Public Transp (2017) 9:217–241
DOI 10.1007/s12469-016-0128-9

ORIGINAL PAPER

The integrated dial-a-ride problem with timetabled fixed route service

Marcus Posada[1] · **Henrik Andersson**[2] · **Carl H. Häll**[1]

Accepted: 21 June 2016 / Published online: 24 November 2016

Abstract This paper concerns operational planning of door-to-door transportation systems for the elderly and/or disabled, who often need a more flexible transportation system than the rest of the population. Highly flexible, but very costly direct transportation is often offered as a complement to standard fixed route public transport service. In the integrated dial-a-ride problem (IDARP), these modes of transport are combined and certain legs of the passengers journeys may be performed with the fixed route public transport system. We extend the IDARP and include timetables for the fixed route services, forcing the fleet of vehicles to schedule the arrival at transfer locations with care. Two mixed integer linear programming formulations of the integrated dial-a-ride problem with timetables are presented and analyzed. The key modeling challenge is that of the transfers between the fleet of vehicles and the fixed route public transport system. The formulations differ in how the transfers are modeled and the differences are thoroughly discussed. The computational study compares the formulations in terms of network size, computational time and memory usage and conclusions about their performances are drawn.

Keywords Dial-a-ride · Operational planning · Paratransit · Public transport · Optimization · Mobility · Multimodal transport · Cost effectiveness

✉ Marcus Posada
marcus.posada@liu.se

[1] Department of Science and Technology (ITN), Linköping University, 60174 Norrköping, Sweden

[2] Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, 7491 Trondheim, Norway

1 Introduction

The elderly and/or persons with disabilities often need a more flexible transportation system than the rest of the population, due to problems using, or getting to and from, the fixed route public transport systems. Therefore, highly flexible door-to-door transport is offered as a complement to standard fixed route public transport to the elderly and/or the disabled in many parts of the world. Such systems are often called special transport systems or paratransit systems. These transport systems are often very costly and the fleets under-utilized. This paper concerns the operational planning of systems where door-to-door transport and fixed route public transport are integrated.

The dial-a-ride problem (DARP) concerns transportation of people and is an instance of the pickup and delivery problem (PDP). The DARP is to design vehicle routes and schedules to meet a number of requests for door-to-door transport under a number of side constraints, typically concerning trip duration, time windows, and vehicle capacity. This kind of service is demand responsive, in contrast to the fixed route public service. Vehicle routes are designed based on a set of origins, destinations, vehicle fleet information, and desired departure/arrival times. Commonly, the objective of this operational planning is to reduce the operating costs of the vehicles while keeping the level of service acceptable. The DARP, the PDP and other vehicle routing problems are further described in e.g. Toth and Vigo (2014). A review of the scientific literature on both exact and heuristic solution methods for the DARP is given in Cordeau and Laporte (2007).

The integrated dial-a-ride problem (IDARP) is an extension of the DARP in which certain legs of the passengers' journeys may be performed with a fixed route public transport system. Since demand responsive services are very costly compared with fixed route public transport, operators of demand responsive transport could possibly reduce their operating costs by using existing fixed route services. The IDARP is in many aspects similar to the pickup and delivery problem with transshipments (PTPT) and the dial-a-ride problem with transfers (DARPT). These problems are discussed in Sect. 3.

Figure 1 shows an example of when the integration of a demand responsive system and fixed route transport could be beneficial. A request is made for transportation from node **a** to node **b**. In a (non-integrated) dial-a-ride system, one of many possible solutions could be driving directly from the pick-up location to the drop-off location, this is represented by the two parallel lines. The solid curves represent a possible solution in an integrated system. Here, the request is picked up at **a** by a demand responsive vehicle, and driven to a transfer location (i.e. a fixed route stop where transfers between the demand responsive vehicles and the fixed route transport system is allowed). The request then travels by the public transport system to another transfer location and is there picked up by a different demand responsive vehicle and driven the final leg of the trip to **b**. It is clear that the distance traveled by demand responsive vehicles is shorter in the latter solution. It is therefore a cheaper solution and most likely preferable from an operational point of view. On the other hand, the service level of the trip is diminished since two

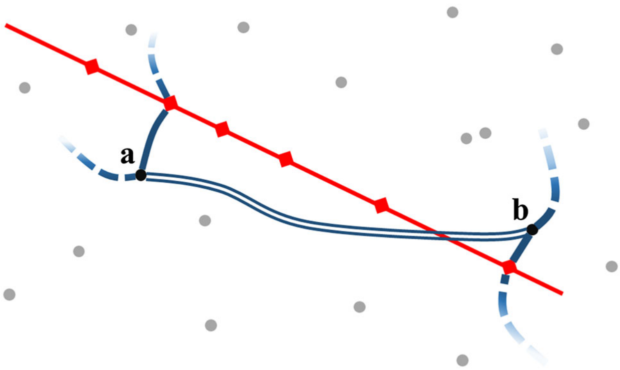

Fig. 1 This is an example of a scenario when integrating demand responsive and fixed route systems could be beneficial. A request has been made for transportation from **a** to **b**. The *dots* represent nodes belonging to requests (could be either pick-up or drop-off). The *straight line* represents a fixed route public transport system (e.g. a bus or tram line) with transfer locations at the squares. Both the *dashed* and *parallel lines* represent demand responsive vehicle routes

transfers have to be made between different travel modes. Note that the first (non-integrated) solution is a possible solution also in an IDARP model. Also note that it is not predetermined which transfer locations, if any, to use for each request. Which requests that are served by which combination of the two transport modes and which transfer nodes each of these requests will use is determined by minimizing the overall cost (e.g. operating cost of the demand responsive fleet and/or generalized cost of the passengers).

The purpose of this paper is to present a richer version of the IDARP than previous versions, significantly more applicable to real-world planning situations. Firstly, the new problem, called the integrated dial-a-ride problem with timetables (IDARP-TT) includes timetables for the fixed route service. Secondly, each request is described through different resources, such as the number of passengers and the number of wheelchairs. Thirdly, a fleet of heterogeneous vehicles, where the capacity regarding the different resources may differ is used for the demand responsive service. Fourthly, the speed of the demand responsive vehicles, as well as the vehicles on the fixed route lines may differ. Finally, not all pick-up and drop-off locations have to be visited by demand responsive vehicles, but a request can start or end with a fixed route transport. The key modeling challenge is that of the transfers between the demand responsive vehicles and the fixed route public transport system. We present two mixed integer linear programming models for the IDARP-TT that differ with regards to how the transfers are modeled.

The paper is organized as follows. In Sect. 2 some of the practical implementation issues of an integrated dial-a-ride-system are discussed. Section 3 provides a brief literature review of previous work on the IDARP and other relevant problems. The two formulations presented are given in Sect. 4 together with a small illustrative example. Methods to strengthen these formulations are described in Sect. 5. The rest of the paper concerns comparisons of the two models. Section 6 provides a theoretical comparison of the two formulations. Section 7 describes the specific evaluation scenario used and details some computational results. Conclusions and suggestions for future research are accounted for in Sect. 8.

2 Practical implementation issues

There are several possible practical issues associated with integrating public transportation lines into a demand responsive service, especially when the service is aimed at the elderly and/or disabled. In this section we discuss three of these issues which we consider to be especially relevant: lower levels of service due to transfers, delays, and adaptation costs.

