Urban Transportation Networks*: Prentice-Hall, Englewood Cliffs.

Stelzer, A., Englert, F., Hörold, S., & Mayas, C. (2016). Improving service quality in public transportation systems using automated customer feedback. *Transportation Research Part E, 89*, 259-271.

Suhl, L., Biederbick, C., & Kliewer, N. (2001). Design of customer-oriented dispatching support for railways. In: Voss., Daduna (Eds.), *Computer-Aided Scheduling of Public Transport,* LNES, vol 505. Springer, Berlin, Heidelberg.

Tong, L. C., 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, 85*, 451-475.

van Engelen, M., Cats, O., Post, H., & Aardal, K. (2018). Enhancing flexible transport services with demand-anticipatory insertion heuristics. *Transportation Research Part E, 110*, 110-121.

Yu, Y., Machemehl, R. B., & Xie, C. (2015). Demand-responsive transit circulator service network design. *Transportation Research Part E, 76*, 160-175.

Zhang, J., Wang, D. Z., & Meng, M. (2017). Analyzing customized bus service on a multimodal travel corridor: an analytical modeling approach. *Journal of*

Transportation Engineering Part A, 143(11), 1-12.