

Design of Multimodal Transit Services with Stochastic Demand

Khaled Saeed¹ and Fumitaka Kurauchi^{2*}

¹*Graduate School of Civil Engineering, Gifu University, Gifu, Japan*

²*Department of Civil Engineering, Gifu University, Gifu, Japan*

¹*eng_kh.saeed@yahoo.com, ^{2*}kurauchi@gifu-u.ac.jp*

Abstract

The main challenges faced by public transport planners are low passenger loads and variations in demand. Local authorities aim to provide sustainable public transit systems that can accommodate variations in demand. Our aim in this study was to design a multimodal service that combines fixed-route services with flexible transit systems under stochastic demand. In previous work, the design of only one type of service was considered, while the design of other services was not considered or approximated. We took the perspectives of both operators and users into account when designing two services accurately in our model. Our results show that the total costs can be reduced for both the operator and passengers by implementing a multimodal service. We also highlight the effect of considering the accurate design of both services on the optimal combination and associated total costs.

Keywords: *Bus Service, Dial-A-Ride, Multimodal, Stochastic Demand, Sustainable, Transit Network Design*

1. Introduction

Public transit is widely recognized as a key factor in the economic growth and sustainability of a city. The efficiency of public transit systems depends on many factors, including the level of demand, and system planning and operation. Traditional public transit services tend to be more convenient in densely populated areas where demand is high. However, these services are not cost-effective when passenger loads are low, such as in rural areas and late at night. Many local authorities are required to subsidize these types of services to keep them running.

Public authorities consistently improve the quality of public transit services to encourage people to use them. These efforts focus on improving many parameters that affect the efficiency of public transport networks, including the service coverage, frequency, quality, and comfort, and seat availability. The design of such public transportation networks is called the transit network design problem (TNDP). The goal of the TNDP is to define transit line layouts and associated operational characteristics that optimize specific objective functions, such as minimizing the weighted sum of operators' and users' costs [1]. Guihaire and Hao [2] stated that from a users' perspective, a transit network should cover a large area, be highly accessible, offer numerous direct destinations and rarely deviate from the shortest path. Meanwhile, from the operator's perspective, costs should be minimized by offering the lowest number of routes or the shortest total route length.

The TNDP is considered a complex variant of a non-deterministic polynomial-time hard (NP-hard) problem. Many researchers have addressed the TNDP and several approaches have been considered; [1, 3, 4] are recent detailed reviews of this problem. Shimamoto *et al.*, [5] considered the transit fare problem to be a bi-level problem, where the operator has the option to increase fares on crowded lines. Their numerical results show that this approach alleviates the congestion-mitigation

effect; however, some origin/destination (OD) pairs may suffer from heavier congestion. Fan and Machemehl [6] approached the spatial equity issue by applying a bi-level model to the transit network redesign problem. The upper-level of their model minimizes the operator cost, user cost, and unsatisfied demand, while the lower-level addresses the transit assignment problem. Their results show that total travel costs increase when network designs take spatial equity into account. Szeto and Jiang [7] proposed a bi-level model where the upper-level objective was to minimize the number of transfers and the lower-level solved the transit assignment problem with capacity constraints. Waiting times were incorporated into the model of users' behavior, but were not included in the planner's objective function. Guan *et al.*, [8] proposed a binary linear integer program to optimize transit line layouts and passenger line assignments simultaneously. However, they did not determine the optimal frequencies and travelers' waiting times were not considered.

Public transport services are usually fixed-route services; *i.e.*, they operate according to a fixed schedule along a fixed route with predetermined stops. However, passengers on different schedules often wish to be transported to different locations, and transport providers desire to increase their profits. Hence, more flexible transport services are needed. Flexible transit services include a wide range of hybrid services that do not have fully fixed routes and can respond to demand. Demand-responsive services (also known as "dial-a-ride", DAR) are a form of flexible transit service consisting of passenger cars (*e.g.*, vans or small buses) that go to certain locations to pick up or drop off passengers in response to requests [9]. According to Kurauchi *et al.*, [10] it is difficult to provide these services for the same fare as that of a normal bus journey, so the providers of these services must charge higher fares to be profitable. DAR vehicles do not run according to fixed routes or timetables. Instead, they follow itineraries calculated by routing software. The problem of determining an optimal itinerary is called the DAR problem (DARP). The DARP involves designing vehicle routes and schedules to collect a certain number of users and take them to specified points within a given time window (TW). The complexity of this problem increases when users' desires are considered, as these introduce additional operational constraints, such as the pickup/delivery time and maximum ride time [11].

As the DARP is a complex NP-hard problem, it is difficult to find the optimal solution. Therefore, most researchers develop solutions based on heuristics and metaheuristics [12, 13, 14, 15]. Due to its complexity, few studies have provided exact algorithms for solving the DARP. Cordeau [16] proposed a three-indexed formulation for the DARP and developed a branch-and-cut algorithm to find the optimal solution. Using their numerical approach, they were able to find optimal solutions for up to 36 requests. Hu and Chang [17] applied the branch-and-price approach to solve the DARP. They showed that the length of the TW has a significant effect on the objective value and computation time. Saeed and Kurauchi [11] used a branch-and-cut algorithm to solve the DARP for up to 65 requests within 2 hours. Their objective function included constraints imposed by considering both the operator and the passengers. They concluded that the size of the TW and the maximum allowable number of rides imposes a trade-off between the operator's desires and users' needs.

Fixed-route services have high construction and operating costs, so they are not cost-effective when passenger loads are low. DAR services are more flexible, and are cost-effective when demand is low. However, fixed-route services have a lower unit cost per passenger than DAR services. From this point of view, by combining fixed routes with DAR services, the advantages of both services can be enjoyed to serve passengers at a reasonable cost.

Few research papers have considered the possibility of combining fixed-route and flexible transit services, and most that have considered this approach have presented operational planning models. These models have considered the design of the flexible service, but not considered the design of the fixed-route service. Examples of these models can be found in [18, 19, 20]. They assume that the design parameters of fixed services, such as line alignments, vehicle capacities and line frequencies or timetables, are fixed and then calculate the design parameters of the flexible service. Häll *et al.*, [21] introduced a formulation for an integrated DARP. The aim of their work was to design vehicle routes and schedules for DAR requests where part of each request may be fulfilled by a fixed-route service. Unfortunately, their model was able to solve only small-scale cases optimally and they did not consider the design of fixed-route services. Another approach to combining fixed-route and flexible transit services is to consider only the design of the fixed-route service. Examples of this approach can be found in studies by An and Lo [22, 23]. They investigated the optimal way to combine regular services operated on fixed routes and schedules with flexible services under stochastic demand. However, they did not allow passengers to transfer between the two modes, and their model did not consider the design of the flexible service. Instead, they calculated the approximate cost of the flexible service.

In summary, this review shows that combination of fixed and flexible transport services remains an ongoing topic of research. There is a range of approaches to combine the two services, and several models based on differing assumptions have been proposed. However, little attention has been paid to designing the two services as one multimodal service. In this paper, we reconcile different objectives of users and operators in one multimodal service. Our aim is to determine the amount of flexible service needed to complement a fixed service and minimize construction costs, operating costs and passengers' costs. In addition, accurate design of both fixed and flexible services that serve all OD pairs under stochastic demand is considered.