There is discomfort associated with transfers. This diminishes the level of service of the special service passengers whose trips are integrated with the public transport system. How great this reduction in level of service is has not been studied and falls outside the scope of the current paper. In a practical implementation of integrated special services, the passengers whose trips can be made more efficient by integrating a public transport leg could be given the choice of which variant of their trip they wish to take. Most likely the integrated trip would be offered at a lower price. Not all special service passengers can be offered integrated trips, due to very special needs regarding equipment or levels-of-service. This is not an issue in the planning process, but affects the size of the system-wide improvements possible when implementing an integrated system.

The public transport vehicles (buses, trams, etc) could be delayed due to late demand responsive vehicles. One way of dealing with this would be to design the system so that the fixed route public transport never waits for delayed demand responsive vehicles. Another, similar issue is that delayed public transport vehicles could affect the operational planning of the demand responsive vehicles. Delays in the schedule of a demand responsive vehicle could propagate to other vehicles, since the passengers can make use of several vehicles in an integrated trip. Thus, in cases with large headways in the public transport timetable, it might be preferable for the dial-a-ride service to plan robustly. Although delays for the public transport passengers due to the additional passengers from the demand responsive service are to be expected, there is little reason to expect longer delays than those caused by, for example, strollers, other elderly passengers, or school classes.

There is a cost associated with adapting the public transport vehicles (and possibly transfer points) to make them accessible to users of special services. The cost of this adaptation will of course vary. The greater the level of adaptation in the public transport system, the larger the share of the special service passengers who can use integrated trips is likely to be. In Sweden, for example, already 75% of public transit buses are equipped with both lifts/ramps, space for wheelchairs and audiovisual information systems (The Swedish Bus and Coach Federation 2015).

Fixed costs and lowered level-of-service due to issues such as these have to be weighed against the possible lower operational costs of an integrated demand responsive system, but this analysis falls outside the scope of the current paper.

3 Literature review

The problem of combining demand responsive systems with fixed route public transport modes was first introduced in Wilson et al. (1976), a paper focusing on maximizing passenger utility rather than explicitly minimizing the operating costs. Another early paper is Potter (1976) which describes an integrated transport system in Ann Arbor, Michigan. Forty-five demand responsive vehicles serve as feeders to 36 express buses. In this system, connections between different demand responsive vehicles are allowed, in contrast with the system which is described in this paper. Both exact and heuristic methods have been presented since then, for static as well as dynamic versions of the problem. In static versions of the problem, all requests are known in advance, prior to the planning process, while in dynamic versions of the problem new requests are included in the solution as they come in, in real time. Whether static or dynamic versions of the problem are most useful in practice depends on the rules applied for how long in advance a request for transportation has to be made. These rules vary between different nations and different dial-a-ride systems. There are also methods that combine static and dynamic solution techniques. Liaw et al. (1996) formulate an optimization model and heuristic scheduling method for a dial-a-ride system with two transportation modes. Hickman and Blume (2001) present a two-stage scheduling heuristic for a similar problem. Another major contribution is made in Aldaihani and Dessouky (2003). The authors do not provide another model for the problem but present an insertion heuristic that builds on the work of Liaw et al. (1996) and Hickman and Blume (2001). All these heuristics are tested on case studies with real data. More recently, a Network Inspired Framework for Integrated Transport Systems (NITS) has been presented in Edwards et al. (2011). The NITS routes passengers through a transportation network in a manner inspired by the way in which data packets are routed through telecommunications networks. In this analogy, passengers are the data packets and the demand responsive service areas are subnetworks, for example. The NITS uses demand responsive vehicles in the last or first part of each trip if it is optimal. This system is then tested using simulation in Edwards et al. (2012) with promising results regarding quality of service and operating costs in low density urban areas. It can be noted that the focus of these two articles is that of transit in general low density urban areas, rather than transportation for the elderly and/or disabled passengers.

Other similar systems which integrate demand responsive and fixed route transport systems are described in for example Uchimura et al. (2002) and Crainic et al. (2001). Uchimura et al. (2002) develop a three-level hierarchical public transport system. The first two levels are defined as regional lines and express bus services. The third level of the system is a dial-a-ride system which is integrated with the higher levels of the system by providing both intracommunity transportation and a feeder system to regional transit. They solve the planning problem using a genetic algorithm. Crainic et al. (2001) analyze a transportation system where timetabled fixed route lines are integrated with lines with flexible itineraries and timetables. The aim of the system is to create a transportation system that provides a higher level of personalized service than a conventional transit line to a larger set of customers without the need for the large overhead required by door-to-door systems. Horn (2004) describes planning procedures designed for use in a real-time traveler information system in an urban environment where both demand responsive (taxi) and public transport modes are available. A simulation study shows the viability of the proposed framework. The public transport modeling system LITRES-2 described in Horn (2002) is used in Häll et al. (2008) to simulate and evaluate an integrated public transportation system. The results of the simulation study show that the attractiveness of an integrated system is greatly dependent on the pricing policy and the number of transfer nodes.

The IDARP is closely related to the pickup and delivery problem with transfers (PDPT), which is described in e.g. Cortés et al. (2010). In the PDPT and the similar dial-a-ride problem with transfers (DARPT), users may change vehicles during the trip. Cortés et al. (2010) give an arc-flow formulation of the problem and describe an exact solution method based on a branch-and-cut algorithm. The method used to model the transfer locations in Cortés et al. (2010) is conceptually quite similar to one of the formulations in this paper. The main difference between the IDARP and DARPT is that the users change mode at the transfer points in the IDARP, and then travel a specified distance with the public transport (PT). In Masson et al. (2014), an adaptive large neighborhood search method is used for solving the DARPT. It also describes the problem of how to check if a solution is feasible or not, and how this differs from how the feasibility is checked for a dial-a-ride service without transfers. Experiments with real-life data show savings of up to 8%, though the authors note that the passengers most likely experience a lowered level of service due to the transfers, which is not taken into account in the solution method. Deleplanque and Quilliot (2013) propose an algorithm for the DARPT which uses insertion heuristics and constraint propagation. In contrast with the fixed transfer points in e.g. Masson et al. (2014), the transfers in Deleplanque and Quilliot (2013) can occur anywhere. A problem similar to the DARPT is the pickup-and-delivery problem with transshipments (PDPT). A model for the PDPT is given in Rais et al. (2014) together with a small computational example using problem instances from Li and Lim (2003) and a review of the PDPT literature.

The Pickup and Delivery Problem with Time Windows and Scheduled Lines (PDPTW-SL) is examined in Ghilas et al. (2016). The PDPTW-SL concerns integrating short-haul passenger and freight transportation using public transport and thus has many similarities with the IDARP. Ghilas et al. (2016) present an Adaptive Large Neighborhood Search metaheuristic to efficiently solve the PDPTW-SL. They test the metaheuristic on generated instances with up to 100 freight requests.

Based on an arc-flow formulation of the DARP from Cordeau (2006), a mathematical programming formulation of the IDARP is given in Häll et al. (2009). Using this formulation the problem is solved exactly for small instances, up to ten requests. Finding an optimal feasible solution to the IDARP is an NP-hard problem, since it generalizes the DARP, which in turn is a generalization of the traveling salesman problem with time windows (see e.g. Cordeau 2006). Due to the design of the network, the size of the problem increases very quickly with both the number of requested trips and the number of transfer locations included. To counteract this rapid increase in problem size, the model is strengthened using arc elimination rules, variable substitutions and subtour elimination constraints. Some of these rules are specific to the IDARP while some are also used for the dial-a-ride, pickup and delivery, vehicle routing, and travelling salesman problems.