The outline of this paper is as follows: In Section 2 we provide the mathematical formulation of the model, define the key variables and describe the networks. The procedure for solving the model is described in Section 3. In Section 4, we present our numerical results and describe the advantages conferred by using a multimodal service. Finally, we conclude our work in Section 5 and describe our plans for future work.

2. Model Formulation

2.1. Networks Description

2.1.1. Bus Service Network: We define the bus service network as a graph $G(N, A)$, where N is the set of nodes and A is the set of arcs. Each arc $(i, j) \in A$ connects node i to node j , where $i, j \in N$, $i \neq j$ and the link distance is l_{ij} . Each physical bus station $i \in N$ contains one replicated node per bus line, so it consists of L replicated nodes as in Häll *et al.*, [21]. This enables us to represent the boarding and alighting demand for every bus line, and the transfer demand between different bus lines. Each replicated node is denoted ia , where i represents the original station and a represents the bus line associated with this replicated node. Note that every replicated node is connected to only one bus line. As shown in Figure 1, node $2a$ represents the replicated node for line a at bus station 2, while node $3b$ represents the replicated node for line b at bus station 3. Each bus line $l \in L$ has a specific frequency $f_l \in f$ and the route frequency is constrained by lower and upper bounds, f_{min} and f_{max} , respectively. The maximum capacity of a bus is Q_b . We introduce a set of dummy start nodes $D_s = \{D_{sl}, l \in L\}$ and a set of dummy end nodes $D_e = \{D_{el}, l \in L\}$.

$L\}$ so that every bus line l has a fixed dummy start (end) node. This enables us to select the start and end nodes of each bus line flexibly. We assume that each route serves both directions with the same line frequency. We define c_{ij}^1 and c_{ij}^2 as the bus operation and construction costs for link ij , respectively. c_{ij}^3 is the passengers' travel cost for link ij . c^4 and c^5 are, respectively, the passenger waiting time and transfer time costs.

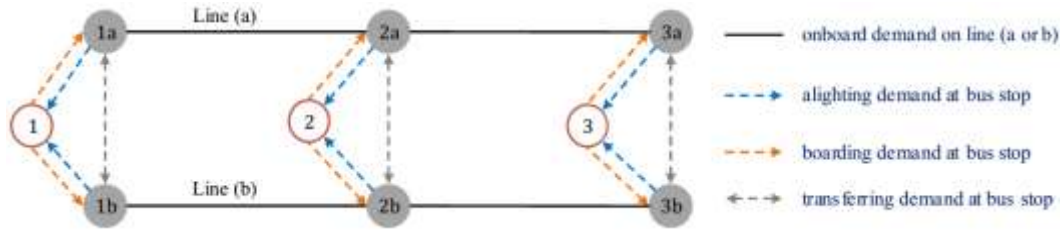


Figure 1. Example of Two Bus Lines with Three Bus Stops

2.1.2. DAR Service Network: We define the DAR service network on a directed graph $G(\bar{N}, \bar{A})$, where \bar{N} is the set of nodes and \bar{A} is the set of arcs. Each arc $(i, j) \in \bar{A}$ is associated with a DAR vehicle operating cost c_{ij}^6 and travel time t_{ij} . The set of nodes, \bar{N} , is divided into four subsets: pickup nodes ($P = \{i, \dots, R\}$), delivery nodes ($D = \{R+i, \dots, 2R\}$), the origin depot node ($D_o = \{0\}$), and the destination depot node ($D_d = \{2R+1\}$). We define R as the set of requests and K as the set of DAR vehicles. Every DAR vehicle has maximum capacity Q_k and maximum route duration time T_k . Every request is associated with two nodes $(i, R+i)$, where $i \in P$ and $R+i \in D$ and has load q_r . Each node $i \in \bar{N}$ is associated with a time window $[e_i, l_i]$, where e_i and l_i represents the earliest and latest time at which the service may begin at node i , respectively. Also, a service duration d_i for loading/unloading passengers is associated with each node $i \in \bar{N}$. The passenger load q_i , is positive at pickup nodes and negative at delivery nodes. The binary three-index variable x_{ij}^k is equal to 1 if arc (i, j) is traversed by vehicle k . We assume that the DAR services will not have big influence on the roadway congestion as they are supplementary services.

2.2. Variable Definitions

In this paper, we aim to minimize the total cost of the multimodal service while serving all OD pairs under stochastic demand. This cost includes the bus costs, DAR costs and passengers' cost. To address this problem, we determine the variables that are relevant to the bus and DAR services. In the case of the bus service, we consider: bus arc flows (\mathbf{Y}), which specify the bus line alignments; passenger arc flows (\mathbf{Z}), which represent the on-board passenger demand on every bus line; passenger arc flows (\mathbf{U}), which represent the passenger demand for boarding (alighting) on (from) different bus lines; passenger arc flows (\mathbf{T}), which represent the passenger demand for transferring between different bus lines; and frequencies (\mathbf{F}), which represent the frequency of each bus line. The DAR service is associated with the following variables: DAR arc flows (\mathbf{X}), which specify the DAR vehicle itineraries; DAR vehicle capacities (\mathbf{Q}), which represent the number of passengers in the DAR vehicle after it leaves a node; and DAR service times (\mathbf{B}), which correspond to when the DAR vehicles begin their service at any node, whether loading at a pick-up node or unloading at a delivery node. Our model can be applied to networks with complex topologies. Passengers can transfer between different bus lines, but they cannot transfer between the bus and DAR services.

2.2.1. Bus Service Network: This model uses the following notation:

N^* Total number of nodes on one bus line, including dummy nodes;

s^v	Service ratio for OD pair v ;
\bar{B}^v	Amount of random demand carried by the bus service for OD pair v ;
Ψ_v^{-1}	Cumulative distribution function of the random demand for OD pair v ;
$w_{il,jl}^l$	Continuous variable representing the link frequency for link il, jl on bus line l ;
λ_i^l	Binary variable indicating the sequence of station i on bus line l ;
f_l	Continuous variable representing the frequency of bus line l ;
$Y_{il,jl}^l$	Binary variable indicating whether link il, jl is on bus line l ;
Z_{imjm}^v	Continuous variable representing the passenger flow from replicated node im to replicated node jm on line m for OD pair v ;
$U_{i,im}^v$	Continuous variable representing the boarding passenger flow inside station i between the origin station node i and the replicated node im on line m for OD pair v .

2.2.2. DAR Service Network: This model uses the following notation:

T^r	Request travel time;
M	Large positive number;
B_i^k	Continuous variable representing the time at which vehicle k starts its service at node i ;
Q_i^k	Continuous variable representing the number of passengers on vehicle k after leaving node i ;
x_{ij}^k	Binary variable denoting whether vehicle k travels from node i to node j ;
E^v	Continuous variable representing the demand carried by the DAR service for OD pair v .