Häll et al. (2009) assume that the frequency on the fixed route lines is so high that customer waiting times at transfer locations can be disregarded. As the real-world environments where an integrated demand responsive service for the elderly and/or disabled could be useful are likely to be low density urban areas and more rural areas, this assumption could well limit the practical applicability of the model. In Ronald et al. (2015) this lack of synchronization on timetables is pointed out as an issue with the model. Also, the fixed route lines that have the largest potential for lowering operational costs are lines covering large distances, making the assumption of short waiting times possibly unrealistic. Additionally, the travel times of all vehicles, both demand responsive and fixed route, are assumed to be the same in the model presented in Häll et al. (2009). This is quite a rigid framework that does not allow for the introduction of, for example, commuter trains (possibly a lot faster than demand responsive vehicles) or bus lines with convoluted, non-direct, paths (possibly a lot slower than demand responsive vehicles taking the shortest path).

4 Formulations of the IDARP-TT

We extend the previous problem descriptions of the IDARP with the aim of addressing the issues described at the end of Sect. 3. With this aim in mind, timetables for the fixed route system are introduced. This is a very reasonable addition to the problem description since disregarding timetables could severely diminish the applicability of the model. We also introduce a fleet of heterogeneous demand responsive vehicles with different speeds, operational costs, and capacities.

Another feature of the new formulations is that the fixed route vehicles do not need to have the same speeds as the demand responsive vehicles. This is reasonable and extends the applicability of the model but makes some of the arc elimination rules for the DARP and IDARP introduced in Cordeau (2006) and Häll et al. (2009) invalid, since the triangle inequality for the demand responsive and fixed route vehicles is no longer true in all cases.

In previous DARP and IDARP models, all pick-up and drop-off nodes must be visited to serve a request. This is a sensible constraint for the DARP. However, in many practical cases that the IDARP describes, scenarios can be found where it is useful and reasonable to allow a request to end its trip at a transfer node that lies close to its drop-off node, without having to take a demand responsive vehicle the last leg of the trip. Similarly, cases can be found where it makes sense from both a cost minimizing and user-inconvenience standpoint to let the trips begin at transfer nodes instead of at the original pick-up nodes. Previous IDARP models do not allow this and, similarly to the DARP, require every pick-up and drop-off node to be visited. These restrictions have been relaxed in the two models presented in this paper. In these models, requests can under certain circumstances begin or end their journeys at transfer nodes that lie sufficiently close to their pick-up or drop-off nodes without involving demand responsive vehicles in the first or last leg of the journey.

4.1 Problem description

The IDARP-TT concerns the routing and scheduling of a fleet of demand responsive vehicles to serve a set of transport requests. Each request has a given origin, destination and demand for a set of resources, such as regular seats, places for wheelchairs and luggage. Time windows for the departure from the origin and the arrival at the destination as well as a maximum travel time are also defined for each request. A request can be transported from the origin to the destination by a single demand responsive vehicle or it can be transferred between a demand responsive vehicle and a fixed route transport system. A request may not be split between several vehicles, even though it may comprise several people. The fixed route system is defined over a set of transfer locations, and timetables dictate when it is possible to travel between the transfer locations. A heterogeneous fleet of vehicles with different speeds, operation costs, and capacities is located at a depot and used to transport the requests. The fleet is divided into different vehicle classes. The vehicles within each class are homogeneous. A route for a specific demand responsive vehicle is feasible if it begins and ends at the depot and if all requests which are picked up are also dropped off, either at their respective drop-off nodes or at a transfer node, while the capacity of the vehicle is not exceeded. For the set of all vehicle routes to be feasible, we require that each request leaves its origin and arrives at its destination inside the specified time windows. The objective is to find vehicle routes which minimize the operational cost of the demand responsive service and the usage cost of the fixed route transport system.

4.2 Two models

Studying the problem structure of the IDARP-TT through explicit modeling, and solving it to optimality, is of value since having an understanding of the problem facilitates the successful designing of heuristic solution methods. A heuristic solution method is needed to solve larger problem instances since the IDARP-TT is NP-hard. Thus, the usefulness of any explicit model solved to optimality is low in the sense that the solution time for any problem instance of realistic size will be prohibitively long.

In the model presented in Häll et al. (2009), each transfer location is modeled by one node for each request. Thus, each request adds $2 + g$ nodes to the network, where g is the number of physical transfer locations. This model design results in a large number of nodes to model all possible visits to every transfer location, making the network too large for the problem to be solved in a reasonable time for problem sizes above a few requests and a few transfer locations. One way of studying the properties and complexities of the problem is to change the way the transfer locations are modeled. Therefore, this paper presents two alternative formulations of the IDARP-TT, the first of which retains the basic structure of the model presented in Häll et al. (2009) in order to facilitate a comparison between the two ways of modeling the transfer locations. The second model, however, has a different transfer node structure. This node structure is similar to the one used in the model presented in Stålhane et al. (2014) describing a routing and scheduling problem faced by tramp shipping companies. In that model each node has two indices, the first represents the physical location and the second the visit number. This opens up for a location being visited multiple times. Applying the same basic node structure to the IDARP-TT significantly reduces the number of binary variables needed, as is shown in Sect. 6. The first model is presented in Sect. 4.3 and the second model is presented in Sect. 4.4.

4.3 Mathematical formulation of Model 1

The IDARP-TT is formulated over a directed graph $G = (\mathcal{N}, \mathcal{A})$ where $\mathcal{N}$ is the set of all nodes and $\mathcal{A}$ is the set of arcs connecting those nodes. The node set includes the vehicle depot, pick-up nodes, drop-off nodes, and transfer nodes. The depot is modeled as two nodes, a start node 0 and an end node $2\bar{r} + 1$ where $\bar{r}$ is the number of requests. The pick-up nodes are denoted $\mathcal{N}^P = \{1, \ldots, \bar{r}\}$ which is identical to the set of requests $\mathcal{R}$. The drop-off nodes are denoted $\mathcal{N}^D = \{\bar{r} + 1, \ldots, 2\bar{r}\}$. Thus, each pick-up node $i \in \mathcal{N}^P$ is associated with a drop-off node $i + \bar{r} \in \mathcal{N}^D$. The transfer nodes are modeled in the same way as in Häll et al. (2009); for each physical transfer location (e.g. a bus stop) $\bar{r}$ artificial transfer nodes are created. The set of artificial transfer nodes is denoted $\mathcal{N}^G = \{2\bar{r} + 2, \ldots, 2\bar{r} + 1 + \bar{r}g\}$ where g is the number of transfer locations in the fixed route transportation system. The sets $\mathcal{N}^G_r \subseteq \mathcal{N}^G$ contain the transfer nodes corresponding to request r. The set of arcs connecting artificial transfer nodes is denoted $\mathcal{A}^G$. The set of timetabled departures between the nodes i and j connected by arcs in $\mathcal{A}^G$ is denoted $\mathcal{D}_{ij}$. Thus, there are no departures between transfer nodes that are not connected in the fixed route transportation system or that correspond to different requests.