2.3. Mathematical Formulation

The objective of this formulation is to find the optimal combination (service ratio) of bus and DAR services, such that total costs are minimized and all the OD pairs are served. The service ratio (\mathbf{S}) is defined as the ratio of demand carried by the bus service. It is a vector containing the service ratio of each OD pair, and different OD pairs may have different service ratios. We start by fixing the values of \mathbf{S} for the multimodal service. This enables us to design the bus service easily, as the demand is known. We design the DAR service to meet the excess demand when the bus service is operating at full capacity.

2.3.1. Bus Service Design

$$\begin{aligned} \text{Min}_{f,Y,Z,U,T} \quad & \sum_{l \in L} \sum_{ij \in A} c_{ij}^1 f_l Y_{il,jl}^l + \sum_{l \in L} \sum_{ij \in A} c_{ij}^2 Y_{il,jl}^l + \sum_{v \in V} \sum_{ij \in A} c_{ij}^3 Z_{il,jl}^v \\ & + c^4 \sum_{v \in V} \sum_{l \in L} \frac{60}{2f_l} U_{i,il}^v + c^5 \sum_{v \in V} \sum_{l \in L} (\theta + \frac{60}{2f_l}) T_{in,il}^v \end{aligned} \quad (1)$$

$$\sum_{l \in L} \sum_{j \in N} Y_{il,jl}^l = 1 \quad \forall i \in D_s \quad (2)$$

$$\sum_{j \in N} Y_{il,jl}^l \leq 1 \quad \forall i \in N, l \in L \quad (3)$$

$$\sum_{j \in N} Y_{jl,il}^l \leq 1 \quad \forall i \in N, l \in L \quad (4)$$

$$\sum_{l \in L} \sum_{j \in N} Y_{jl,il}^l = 1 \quad \forall i \in D_e \quad (5)$$

$$\sum_{i \in N} Y_{il,gl}^l - \sum_{j \in N} Y_{gl,jl}^l = 0 \quad \forall l \in L, g \in N \quad (6)$$

$$\sum_{l \in L} Y_{il,jl}^l + Y_{jl,il}^l \leq 1 \quad \forall ij \in A \quad (7)$$

$$\sum_{l \in L} U_{g,gl}^v - \sum_{l \in L} U_{gl,g}^v = \begin{cases} \bar{B}^v, & \text{if } (g) \text{ is the origin of OD pair } v \\ -\bar{B}^v, & \text{if } (g) \text{ is the origin of OD pair } v \\ 0, & \text{otherwise} \end{cases} \quad \forall g \in N, v \in V \quad (8)$$

$$Z_{il,jl}^v + U_{il,i}^v + \sum_{n \in L} T_{il,in}^v - Z_{gl,il}^v - U_{i,il}^v - \sum_{n \in L} T_{in,il}^v = 0 \quad \forall i \in N, l \in L, v \in V \quad (9)$$

$$\sum_{v \in V} Z_{il,jl}^v \leq Q_b (Y_{il,jl}^l + Y_{jl,il}^l) f_l \quad \forall ij \in A, l \in L \quad (10)$$

$$w_{il,jl}^l - f_l \leq 0 \quad \forall ij \in A, l \in L \quad (11)$$

$$w_{il,jl}^l - M Y_{il,jl}^l \leq 0 \quad \forall ij \in A, l \in L \quad (12)$$

$$M(Y_{il,jl}^l - 1) - w_{il,jl}^l + f_l \leq 0 \quad \forall ij \in A, l \in L \quad (13)$$

$$f_{\min} \leq f_l \leq f_{\max} \quad \forall l \in L \quad (14)$$

$$\lambda_{il}^l - \lambda_{jl}^l + |N^*| \times Y_{il,jl}^l \leq |N^*| - 1 \quad \forall ij \in A, l \in L \quad (15)$$

$$w_{il,jl}^l = f_l Y_{il,jl}^l \quad \forall ij \in A, l \in L \quad (16)$$

$$Y_{il,jl}^l = 0 \text{ or } 1 \quad \forall l \in L, ij \in A \quad (17)$$

$$\bar{B}^v = \Psi_v^{-1}(s^v) \quad \forall v \in V \quad (18)$$

The objective function (1) minimizes the sum of its components: the operating costs, construction costs, and passengers' in-vehicle travel time costs, pre-boarding waiting time costs and transfer time costs. The frequency of the bus lines determines how long passengers must wait at bus stops, so the waiting time can be set to $(0.5 \times 60 / \text{frequency})$ [24]. The transfer penalty is denoted θ . Constraint (2) ensures that each line is connected to only one start dummy node. We ensure that station i is connected to one upstream station and one downstream station per bus line by setting constraints (3) and (4), respectively. Constraint (5) states that each line is connected to only one end dummy node. Constraint (6) rules out connectivity of each station on the bus line l to be connected by two links on the same line l . Constraint (7) ensures that every link is occupied by at most one bus line. Constraint (8) represents the passenger flow balancing condition on the original station nodes. This captures the passenger demand for boarding (alighting) different bus lines. Constraint (9) represents the passenger flow balancing condition on every replicated node il . This captures the passenger demand for transferring from (to) line l to (from) another bus line. We guarantee that there is sufficient capacity to meet the passenger demand to travel from station i to station j on different bus lines by setting Constraint (10). The constraints (11)–(13) are a linear relaxation of the nonlinear constraint (16). This is shown here as an illustration and will be removed from the calculation process later. The boundaries of the bus line frequencies are set by constraint (14). Constraint (15) is the sub-tour elimination constraint. Constraint (18) calculates how much of the demand is met by the bus service for each OD pair.

2.3.2. DAR Service Design

$$\text{Min}_{X, B, Q} \sum_{k \in K} \sum_{i \in N} \sum_{j \in N} c_{ij}^k x_{ij}^k + c^7 \sum_{r \in R} q_r T^r \quad (19)$$

$$\sum_{k \in K} \sum_{j \in N} x_{ij}^k = 1 \quad \forall i \in P \cup D \quad (20)$$

$$\sum_{j \in N} x_{ij}^k - \sum_{j \in N} x_{R+i,j}^k = 0 \quad \forall i \in P, k \in K \quad (21)$$

$$\sum_{j \in N} x_{0j}^k = 1 \quad \forall k \in K \quad (22)$$

$$\sum_{j \in N} x_{i,2R+1}^k = 1 \quad \forall k \in K \quad (23)$$

$$\sum_{j \in N} x_{ji}^k - \sum_{j \in N} x_{ij}^k = 0 \quad \forall i \in P \cup D, k \in K \quad (24)$$

$$Q_j^k \geq (Q_i^k + q_j) - M(1 - x_{ij}^k) \quad \forall i \in \bar{N}, j \in \bar{N}, k \in K \quad (25)$$

$$\max\{0, q_i\} \leq Q_i^k \leq \min\{Q_k, Q_k + q_i\} \quad \forall i \in \bar{N}, k \in K \quad (26)$$

$$B_j^k \geq B_i^k + d_i + t_{ij} - M(1 - x_{ij}^k) \quad \forall i \in \bar{N}, j \in \bar{N}, k \in K \quad (27)$$

$$B_{2R+1}^k - B_0^k \leq T_k \quad \forall k \in K \quad (28)$$

$$t_{i,R+i} \leq B_{R+i}^k - (B_i^k + d_i) \leq h_r(t_{i,R+i}) \quad \forall r \in R, i \in P, k \in K \quad (29)$$

$$e_i \leq B_i^k \leq l_i \quad \forall i \in \bar{N}, k \in K \quad (30)$$

$$x_{ij}^k \in \{0,1\} \quad \forall i \in \bar{N}, j \in \bar{N}, k \in K \quad (31)$$