As mentioned before, one extension introduced in this model is the possibility of allowing requests to end or begin their trip at a fixed route stop. A natural example is at a hospital where the distance from the bus stop to the main entrance is short. These special cases are handled by introducing the appropriate pick-up or drop-off nodes to the set of artificial transfer nodes. The set $\mathcal{N}^G_{PD}$ constitutes a set of pick-up and drop-off nodes which are associated with a public transport stop. That is, these nodes comprise the pick-up and drop-off nodes of the request that are allowed to begin or end their trip at a specific fixed route stop. Overlap between $\mathcal{N}^P \cup \mathcal{N}^D$ and $\mathcal{N}^G$ is solved by removing certain elements of the latter set.

The heterogeneous vehicle fleet is divided into homogeneous vehicle classes. The set of vehicle classes is denoted $\mathcal{V}$ and the set of vehicles of class $v \in \mathcal{V}$ is denoted $\mathcal{K}_v$. A set of resources is defined as $\mathcal{S}$. Each vehicle class v has a capacity Q_{vs} of resource $s \in \mathcal{S}$ and each request r has a demand L_{rs} of resource s.

Sets

$\mathcal{R}$ Set of requests
$\mathcal{N}$ Set of all nodes, including pick-up nodes, drop-off nodes, depot nodes, and transfer nodes
$\mathcal{A}$ Set of arcs connecting the nodes
$\mathcal{A}^G$ Set of arcs connecting transfer nodes
$\mathcal{N}^P$ Set of pick-up nodes
$\mathcal{N}^D$ Set of drop-off nodes
$\mathcal{N}^G$ Set of transfer nodes
$\mathcal{N}^G_{PD}$ Subset of pick-up and drop-off nodes which are directly associated with a transfer node
$\mathcal{N}^G_r$ Subset of the transfer nodes that are associated with request r
$\mathcal{V}$ Set of demand responsive vehicle classes
$\mathcal{K}_v$ Set of vehicles of class v
$\mathcal{S}$ Set of resources
$\mathcal{D}_{ij}$ Set of departures between transfer node i and j

Variables

$$x_{ijvk} = \begin{cases} 1, & \text{if vehicle } k \text{ of class } v \text{ traverses arc } (i,j) \\ 0, & \text{otherwise} \end{cases}$$

$$y_{ijr} = \begin{cases} 1, & \text{if request } r \text{ travels by a demand responsive vehicle from node } i \\ & \text{to node j} \\ 0, & \text{otherwise} \end{cases}$$

$$z_{ijd} = \begin{cases} 1, & \text{if the fixed route from node } i \text{ to node } j \text{ using departure } d \text{ is used} \\ 0, & \text{otherwise} \end{cases}$$

t_i Time service at node i starts

Parameters

$\underline{T}_i$ Earliest start of service at node $i \in \mathcal{N}^P \cup \mathcal{N}^D$
$\overline{T}_i$ Latest service time at node $i \in \mathcal{N}^P \cup \mathcal{N}^D$
B_i Maximum travel time of request i
C_{ijv} Travel cost associated with arc (i, j) and vehicle class v
T_{ijv} Travel time associated with arc (i, j) and vehicle class v
L_{rs} Demand for resource s of request r
Q_{vs} Capacity of vehicle class v regarding resource s
T^D_{ijd} Departure time of departure d from node i to node j
T^A_{ijd} Arrival time at node j corresponding to departure d from node i
C_{ijd} Cost associated with departure d between transfer nodes i and j

$$F_{ir} = \begin{cases} 1, & \text{if node } i \text{ is the pick-up node of request } r \\ -1, & \text{if node } i \text{ is the drop-off node of request } r \\ 0, & \text{otherwise} \end{cases}$$

M Large positive number

Formulation

$$\min \sum_{(i,j)\in\mathscr{A}} \sum_{v\in\mathscr{V}} \sum_{k\in\mathscr{K}_v} C_{ijv} x_{ijvk} + \sum_{(i,j)\in\mathscr{A}^G} \sum_{d\in\mathscr{D}_{ij}} C_{ijd} z_{ijd} \tag{1}$$

Subject to:

$$\sum_{j\in\mathscr{N}} \sum_{v\in\mathscr{V}} \sum_{k\in\mathscr{K}_v} x_{ijvk} = 1, \quad i \in \mathscr{N}^P \cup \mathscr{N}^D \setminus \{\mathscr{N}^G_{PD}\} \tag{2}$$

$$\sum_{j\in\mathscr{N}} x_{0jvk} = 1, \quad v \in \mathscr{V}, k \in \mathscr{K}_v \tag{3}$$

$$\sum_{j\in\mathscr{N}} x_{ijvk} - \sum_{j\in\mathscr{N}} x_{jivk} = 0, \quad i \in \mathscr{N}^P \cup \mathscr{N}^D \cup \mathscr{N}^G, v \in \mathscr{V}, k \in \mathscr{K}_v \tag{4}$$

$$\sum_{i\in\mathscr{N}} x_{i,2\bar{r}+1,vk} = 1, \quad v \in \mathscr{V}, k \in \mathscr{K}_v \tag{5}$$

$$\sum_{j\in\mathscr{N}} y_{ijr} - \sum_{j\in\mathscr{N}} y_{jir} + \sum_{j\in\mathscr{N}^G_r} \sum_{d\in\mathscr{D}_{ij}} z_{ijd} - \sum_{j\in\mathscr{N}^G_r} \sum_{d\in\mathscr{D}_{ji}} z_{jid} = F_{ir}, \quad r \in \mathscr{R}, i \in \mathscr{N}^G_r \tag{6}$$

$$\sum_{j\in\mathscr{N}\setminus\{0,2\bar{r}+1\}} y_{ijr} - \sum_{j\in\mathscr{N}\setminus\{0,2\bar{r}+1\}} y_{jir} = F_{ir}, \quad r \in \mathscr{R}, i \in \mathscr{N} \setminus \mathscr{N}^G_r \tag{7}$$

$$\sum_{r\in\mathscr{R}} L_{rs} y_{ijr} \le \sum_{v\in\mathscr{V}} \sum_{K\in\mathscr{K}_v} Q_{vs} x_{ijvk}, \quad (i,j) \in \mathscr{A}, s \in \mathscr{S} \tag{8}$$

$$t_j \ge T_{0jv} \sum_{k\in\mathscr{K}_v} x_{0jvk}, \quad j \in \mathscr{N}, v \in \mathscr{V} \tag{9}$$

$$t_j - t_i - T_{ijv} + M(1 - \sum_{k\in\mathscr{K}_v} x_{ijvk}) \ge 0, \quad (i,j) \in \mathscr{A}, v \in \mathscr{V} \tag{10}$$

$$M\left(1 - \sum_{i\in\mathscr{N}^G} \sum_{d\in\mathscr{D}_{ij}} z_{ijd}\right) + \sum_{i\in\mathscr{N}^G} \sum_{d\in\mathscr{D}_{ij}} T^D_{ijd} z_{ijd} - t_i \ge 0, \quad j \in \mathscr{N}^G \tag{11}$$

$$t_j - \sum_{i\in\mathscr{N}^G} \sum_{d\in\mathscr{D}_{ij}} T^A_{ijd} z_{ijd} \ge 0, \quad j \in \mathscr{N}^G \tag{12}$$

$$\underline{T}_i \le t_i \le \overline{T}_i, \quad i \in \mathscr{N}^P \cup \mathscr{N}^D \tag{13}$$