The objective function (19) minimizes the total cost incurred by operating the DAR vehicles and the passengers' travel costs, where c^7 is the passenger travel cost per unit time. Constraint (20) ensures that each request is only served once. Constraint (21) ensures that the pickup and delivery nodes for each request are served by the same vehicle. Constraints (22) and (23) guarantee that the route for each vehicle k will start from the origin depot and end at the destination depot. Constraint (24) ensures that the DAR vehicle leaves each node after arriving. Constraint (25) is the capacity criterion for accepting a particular request. Constraint (26) ensures that the maximum vehicle capacity is not exceeded. Constraint (27) is the time condition to accept a request. Constraint (28) ensures that the duration of each vehicle itinerary is less than the maximum route duration time for the DAR vehicle. Constraint (29) limits the allowable duration of each request. The parameter h_r allows us to control the upper limit of the ride time of a particular request. Constraint (30) ensures that all nodes are serviced within their TW. Constraints (27) and (30) also guarantee that no sub-tours exist.

2.3.3. Multimodal Service Design

The objective is to minimize the total cost of the multimodal transport service. The total cost includes the operating and construction costs and the passengers' in-vehicle travel time, waiting time and transfer time costs. We assume that the operator chooses the service to provide, bus or DAR, and that passengers cannot choose which mode to use (*i.e.*, system optimal). However, the operator's decision will consider passengers' perspective in terms of minimizing passengers' in-vehicle travel times, waiting times, and transfer times. The multimodal service is designed in two stages; in the first, we design the bus service; in the second, we calculate the passengers' costs and the demand to be met by the DAR service, and then design the DAR service. The details of the two stages are described as follows:

- First stage: We first fix the value of \mathbf{S} so that we can use Equation (18) to determine how much demand the bus service meets for each OD pair. Once the demand is fixed, we can design the bus service and calculate the bus lines' alignment (\mathbf{Y}), associated frequencies (\mathbf{F}) and bus service costs. We assume that the probability distribution of

the stochastic demand is known. The terms for the travel time, waiting time and transfer time costs are removed from Equation (1) as these will be calculated accurately in the next stage. The costs for the bus service can be calculated from Equation (32), subject to (2)–(17).

$$\underset{\mathbf{F}, \mathbf{Y}}{\text{Min}} \omega_1(S) = \sum_{l \in L} \sum_{ij \in A} c_{ij}^1 f_l Y_{il,jl}^l + \sum_{l \in L} \sum_{ij \in A} c_{ij}^2 Y_{il,jl}^l \quad (32)$$

- Second stage: We aim to determine how much demand the DAR will meet when the realized demand exceeds the capacity of the bus service, and the cost of meeting this demand. We express the possible stochastic demand situations using a scenario simulation method. This enables us to calculate how much demand the DAR service will meet and the expected associated costs by generating many possible scenarios from the continuous probability distribution of the stochastic demand. Assume that H scenarios are generated in total and each $h \in H$ is generated with probability $p_h=1/m$, so $\sum_{h \in H} P_h=1$. For each generated scenario $h \in H$, if the realized demand exceeds the capacity of the bus service, the excess demand is met by the DAR service, which serves the people who cannot be served by the bus.

Phase (1): We use Equations (33)–(36) to calculate how much demand the DAR service can meet. The general idea is to recheck whether we can meet some of the excess demand on the bus service by reassigning \mathbf{Z} . The cost of providing the bus service does not increase, as the operating and construction costs for the bus service are fixed during the first stage.

$$\underset{\mathbf{Z}, \mathbf{U}, \mathbf{T}, \mathbf{E}}{\text{Min}} \xi(S) = \sum_{v \in V} \sum_{ij \in A} c_{ij}^3 Z_{il,jl}^v + c^4 \sum_{v \in V} \sum_{l \in L} \frac{60}{2f_l} U_{i,il}^v + c^5 \sum_{v \in V} \sum_{l \in L} (\theta + \frac{60}{2f_l}) T_{in,il}^v + \sum_{v \in V} c^8 E_h^v \quad \forall h \in H \quad (33)$$

$$\sum_{l \in L} U_{g,gl}^v - \sum_{l \in L} U_{gl,g}^v = \begin{cases} \bar{B}^v - E_h^v, (g) \text{ is origin of OD pair } v \\ E_h^v - \bar{B}^v, (g) \text{ is origin of OD pair } v \\ 0, \text{ otherwise} \end{cases} \quad \forall g \in N, v \in V \quad (34)$$

$$Z_{il,jl}^v + U_{il,i}^v + \sum_{n \in L} T_{il,in}^v - Z_{gl,il}^v - U_{i,il}^v - \sum_{n \in L} T_{in,il}^v = 0 \quad \forall i \in N, l \in L, v \in V \quad (35)$$

$$\sum_{v \in V} Z_{il,jl}^v \leq Q_b (Y_{il,jl}^l + Y_{jl,il}^l) f_l \quad \forall ij \in A, l \in L \quad (36)$$

The objective function (33) minimizes the sum of the passengers' costs and the amount of demand met by the DAR service, where c^8 is the approximate cost of using the DAR service. Constraint (34) guarantees that when the demand exceeds the capacity of the bus service, the DAR service will meet this demand. Constraint (35) represents the passenger flow balancing condition on every replicated node. As \mathbf{Y} and \mathbf{F} are fixed in the first stage, Constraint (36) imposes an upper limit on the passenger flow for the bus service. It is easy to calculate the sum of the passengers related costs from Equation (37).

$$\omega_2(S) = \xi(S) - \sum_{v \in V} c^8 E_h^v \quad \forall h \in H \quad (37)$$

Phase (2): The DAR service can now be designed using Equations (19)–(31) and (38). Equation (38) ensures that, for every scenario $h \in H$ generated, the total demand for all requests is equal to the level of demand met by the DAR.

$$\sum_{v \in V} E_h^v = \sum_{r \in R} q_r, \forall h \in H \quad (38)$$

Equation (39) gives the expected cost of the DAR:

$$\phi(S) = \sum_{h \in H} p_h \sum_{k \in K} \sum_{i \in N} \sum_{j \in N} c_{ij}^k x_{ij}^k + c^7 \sum_{r \in R} q_r T^r \quad (39)$$

The overall cost can be calculated as follows:

$$\underset{\substack{f, Y, Z, U, T, \\ X, B, Q, E, S}}{\text{Min}} \Phi(S_n) = \omega_1(S) + \omega_2(S) + \phi(S) \quad (40)$$

We solve the problem using the SR-based gradient solution procedure presented in Lo *et al.*, [25]. This solution procedure determines the descent direction in each step. Hence, the total cost decreases gradually until a locally optimal solution is obtained. The locally optimal solution is defined as the zero cost derivative with respect to \mathbf{S} . We have improved the efficiency of this procedure, and will describe the method for developing the updated gradient solution over \mathbf{S} in the next section, where we solve stages (1, 2) repeatedly until a stopping criterion is fulfilled.

3. Solution Procedure

The SR-based gradient solution procedure determines the descent direction based on \mathbf{S} . As stated earlier, once \mathbf{Y} is fixed, the minimum cost can be calculated. Due to the discrete nature of \mathbf{Y} , it is not possible to conduct a line search directly over \mathbf{Y} by considering the effect of all of its elements at the same time. Instead, we use \mathbf{S} to determine the gradient descent direction that best minimizes the objective function. In the following steps, we conduct a line search to optimize \mathbf{S} , which is a vector with different elements for different OD pairs. The iteration number is denoted by the subscript n .

1. Set $n = 1$, initialize $\mathcal{S}_n = (s^1, s^2, \dots, s^t)$, (t) is the number of OD pairs.
2. In stage (1), design a bus service for a fixed demand, which corresponds to fixed \mathbf{S} , and calculate the corresponding values of \mathbf{Y} and \mathbf{F} .
3. Solve stage (2) for all the generated scenarios $h \in H$ and calculate $\omega_2(S)$ and $\phi(S)$.
4. Calculate the total service costs $\Phi(\mathcal{S}_n)$ from Equation (40).
5. If $\Phi(\mathcal{S}_n) < \Phi(\mathcal{S})^*$, update the interim optimal solution $(\Phi(\mathcal{S})^*, \mathbf{S}^*, \mathbf{Y}^*, \mathbf{F}^*, \mathbf{Z}^*, \mathbf{U}^*, \mathbf{T}^*, \mathbf{X}^*, \mathbf{Q}^*, \mathbf{B}^*)$, where $\Phi(\mathcal{S})^*$ is the minimum total service cost so far.
6. If $|\Phi(\mathcal{S}_n) - \Phi(\mathcal{S}_{n-1})| \leq \varepsilon$, stop. Otherwise, determine the optimal value for \mathbf{S}^* by proceeding to step (7).
7. Calculate the partial derivative of $\phi(\mathcal{S})$ with respect to every element in \mathbf{S} with respect to s^v . It is not possible to find this partial derivative analytically as the elements of \mathbf{Y} are integers, so we calculate the partial derivatives using perturbation analysis, as follows:
 - a) Given $\mathcal{S}_n = (s^1, s^2, \dots, s^t)$ and its corresponding solution $\mathbf{Y}_n, \mathbf{f}_n, \Phi(\mathcal{S}_n)$.
 - b) Set $\mathcal{S}_n^v = (s^1, s^2, s^v + \sigma^v, \dots, s^t)$, $\sigma^v = 0.2$.
 - c) If $s^v + \sigma^v > 1$, go to (e). Otherwise, solve stage (1) for a given \mathcal{S}_n^v to obtain solution $(\mathbf{Y}_n^v, \mathbf{f}_n^v)$.
 - d) If $\mathbf{Y}_n^v \neq \mathbf{Y}_n$ or $\mathbf{f}_n^v \neq \mathbf{f}_n$ go to step (8), otherwise increase $\sigma^v = \sigma^v + 0.2$ and go back to step (b).

- e) Set $\mathcal{S}_n^v = (s^1, s^2, s^v - \sigma^v, \dots, s^t)$, $\sigma^v = 0.2$.
- g) Solve stage (1) for a given \mathcal{S}_n^v to obtain solution $(\mathbf{Y}_n^v, \mathbf{f}_n^v)$.
- h) If $\mathbf{Y}_n^v \neq \mathbf{Y}_n$ or $\mathbf{f}_n^v \neq \mathbf{f}_n$ go to step (8), otherwise increase $\sigma^v = \sigma^v + 0.2$.
- i) If $s^v - \sigma^v > 0$, go to (e). Otherwise, set the sensitivity of s^v to zero.
8. Solve stage (2) for all generated scenarios $h \in H$ and calculate $\omega_2(S)$ and $\phi(S)$.
9. Apply Equation (40) to calculate total costs $\Phi(\mathcal{S}_n^v)$.
10. If $\Phi(\mathcal{S}_n^v) < \Phi(\mathcal{S})^*$, update the interim optimal solution
11. Calculate the sensitivity of s^v using $\frac{\Phi(\mathcal{S}_n^v) - \Phi(\mathcal{S}_n)}{\sigma^v}$.
12. If $v \neq t$, set $v = v + 1$, and go to step (b). Otherwise we now have all of the sensitivities of $\Phi(\mathcal{S})$ with respect to every element in \mathbf{S} . These are denoted $\nabla\Phi(\mathcal{S}_n)$.
13. Calculate the new service ratio vector for the next iteration, as follows:
- j) Take the negative sensitivity vector $-\nabla\Phi(\mathcal{S}_n)$ as the search direction.
- k) Calculate the step size in a similar way to the sub-gradient methods used for non-differentiable optimization [26, 27]: $\pi_n = \lambda_n \frac{\Phi(\mathcal{S}_n) - \tau\Phi(\mathcal{S})^*}{\|\nabla\phi(\mathcal{S}_n)\|^2}$. Where $\tau\Phi(\mathcal{S})^*$ is the estimated minimum value of $\Phi(\mathcal{S})$ and $\|\nabla\phi(\mathcal{S}_n)\|$ is the norm of the sensitivities. λ_n is a scaling factor. The values of (λ_n, τ) are problem-dependent and can be set to increase the efficiency of the solution process.
- m) Calculate the service ratio vector for the next iteration $\mathcal{S}_{n+1} = \mathcal{S}_n - \pi_n \nabla\Phi(\mathcal{S}_n)$.
- n) If $\Phi(\mathcal{S})^*$ is not improved after a specified number of iterations, go to (p), otherwise go to step (14).
- p) Use the optimal \mathbf{S}^* vector found so far as the starting point for the next iteration (i.e., $\mathcal{S}_{n+1} = \mathbf{S}^*$). Go to step (14). If $\Phi(\mathcal{S})^*$ is not improved after a specified number of iterations, go to (q).
- q) Divide the iteration scaling factor, λ_n , by a constant then go back to (k) and calculate the step size using the new scaling factor λ_{n+1} rather than the old scaling factor λ_n .
14. Set $n = n + 1$ and use the new service ratio vector \mathcal{S}_{n+1} when going back to step (2).