$$t_{\bar{r}+i} - t_i \leq B_i, \quad i \in \mathcal{N}^P \tag{14}$$

$$x_{ijvk} \in \{0, 1\}, \quad (i,j) \in \mathcal{A}, v \in \mathcal{V}, k \in \mathcal{K}_v \tag{15}$$

$$y_{ijr} \in \{0, 1\}, \quad r \in \mathcal{R}, (i,j) \in \mathcal{A} \tag{16}$$

$$z_{ijd} \in \{0, 1\}, \quad (i,j) \in \mathcal{A}^G, d \in \mathcal{D}_{ij} \tag{17}$$

$$t_i \geq 0, \quad i \in \mathcal{N} \tag{18}$$

The objective function (1) minimizes the combined operational costs of the transportation service, both for the demand responsive vehicles and the fixed route service. The total vehicle distance is used as a proxy for the operational costs of the demand responsive vehicles. No extra components are added to the objective for the customer level of service since it is assumed that all feasible solutions have an acceptable level of service. The expansion of the objective function to include various measures of levels of service, such as the number of transfers or deviation from minimum possible travel time, is straightforward. Note that the set of requests is heterogeneous and therefore the fixed route costs are unique to the requests. Constraints (2)–(5) are typical DARP vehicle constraints. Constraint (2) says that every pick-up and drop-off node has to be visited by one vehicle, with the exception of the special cases where the requests are allowed to reach their final destinations or begin their journeys using public transport. Constraints (3) and (5) force the vehicles to begin and end their routes at the depots while (4) makes sure that each vehicle that visits a node also leaves it. Constraint (6) is a node balancing constraint for the transfer nodes. Constraint (7) is a node balancing constraints for the rest of the nodes. Constraint (8) guarantees that the vehicle capacities are not exceeded. Constraints (9) and (10) ensure that the travel times of the demand responsive vehicles are consistent. Both constraints (11) and (12) connect the demand responsive vehicles with the fixed route time tables. Constraint (11) ensures that the departure with the public transport vehicle occurs after the start of service at the transfer node. Constraint (12) serves the same purpose, but for arriving public transport vehicles. Constraint (13) guarantees that the service times at pick-up and drop-off nodes are within the defined time windows and constraint (14) makes sure that the maximum travel times of the requests are not exceeded. The final group of constraints, (15)–(18), defines the binary variables and the continuous time variables.

4.4 Mathematical formulation of Model 2

In Model 1 the problem size grows very quickly as a function of the number of requests. One way to diminish this effect is to create a model without request-specific nodes at the transfer locations.

Instead of allowing each request to have a specific node at each transfer location, this model has nodes representing every visit to that specific transfer location, regardless of which request or vehicle that makes the visit. Thus, several requests can use the same transfer node and all transfer nodes can be used by all requests. This diminishes the total number of nodes and we now know the order in which those nodes may be active, which decreases the number of arcs in the network. The magnitude of this decrease is described in Sect. 6.

This difference in the node structure leads to some differences in how the nodes are represented in Model 2 compared with Model 1. Each node is denoted (i, m) where i represents the physical location and m is the visit number. For example, the node (16, 3) represents the third visit to location 16. Observe that for the pick-up or drop-off nodes only one visit is possible and thus they are denoted $(i, 1)$.

In Model 2 we define the problem on a directed graph $\mathscr{G} = (\mathscr{N}, \mathscr{A})$ where $\mathscr{N}$ is the set of nodes and $\mathscr{A}$ is the set of arcs. Each node in $\mathscr{N}$ is denoted (i, m) where $i \in \{0, \ldots, 2\bar{r} + 1 + g\}$ is the physical locations and $m \in \{1, \ldots, \overline{M}_i\}$ the visit number, where $\overline{M}_i$ is an upper bound on the number of visits to location i. The depot is modeled as two nodes, a start node (0, 1) and an end node $(2\bar{r} + 1, 1)$. The pick-up nodes are denoted $\mathscr{N}^P = \{(1, 1), \ldots, (\bar{r}, 1)\}$ and the drop-off nodes are denoted $\mathscr{N}^D = \{(\bar{r} + 1, 1), \ldots, (2\bar{r}, 1)\}$. The set of nodes representing the transfer location g is denoted $\mathscr{N}^G_g = \{(g, 1), \ldots, (g, \overline{M}_g)\}$.

The special cases where a request is allowed to begin or end its trip at a transfer location are modeled by introducing a variable w_{imjnr} which is activated if the request walks the first or last leg of its trip. One extra set of variables, a_{im}, is needed to indicate if an artificial transfer node (i, m) is visited or not. This is necessary to make sure that the transfer nodes are visited in the correct order.

Sets

$\mathscr{G}$ Set of transfer locations
$\mathscr{N}$ Set of all nodes
$\mathscr{A}$ Set of arcs connecting the nodes
$\mathscr{N}^P$ Set of pick-up nodes
$\mathscr{N}^D$ Set of drop-off nodes
$\mathscr{N}^G$ Set of transfer nodes
$\mathscr{N}^P_g$ Subset of pick-up nodes directly connected to transfer location g
$\mathscr{N}^D_g$ Subset of drop-off nodes directly connected to transfer location g
$\mathscr{R}$ Set of requests
$\mathscr{V}$ Set of demand responsive vehicle classes
$\mathscr{K}_v$ Set of vehicles of class v
$\mathscr{S}$ Set of resources
$\mathscr{D}_{ij}$ Set of departures between the transfer locations i and j

Variables

$$x_{imjnvk} = \begin{cases} 1, & \text{if vehicle } k \text{ of class } v \text{ travels from node } (i,m) \text{ to node } (j,n) \\ 0, & \text{otherwise} \end{cases}$$

$$y_{imjnr} = \begin{cases} 1, & \text{if request } r \text{ travels by } a \text{ demand responsive vehicle from node} (i,m) \\ & \text{ to node } (j,n) \\ 0, & \text{otherwise} \end{cases}$$

$$z_{imjndr} = \begin{cases} 1, & \text{if the fixed route departure } d \text{ from node } (i,m) \text{ to node } (j,n) \text{ is used} \\ & \text{by request } r \\ 0, & \text{otherwise} \end{cases}$$

$$w_{imjnr} = \begin{cases} 1, & \text{if request } r \text{ walks from node } (i,m) \text{ to node } (j,n) \\ 0, & \text{otherwise} \end{cases}$$

$$a_{im} = \begin{cases} 1, & \text{if artificial transfer node } (i,m) \text{ is visited} \\ 0, & \text{otherwise} \end{cases}$$

t_{im} Time service at node (i, m) starts

Parameters

$\underline{T}_i$ Earliest start of service at location $i \in \mathcal{N}^P \cup \mathcal{N}^D$
$\overline{T}_i$ Latest start of service at location $i \in \mathcal{N}^P \cup \mathcal{N}^D$
B_r Maximum travel time of request r
C_{ijv} Travel cost associated with travelling between locations i and j with vehicle class v
T_{ijv} Travel time between location i and j with vehicle class v
L_{rs} Load of request r regarding resource s
Q_{vs} Capacity of vehicle class v regarding resource s
T^D_{ijd} Departure time of departure d from location i to location j
T^A_{ijd} Arrival time at location j corresponding to departure d from location i
C_{ijdr} Cost associated with departure d between locations i and j for request r

$$F_{ir} = \begin{cases} 1, & \text{if node } (i,1) \text{ is the pick-up node of request } r \\ -1, & \text{if node } (i,1) \text{ is the drop-off node of request } r \\ 0, & \text{otherwise} \end{cases}$$

M Large positive number

Formulation

$$\min \sum_{(i,m)\in\mathcal{N}} \sum_{(j,n)\in\mathcal{N}} \sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{K}_v} C_{ijv} x_{imjnvk} + \sum_{(i,m)\in\mathcal{N}^G} \sum_{(j,n)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} \sum_{r\in\mathcal{R}} C_{ijdr} z_{imjndr} \tag{19}$$

Subject to:

$$\sum_{(j,n)\in\mathcal{N}} \sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{K}_v} x_{i1jnvk} = 1, \quad (i,1) \in \mathcal{N}^P \cup \mathcal{N}^D \setminus \{\mathcal{N}^P_g \cup \mathcal{N}^D_g : g \in \mathcal{G}\} \tag{20}$$

$$\sum_{(j,n)\in\mathcal{N}} x_{01jnvk} = 1, \quad v \in \mathcal{V}, k \in \mathcal{K}_v \tag{21}$$

$$\sum_{(j,n)\in\mathcal{N}} x_{jnimvk} - \sum_{(j,n)\in\mathcal{N}} x_{imjnvk} = 0, \\ (i,m) \in \mathcal{N}^P \cup \mathcal{N}^D \cup \mathcal{N}^G, v \in \mathcal{V}, k \in \mathcal{K}_v \tag{22}$$

$$\sum_{(i,m)\in\mathcal{N}} x_{im2\bar{r}+1,1vk} = 1, \quad v \in \mathcal{V}, k \in \mathcal{K}_v \tag{23}$$

$$\sum_{(j,n)\in\mathcal{N}^G} w_{i1jnr} - \sum_{(j,n)\in\mathcal{N}^G} w_{jni1r} + \sum_{(j,n)\in\mathcal{N}} y_{i1jnr} - \sum_{(j,n)\in\mathcal{N}} y_{jni1r} = F_{ir}, \\ (i,1) \in \mathcal{N}^P \cup \mathcal{N}^D, r \in \mathcal{R} \tag{24}$$

$$\sum_{(j,1)\in\mathcal{N}_i^D} w_{imj1r} - \sum_{(j,1)\in\mathcal{N}_i^P} w_{j1imr} + \sum_{(j,n)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} z_{imjndr} \\ - \sum_{(j,n)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} z_{jnimdr} + \sum_{(j,n)\in\mathcal{N}} y_{imjnr} - \sum_{(j,n)\in\mathcal{N}} y_{jnimr} = 0, \tag{25}$$

$$(i,m) \in \mathcal{N}^G, r \in \mathcal{R}$$

$$w_{i1jni} - \sum_{(k,m)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} z_{jnkmdi} = 0, \quad (j,n) \in \mathcal{N}^G, (i,1) \in \mathcal{N}_j^P \tag{26}$$

$$\sum_{r\in\mathcal{R}} L_{rs} y_{imjnr} - \sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{K}_v} Q_{vs} x_{imjnvk} \le 0, \quad (i,m) \in \mathcal{N}, (j,n) \in \mathcal{N}, s \in \mathcal{S} \tag{27}$$

$$t_{jn} - T_{0jv} + M(1 - \sum_{k\in\mathcal{K}_v} x_{01jnvk}) \ge 0, \quad (j,n) \in \mathcal{N}, v \in \mathcal{V} \tag{28}$$

$$t_{jn} - t_{im} - T_{ijv} + M(1 - \sum_{k\in\mathcal{K}_v} x_{imjnvk}) \ge 0, \quad (i,m) \in \mathcal{N}, (j,n) \in \mathcal{N}, v \in \mathcal{V} \tag{29}$$

$$t_{j1} - t_{im} + M(1 - w_{imj1,j-\bar{r}}) \ge 0, \quad (i,m) \in \mathcal{N}^G, (j,1) \in \mathcal{N}_i^D \tag{30}$$

$$t_{jn} - t_{i1} + M(1 - w_{i1jni}) \ge 0, \quad (j,n) \in \mathcal{N}^G, (i,1) \in \mathcal{N}_j^P \tag{31}$$

$$M + \sum_{(j,n)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} (T_{ijd}^D - M) z_{imjndr} - t_{im} \ge 0, (i,m) \in \mathcal{N}^G, r \in \mathcal{R} \tag{32}$$

$$t_{jn} - \sum_{(i,m)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} T_{ijd}^A z_{imjnd} \ge 0, \quad (j,n) \in \mathcal{N}^G, r \in \mathcal{R} \tag{33}$$

$$t_{im} - t_{i,m+1} \le 0, \quad i \in \mathcal{G}, (i,m) \in \mathcal{N}_i^G \tag{34}$$

$$\underline{T}_i \leq t_{i1} \leq \overline{T}_i, \quad (i,1) \in \mathcal{N}^P \cup \mathcal{N}^D \tag{35}$$

$$t_{\bar{r}+i,1} - t_{i1} \leq B_i, \quad (i,1), \in \mathcal{N}^P \tag{36}$$

$$Ma_{im} \geq \sum_{(j,n)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} z_{imjndr} + \sum_{(j,n)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} z_{jnimdr}, \quad (i,m) \in \mathcal{N}^G, r \in \mathcal{R} \tag{37}$$

$$a_{im} \leq \sum_{(j,n)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} z_{imjndr} + \sum_{(j,n)\in\mathcal{N}^G} \sum_{d\in\mathcal{D}_{ij}} z_{jnimdr}, \quad (i,m) \in \mathcal{N}^G, r \in \mathcal{R} \tag{38}$$

$$a_{i,m+1} - a_{im} \leq 0, \quad i \in \mathcal{G}, (i,m) \in \mathcal{N}_i^G \tag{39}$$

$$x_{imjnvk} \in [0,1], \quad (i,m) \in \mathcal{N}, (j,n) \in \mathcal{N}, v \in \mathcal{V}, k \in \mathcal{K}_v \tag{40}$$

$$y_{imjnr} \in [0,1], \quad (i,m) \in \mathcal{N}, (j,n) \in \mathcal{N}, r \in \mathcal{R} \tag{41}$$

$$w_{i1jnr} \in [0,1], \quad j \in \mathcal{G}, (i,1) \in \mathcal{N}_j^P, (j,n) \in \mathcal{N}_j^G, r \in \mathcal{R} \tag{42}$$

$$w_{imj1r} \in [0,1], \quad i \in \mathcal{G}, (j,1) \in \mathcal{N}_i^D, (i,m) \in \mathcal{N}_i^G, r \in \mathcal{R} \tag{43}$$

$$z_{imjndr} \in [0,1], \quad i,j \in \mathcal{G}, (i,m) \in \mathcal{N}_i^G, (j,n) \in \mathcal{N}_j^G, d \in \mathcal{D}_{ij} \tag{44}$$

$$a_{im} \in [0,1], \quad i \in \mathcal{G}, (i,m) \in \mathcal{N}_i^G \tag{45}$$

$$t_{im} \geq 0, \quad (i,m) \in \mathcal{N} \tag{46}$$

Most of the constraints in Model 2 serve the same purpose as their corresponding constraints in Model 1 although there are some groups of constraints which are needed here but not in Model 1, due to the changes in transfer node structure. These are (30), (31), (34) and (37)–(39). Constraints (30) and (31) ensure that service times at drop-off and pick-up nodes which are being walked to are consistent with the service times of the involved transfer nodes. Constraint (34) makes sure the service times of the transfer nodes are consistent with the ordering. The final group of new constraints, (37)–(39), defines the variable, a_{im} that governs that transfer nodes are visited in correct order.