4. Numerical Results

The solution algorithm described in this study was developed in Matlab (Matlab Inc., Natick, MA, USA), and the optimization solver was the academic version of IBM CPLEX 12.6.1. The experiments were run on a 3.6-GHz i7-4960x computer. We tested our formulation and solution algorithm using a network with 11 nodes and 21 undirected arcs, as shown in Figure 2. The number on each link represents the link cost. Each bus vehicle has a maximum capacity of $Q_b = 25$ passengers. At most, two bus lines are allowed in the network. The upper and lower bounds of the frequency are $f_{max} = 10 \text{ veh/hr}$ and $f_{min} = 2 \text{ veh/hr}$, respectively. We used six DAR vehicles with a maximum capacity $Q_k = 8$ passengers. The unit bus operating cost per unit length is $c^1 = 2$ and the unit bus construction cost per unit length is $c^2 = 6$. We set $c^3 = 0.02$, $c^4 = 0.03$, $c^5 = 0.03$, $c^6 = 0.64$, $c^7 = 0.02$ and $\theta = 1$. The approximate DAR cost per passenger per unit length, c^8 , is set at two times the unit operating cost per bus per unit length. The maximum DAR vehicle route duration is 600 min. The random demand is uniformly distributed such that $B^d \approx U(0, 2b^d)$, where b^d is the average hourly demand, as presented in Table 1.

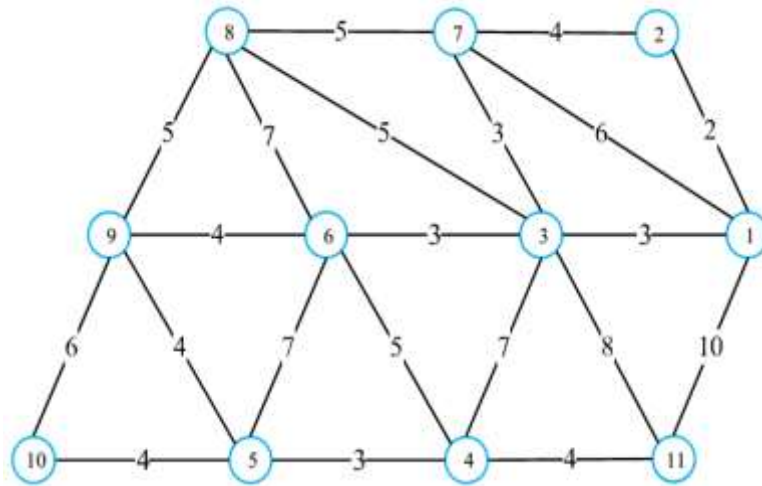


Figure 2. Network Setting

The details of the nine OD pairs are presented in Table 1, alongside the optimal solutions for the multimodal transport system. We demonstrate the advantages of the multimodal service by calculating the optimal solution when only a bus service is provided. The optimal solutions for this system when only a bus service is provided are shown in Table 2. The bus line alignments and passenger flows for the optimal multimodal service and bus service only solution are shown in Figure 3. The number on each link represents the passenger flow for the bus service. By comparing the optimal solutions in Tables 1 and 2, we observe that some OD pairs (e.g., 8-6, 2-4 and 11-1) are still served by buses in both solutions. In the multimodal case, other OD pairs (e.g., 6-9 and 9-2) are served by a combination of buses and DAR vehicles. Also, Table 1 shows that the passengers served by the DAR service might be offered a direct trip (e.g., OD 6-9 and OD 7-11) or an indirect trip (e.g., OD 6-9 and OD 1-5) from their origin to their destination. This decision is made by the operator and depends on many factors, such as the available on-board capacity, operation costs and passenger travel costs.

Table 1. OD Pair Details and Passenger Flow Results for the Multimodal Service

OD Pair	OD Demand (pass/hr)	Served by Bus	Path by Bus	Distance (km)	Served by DAR	Path by DAR	Distance (km)
6→9	50	37	6-9	4	8	6-9	4
					5	6-4-5-9	12
8→6	40	40	8-9-6	9	0	-	-
7→11	200	165	7-2-1-11	16	32	7-11	11
					3	7-2-11	16
2→4	90	90	2-7-3-4	14	0	-	-
1→5	150	29	1-2-7-3-4-5	19	2	1-6-4-5	14
		119	1-3-6-9-5	14			
11→1	50	50	11-1	10	0	-	-
4→8	100	90	4-5-9-8	12	1	4-5-9-8	12
		9	4-3-6-9-8	19			
5→7	110	110	5-4-3-7	13	0	-	-
9→2	120	117	9-6-3-1-2	12	3	9-8-7-2	14

Table 2. OD Pair Details and Passenger Flow Results for the Bus-Only Service

OD Pair	OD Demand (pass/hr)	Served by Bus	Path by Bus	Distance (km)
6→9	50	50	6-9	4
8→6	40	40	8-6	7
7→11	200	90	7-2-1-11	16
		110	7-8-9-5-4-1	21
2→4	90	90	2-1-11-4	16
1→5	150	20	1-2-7-8-9-5	20
		40	1-11-4-5	17
		90	1-2-7-8-6-9-5	26
11→1	50	50	11-1	10
4→8	100	10	4-5-9-8	12
		90	4-5-9-6-8	18
5→7	110	110	5-4-11-1-2-7	23
9→2	120	120	9-8-7-2	14

It is obvious from Figure 3 that the optimal bus line alignments and associated frequencies in the case of the multimodal service are different to those arising in the case of the bus-only service. This is because the existence of the DAR service decreases the on-board passenger flow between many links on the buses. For example, the path 7-2-1-11 on line 1 in Figure 3 has a maximum passenger flow of 220 in the case of the bus-only service and 165 in the case of the multimodal service. As OD 7-11 will be served by the bus and DAR in the multimodal service, some passengers have been shifted to the DAR, so the on-board passenger flow has decreased. These reductions in the passenger flows lead to a reduction in the bus line frequencies from (8.80, 5.20) to (6.60, 4.76).



(a) Solution for the Multimodal Service



(b) Solution for the Bus-Only Service

Figure 3. Resulting Bus Line Configuration and Passenger Flow

Figure 4 shows an example of two DAR vehicle itineraries. The number on each line represents the number of passengers in each DAR vehicle. Some DAR vehicles are shared between many OD pairs (e.g., DAR vehicle 1), while others are not (e.g., DAR vehicle 2). Also, DAR vehicle 1 is not fully occupied with passengers over the course of its entire itinerary 1-6-4-5-9-8-7-2-11, while DAR vehicle 2 is fully occupied for all of its itinerary 6-9. As we explained earlier, passengers may be offered either direct or indirect trips from their origin to their destination. For example, five passengers on OD 6-9 take the route 6-4-5-9 in DAR vehicle 1, while eight passengers with the same OD, 6-9, will be offered a

direct trip in DAR vehicle 2. Due to the maximum capacity restriction for DAR vehicles (8 passengers) the operator cannot offer all 13 passengers a direct trip.