4.5 Illustrative example of the two models

Assume an instance with six requests. The example shows how the same transfer location is represented in Models 1 and 2, respectively, in the same solution. Vehicle 1 arrives at the transfer location at 10.30 to drop off requests 2 and 3 and pick up requests 1 and 5. Vehicle 2 arrives at the transfer location at 10.40 to pick up request 4. Request 4 arrives to the transfer location with fixed route transport at 10.40 while requests 5 and 1 arrive at 10.50. Both requests 2 and 3 use the same departure from the transfer node, at 10.50. The same solution is presented for both models in Fig. 2.

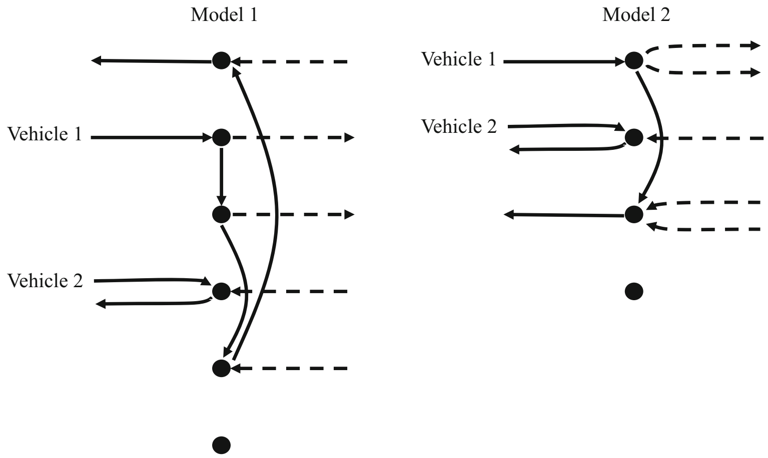

Fig. 2 A comparison of the node structures of the two models for an instance where there are six requests. Model 1 is shown on the *left*, Model 2 on the *right*. In Model 1 the node corresponding to request 1 is on the *top* and the node corresponding to request 6 is at the *bottom*. In Model 2 the nodes are ordered such that the *topmost* node corresponds to the first visit and the *bottom* node to the last visit

Note that

- With six requests, six transfer nodes are needed at the transfer location. Assume that precalculations gave that only four transfer nodes are needed in Model 2. Note that the number of nodes at the transfer location increases linearly with the number of requests in Model 1, while the number of nodes in Model 2 also depends on the number of fixed route transportation departures/arrivals.
- The order of the drop-offs can be interchanged for vehicle 1 in Model 1, the same goes for the pick-ups. This gives four mathematically different solutions that are the same in practice (and that have the same objective function value). Due to the node structure in Model 2 this symmetry problem has been eliminated. Since the arrival time of the fixed route transportation carrying request 4 arrives earlier than the fixed route transportation carrying requests 1 and 5, but later than the departure time of requests 2 and 3, vehicle 2 visits node number 2 and vehicle 1 visits nodes 1 and 3 in Model 2.
- In Model 1, seven variables corresponding to demand responsive vehicles, x-variables, and five variables corresponding to fixed route transportation, z-variables, are non-zero. There are six nodes corresponding to the same transfer location. For Model 2, five x-variables, five z-variables and four nodes are needed.

5 Strengthening the models

To evaluate the performance of the two models, they are compared as both unstrengthened and after some strengthening has been applied. The strengthening of the models is done by arc elimination and by adding some additional sets of constraints. In short, the arc elimination rules used eliminate all arcs that go:

- from-to the same node
- to the node representing the start depot
- from the node representing the end depot
- from start depot to any drop-off node
- from any pick-up node to end depot
- from a drop-off node to the pick-up node for the same request.

These elimination rules makes it possible to eliminate all x-variables (and y-variables) that use any arc (i, j) fulfilling the above statements. There are also some arcs (i, j) that can be used by some vehicles but not by others, meaning that some x-variables can be fixed to 0. These rules are that:

- if: $Q_{vs} < L_{rs}$, then $x_{rjvk} = 0$ and $x_{jrvk} = 0, \quad j \in N, k \in \mathcal{K}_v, v \in \mathcal{V}, s \in \mathcal{S}$
- if: $Q_{vs} < L_{is} + L_{js}$, then $x_{ijvk} = 0$, $x_{i,j+\bar{r},vk} = 0$, $x_{i+\bar{r},j+\bar{r},vk} = 0, \quad k \in \mathcal{K}_v, s \in \mathcal{S}$
- if: $\underline{T}_i + T_{ijv} > \overline{T}_j$ then $x_{ijvk} = 0, \quad i \in \mathcal{N}^P \cup \mathcal{N}^D, j \in \mathcal{N}^P \cup \mathcal{N}^D, k \in \mathcal{K}_v, v \in \mathcal{V}$.

The first rule says that if the capacity of vehicle class v regarding resource s is less than the demand that request r has of resource s, then no vehicle belonging to vehicle class v can be used to visit the pick-up or drop-off node belonging to request r (and in Model 1 nor to/from any transfer nodes of request r). This means, for example, that since a request requiring wheelchair transportation must be served by a vehicle with such capacity, arcs going to or from a pick-up or drop-off node belonging to such a request can be eliminated for all other vehicles.

The second rule says that if the capacity of a vehicle of class v is less than the demand of request i and j (for any specific resource s) then arc (i, j), $(i, j + \bar{r})$ and $(i + \bar{r}, j + \bar{r})$ can be eliminated for all vehicles of class v.

The third rule says that if a vehicle of class v cannot travel from node i to node j and meet the requirements of time windows of both nodes, then arc (i, j) can be eliminated for all vehicles of class v.

After the arc elimination rules have been applied the models are further strengthened by adding subtour elimination constraints. By finding, and eliminating subtours in the LP-relaxation of the problem the lower bound of the original (MILP-formulation) is strengthened. The subtour elimination constraints added are based on the identification of clusters of nodes including at least one pick-up or drop-off node that must be visited by a vehicle. This means that at least one vehicle has to enter each such cluster. Based on the formulation of Model 1, the following set of constraints can be added:

$$\sum_{i \in \varphi} \sum_{j \in \mathcal{N} \setminus \varphi} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} x_{ijvk} \geq 1 \qquad \varphi \in \Phi \tag{47}$$

where Φ is the set of all identified clusters and φ is one such cluster.