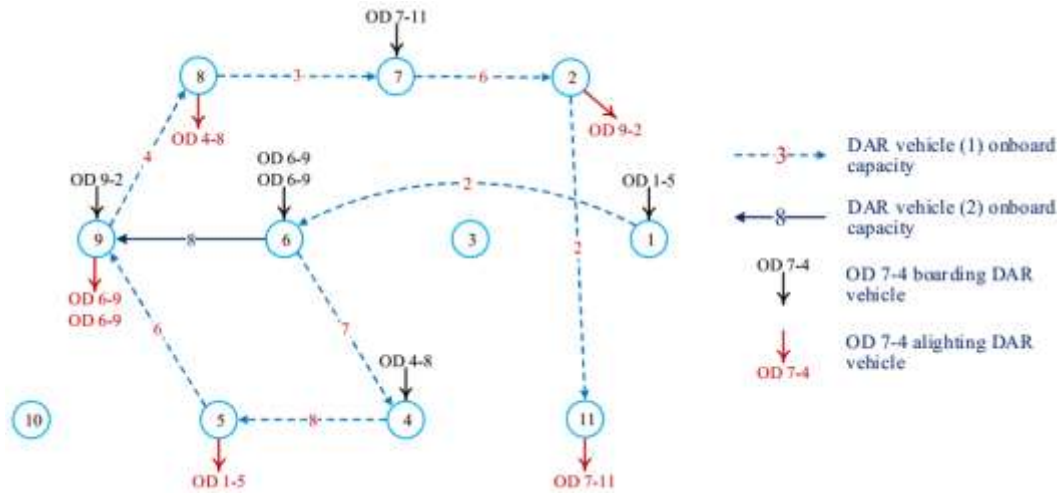


Figure 4. Example of Two DAR Vehicle Itineraries

Table 3 shows the total costs and a detailed cost breakdown for both the operator and the users, along with the total distance travelled by all passengers using the bus and DAR services. The results presented show that the total costs of the multimodal service are 14.13% lower than the costs of the bus-only service. The operator costs decreased from 1,047 to (900 = 841 + 59), and the passengers’ costs decreased from 391 to 334, when the DAR service was used to transport some passengers. As bus line frequencies can be reduced when the DAR service is offered, passenger costs, in particular waiting and transferring costs, decreased. Additionally, as bus line alignments changed so that shorter routes were offered, the total distance travelled by all passengers decreased by 24.17%, from 15,540 km to 11,784 km. Another reason for the reduction in the total distance travelled is that the DAR service travels by shorter paths than the bus service, for example in the case of OD 7-11. In summary, our results show that costs to both operators and users are reduced when multimodal transport services are offered. This proves that both operators and users benefit when operators offer a multimodal service. Also, the users’ perspective is considered when the operator chooses the service to provide.

Table 3. Comparison of the Components of the Total Cost between Different Optimal Solutions

Optimal Solution	Total Cost	Operator Costs		Users Costs		Total Distances Travelled (km)			Comp Time (min)
		Bus	DAR	Bus	DAR	Bus	DAR	Total	
Multimodal Service	1234	841	59	321 224 48 19 ^a	13	11210	574	11784	430
Bus-Only Service	1438	1047	0	391 311 42 38 ^a	0	15540	0	15540	6
An and Lo Model	1185	798	84	272 202 60 10 ^a	31	10104	1525	11629	142

^a Passenger cost components: in-vehicle travel costs| waiting costs| transfer costs

Similar work has been undertaken by An and Lo [22, 23]. However, though they calculated the approximate cost of providing flexible services, they did not incorporate flexible service design into their models. They calculated the approximate DAR cost by multiplying the approximate cost of using the DAR by the shortest path for OD pair.

However, this approximate approach assumes that the DAR vehicle always takes the shortest path, has infinite capacity to serve any number of passengers and is always fully occupied with passengers. As shown before in Figure 4, none of these assumptions are practical. Hence, as the approximated cost will be underestimated, the optimal combination of bus and DAR services found using this method may be incorrect. As the estimated cost of the DAR is lower than the actual cost, the operator will transport more passengers in the DAR service. We highlight the significance of our accurate approach by solving the network shown in Figure 2 using the model developed by An and Lo [23]. The optimal solution obtained for this system using the An and Lo model is shown in Table 4 and the bus line alignments and passenger flows are shown in Figure 5. The results in Table 4 show that, according to this solution, the DAR service carries more passengers than in the solution shown in Table 1, which we obtained using our model with accurate DAR costs. Hence, as shown in Table 3, the operator costs, user costs and total distance travelled are higher. As the process used to calculate the costs of the DAR service in the An and Lo model does not require any computational effort, there are differences in the computation time required to implement the two approaches. Our model is more accurate, but, as it is an NP-hard problem, more time is required to design the DAR service.

Table 4. OD Pair Details and Passenger Flow Results Arising from the an and Lo Model

OD Pair	OD Demand (pass/hr)	Served by Bus	Path by Bus	Distance (km)	Served by DAR	Path by DAR	Distance (km)
6→9	50	0	-	-	50	6-9	4
8→6	40	0	-	-	40	8-6	7
7→11	200	152	7-3-11	11	48	7-11	11
2→4	90	90	2-1-3-6-4	13	0	-	-
1→5	150	55	1-3-6-4-5	14	1	1-5	13
		95	1-2-7-8-9-5	20			
11→1	50	50	11-3-1	11	0	-	-
4→8	100	90	4-5-9-8	12	10	4-8	12
5→7	110	110	5-4-6-3-7	14	0	-	-
9→2	120	34	9-5-4-6-3-1-2	20	32	9-2	12
		54	9-8-7-2	14			

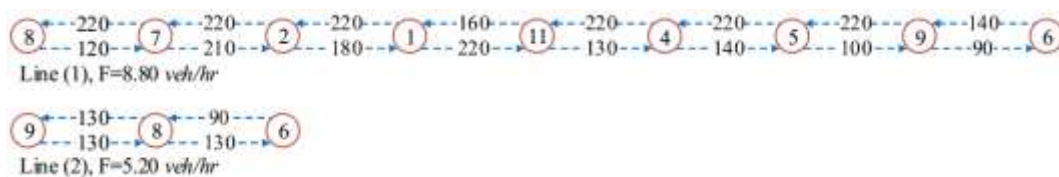


Figure 5. Bus Line Configuration and Passenger Flow Results for the Solution Using the an and Lo Model

We have calculated the approximate and accurate DAR costs, for the optimal solutions obtained from both models, to explain the difference between the approximate and accurate methods for calculating the DAR costs in more detail. As shown in Table 5 for example, according to our solution, the accurate operator cost of offering the DAR service is 59, while the equivalent approximate DAR cost is 31. The approximate cost of offering the DAR service is 84 according to the An and Lo model, but the equivalent accurate cost is 212. The trend is the same for the users' costs, the total distance travelled and the total cost. Table 5 shows that, compared to our accurate approach, the approximate approach in both solutions always underestimates the cost, especially the operator costs. The big

difference in the operator costs is because the DAR vehicles are not always fully occupied. In fact, they may be occupied by a small number of passengers, and sometimes there will be no passengers until a new passenger is picked up. The accurate users' costs and the total distance travelled are higher than the approximate costs, because passengers will sometimes travel to their destination indirectly. Hence, the approximate total costs are lower than the accurate values. The total cost of the solution by the An and Lo model is 1,185, while the equivalent accurate cost is 1,317. In summary, the optimal DAR/bus combinations determined by the approximate models differ from those calculated using the accurate model, and this affects the total cost.

Table 5. Comparison between the Accurate and Approximate Approaches to Calculating the Cost of the DAR Service

Optimal Solution	Operator's Cost (DAR)		Users' Cost (DAR)		Total Distance Travelled (DAR)		Total Costs	
	Acc ^a	App ^b	Acc ^a	App ^b	Acc ^a	App ^b	Acc ^a	App ^b
Multimodal Service	59	31	13	10	574	511	1234	1203
An and Lo Model	212	84	35	31	1548	1512	1317	1185

^a Acc: using accurate approach to calculate DAR costs

^b App: using approximate approach to calculate DAR costs (An and Lo – method)

5. Conclusion

We have developed an exact formulation for a multimodal public transport service comprising fixed and flexible transit services with stochastic demand. We considered the design of the two services to minimize the total costs to both the operator and passengers. Our results show that the advantages to operating a multimodal service including reductions in the total cost to the operator and passengers, and total travel distances for all passengers. The results also show that the accurate design of both services has a large impact on the optimized solution and total associated cost. The work presented in this paper can provide guidelines to the local authorities to determine the optimal combination between fixed and flexible services in serving the OD pairs. We believe that local authorities can maintain public transit services and decrease the negative effects of low passenger loads by combining fixed and flexible services. We will consider the effect of allowing passengers to transfer between the two types of service, and between DAR vehicles, in future studies. Also, the effect of flexible services on roadway congestions will be considered. For future extensions, we will use heuristic and metaheuristic methods to be able to extend this work to solve large networks by

References

- [1] O. J. Ibarra-Rojas, F. Delgado, R. Giesen and J. C. Muñoz, "Planning, operation, and control of bus transport systems: A literature review", *Transportation Research Part B: Methodological*, vol. 77, (2015), pp. 38-75.
- [2] V. Guihaire and J. K. Hao, "Transit network design and scheduling: A global review", *Transportation Research Part A: Policy and Practice*, vol. 42, no. 10, (2008), pp. 1251-1273.
- [3] K. Kepaptsoglou and M. Karlaftis, "Transit route network design problem: review", *Journal of Transportation Engineering*, vol. 135, no. 8, (2009), pp. 491-505.
- [4] A. Chen, Z. Zhou, P. Chootinan, S. Ryu, C. Yang and S. C. Wong, "Transport network design problem under uncertainty: A review and new developments", *Transport Reviews*, vol. 31, no. 6, (2011), pp. 743-768.
- [5] H. Shimamoto, F. Kurauchi, Y. Iida, M. G. Bell and J. D. Schmocker, "Evaluating public transit congestion mitigation measures using a passenger assignment model", *Journal of the Eastern Asia Society for Transportation Studies*, vol. 6, (2005), pp. 2076-2091.

- [6] W. Fan and R. Machemehl, "Bi-level optimization model for public transportation network redesign problem: Accounting for equity issues", *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2263, (2011), pp. 151-162.
- [7] W. Y. Szeto and Y. Jiang, "Transit route and frequency design: Bi-level modelling and hybrid artificial bee colony algorithm approach", *Transportation Research Part B: Methodological*, 67, (2014), pp. 235-263.
- [8] J. F. Guan, H. Yang and S. C. Wirasinghe, "Simultaneous optimization of transit line configuration and passenger line assignment", *Transportation Research Part B: Methodological*, vol. 40, no. 10, (2006), pp. 885-902.
- [9] B. C. Geiger, "Increasing the usage of demand-response transit in rural Kansas", Ph.D. dissertation, Kansas State University, (2009).
- [10] F. Kurauchi, K. Ueshima and Y. Iida, "Study on service characteristics of demand responsive transport using sequential demand assignment algorithm", *International Journal of ITS Research*, vol. 3, no. 1, (2005).
- [11] K. Saeed and F. Kurauchi, "Enhancing the service quality of transit systems in rural areas by flexible transport services", *Transportation Research Procedia*, vol. 10, (2015), pp. 514-523.
- [12] L. Häme, "An adaptive insertion algorithm for the single-vehicle dial-a-ride problem with narrow time windows", *European Journal of Operational Research*, vol. 209, no. 1, (2011), pp. 11-22.
- [13] M. Schilde, K. F. Doerner and R. F. Hartl, "Metaheuristics for the dynamic stochastic dial-a-ride problem with expected return transports", *Computers & Operations Research*, vol. 38, no. 12, (2011), pp. 1719-1730.
- [14] F. Guerriero, F. Pezzella, O. Pisacane and L. Trollini, "Multi-objective optimization in dial-a-ride public transportation", *Transportation Research Procedia*, vol. 3, (2014), pp. 299-308.
- [15] R. Masson, F. Lehuédé and O. Péton, "The dial-a-ride problem with transfers", *Computers & Operations Research*, vol. 41, (2014), pp. 12-23.
- [16] J. F. Cordeau, "A branch-and-cut algorithm for the dial-a-ride problem", *Operations Research*, vol. 54, no. 3, (2006), pp.573-586.
- [17] T. Y. Hu and C. P. Chang, "Exact algorithm for dial-a-ride problems with time-dependent travel cost", *Journal of the Eastern Asia Society for Transportation Studies*, vol. 10, (2013), pp. 916-933.
- [18] A. Trentini, R. Masson, F. Lehuédé, N. Malhéné, O. Péton and H. Tlahig, "A shared" passengers & goods" city logistics system", *Proceedings of the 4th International Conference on Information Systems, Logistics and Supply Chain*, Quebec, Canada, (2012), August.
- [19] V. Ghilas, E. Demir and T. Van Woensel, "The pickup and delivery problem with time windows and scheduled lines", *INFOR: Information Systems and Operational Research*, vol. 54, no. 2, (2016), pp. 147-167.
- [20] A. Lee and M. Savelsbergh, "An extended demand responsive connector", *EURO Journal on Transportation and Logistics*, vol. 6, no. 1, (2017), pp. 25-50.
- [21] C. H. Häll, H. Andersson, J. T. Lundgren and P. Värbrand, "The integrated dial-a-ride problem", *Public Transport*, vol. 1, no. 1, (2009), pp. 39-54.
- [22] K. An and H. Lo, "Service reliability-based transit network design with stochastic demand", *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2467, (2014), pp. 101-109.
- [23] K. An and H. K. Lo, "Two-phase stochastic program for transit network design under demand uncertainty", *Transportation Research Part B: Methodological*, vol. 84, (2016), pp. 157-181.
- [24] B. Yu, Z. Z. Yang, P. H. Jin, S. H. Wu and B. Z. Yao, "Transit route network design-maximizing direct and transfer demand density", *Transportation Research Part C: Emerging Technologies*, vol. 22, (2012), pp. 58-75.
- [25] H. K. Lo, K. An and W. H. Lin, "Ferry service network design under demand uncertainty", *Transportation Research Part E: Logistics and Transportation Review*, vol. 59, (2013), pp. 48-70.
- [26] T. G. Crainic, A. Frangioni and B. Gendron, "Bundle-based relaxation methods for multicommodity capacitated fixed charge network design", *Discrete Applied Mathematics*, vol. 112, no. 1, (2001), pp. 73-99.
- [27] D. Z. Wang and H. K. Lo, "Multi-fleet ferry service network design with passenger preferences for differential services", *Transportation Research Part B: Methodological*, vol. 42, no. 9, (2008), pp. 798-822.