In the same way, constraints can be added saying that if a pick-up node or a drop-off node, belonging to a customer that requires wheelchair transportation is part of a cluster φ, at least one vehicle that can perform such a transportation must enter the cluster. So, if $\hat{v}$ is the only class of vehicles that can transport passengers in wheelchairs, the constraints for such clusters can be strengthened to:

$$\sum_{i \in \varphi} \sum_{j \in \mathcal{N} \setminus \varphi} \sum_{k \in \mathcal{K}_{\hat{v}}} x_{ij\hat{v}k} \geq 1 \qquad \varphi \in \Phi. \tag{48}$$

Since there is no reason for vehicles to travel in both directions between any pair of nodes we can also add the set of constraints:

$$\sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} (x_{ijvk} + x_{jivk}) \leq 1 \qquad (i,j) \in \mathcal{A}. \tag{49}$$

The fact that there is no reason for a demand responsive vehicle to visit a transfer node from/to which no fixed route service is used makes it possible to add the following set of constraints:

$$\sum_{j \in N} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} x_{ijvk} \leq \sum_{j \in \mathcal{N}^G} \sum_{d \in \mathcal{D}_{ij}} (z_{ijd} + z_{jid}) \qquad i \in \mathcal{N}^G \setminus \mathcal{N}^G_{PD}. \tag{50}$$

Since a route performed by a vehicle of class v can be performed by any other vehicle of that class, symmetry breaking constraints can be added to strengthen the formulation. One way of doing this is by adding the following set of constraints saying that the total travel time of a lower-numbered vehicle of a specific class must be higher than a vehicle with a higher number.

$$\sum_{(i,j) \in \mathcal{A}} T_{ijv} x_{ijvk} \geq \sum_{(i,j) \in \mathcal{A}} T_{ijv} x_{ijv,k+1} \qquad v \in \mathcal{V}, k \in \mathcal{K}_v \tag{51}$$

Even though constraints (47)–(51) are described based on the formulation of Model 1, they can just as well be formulated in accordance with Model 2. However, the next set of constraints is based on the fact that it is known what nodes belong to a request in Model 1. This makes it possible to add:

$$\sum_{i \in \mathcal{N}^G_r} \sum_{j \in N} \sum_{v \in \mathcal{V}} \sum_{k \in \mathcal{K}_v} x_{ijvk} \leq 2 \qquad r \in \mathcal{R}. \tag{52}$$

This set of constraints can be added since no request needs to visit more than two of its own transfer nodes. However, this is not applicable in Model 2, since in that model any transfer node can be used by any request.

Several of the above presented strengthening methods were previously described for the IDARP in Häll et al. (2009) and have in this section been reformulated to fit the formulations of the IDARP-TT. For further details regarding these strengthening methods readers are referred to Häll et al. (2009).

6 A theoretical comparison

The idea behind the transfer node structure introduced in Model 2 was to reduce the number of binary variables created in Model 1, since the computational times and use of memory increase rapidly with the number of binary variables. The number of nodes in Model 1 is $|\mathcal{N}_1| = 2\bar{r} + 2 + g\bar{r}$ and the number of arcs is $|\mathcal{A}_1| \propto |\mathcal{N}_1|(|\mathcal{N}_1| - 1)$. Using this, the numbers of binary variables in Model 1 are:

$$\left|x_{ijvk}\right| \propto \bar{K}|\mathcal{A}_1| \qquad \in O(\bar{r}^2) \tag{53}$$

$$\left|y_{ijr}\right| \propto \bar{r}|\mathcal{A}_1| \qquad \in O(\bar{r}^3) \tag{54}$$

$$\left|z_{ijd}\right| \propto d(g-1)\bar{r} \qquad \in O(\bar{r}) \tag{55}$$

where $\bar{K}$ is the total number of vehicles, regardless of class, d is the number of departures with the public transport, and $\left|x_{ijvk}\right|$, $\left|y_{ijr}\right|$ and $\left|z_{ijd}\right|$ denote the number of binary variables of each kind. For Model 2, the number of nodes instead becomes $|\mathcal{N}_2| = 2\bar{r} + 2 + g\min(d, \bar{r})$, from which $|\mathcal{A}_2|$ can be calculated in the same way as for Model 1. The numbers of binary variables are:

$$\left|x_{imjnvk}\right| \propto \bar{K}|\mathcal{A}_2| \qquad \in O(\bar{r}^2) \tag{56}$$

$$\left|y_{imjnr}\right| \propto \bar{r}|\mathcal{A}_2| \qquad \in O(\bar{r}^3) \tag{57}$$

$$\left|z_{imjndr}\right| \propto d(g-1)\min(d, \bar{r})^2 \qquad \in O(\bar{r}). \tag{58}$$

In Model 2, there are two more sets of variables: w_{imjnr} and a_{im}. These are both constants ($O(\bar{r}^0)$) as soon as $\min(d, \bar{r}) = d$ and can therefore be excluded from the total number of binary variables as $\bar{r}$ grows. Figure 3 shows the total number

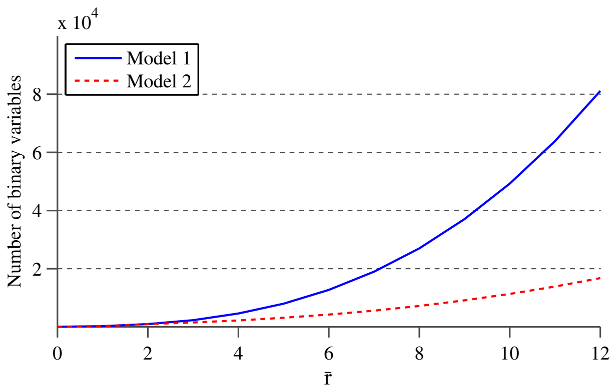

Fig. 3 The number of binary variables in the two models as functions of the number of requests, $\bar{r}$, in the case when there are three demand responsive vehicles, eight transfer locations and two fixed route departures from each transfer location. The *blue solid line* corresponds to Model 1 and the *red dotted line* to Model 2

of binary variables in the two model formulations as functions of the number of requests when there are three available demand responsive vehicles, four transfer locations, and two public transport departures between each of the transfer locations during the relevant time interval. The figure shows the number of variables in the original models, without the strengthening techniques described in Sect. 5.

It can be seen that even though the asymptotical behavior of the numbers of binary variables in the two models is the same as a function of the number of requests, the results for small- to medium-sized problem instances are very different. Model 2 has a much lower rate of growth than Model 1 from the point where $\min(d, \bar{r}) = d$ and onward, as expected. This is due to the sizes of the constants corresponding to the cubic and quadratic terms in expressions (53)–(55) and (56)–(58).

7 Results in an evaluation scenario

This section provides a small evaluation scenario with one public transport line (following a timetable generated to fit the travel times of the demand responsive vehicles). Further details regarding the example are given in Sect. 7.1 and computational results are given in Sect. 7.2.

7.1 Evaluation scenario

Figure 4 shows the evaluation scenario used. Six requests for demand responsive transport are placed in a geographical area (the town of Norrköping, Sweden) together with four public transport transfer locations. The locations shown in the figure are numbered so that the pick-up locations are 1–6 and their corresponding drop-off locations are 7–12. Five of the requests included in this example are special cases in the sense that they are allowed to begin or end their trip with public transport. In the case of request 1, the corresponding pick-up location is situated very close to the leftmost transfer location and the request may therefore perform the first leg of the trip with public transport. For requests 2, 3, 5 and 6, all of their drop-off locations are situated close to transfer locations and they are allowed to perform the last legs of their respective trips with public transport.

This example is used in Sect. 7.2 to illustrate the differences between the two models. The costs and travel times used in the numerical calculations come from the geography of the example town. Note that this is a small example, and in a real world instance, it is likely that an integrated DAR service would have greater benefits in a rural or inter-city setting where the distances are larger.

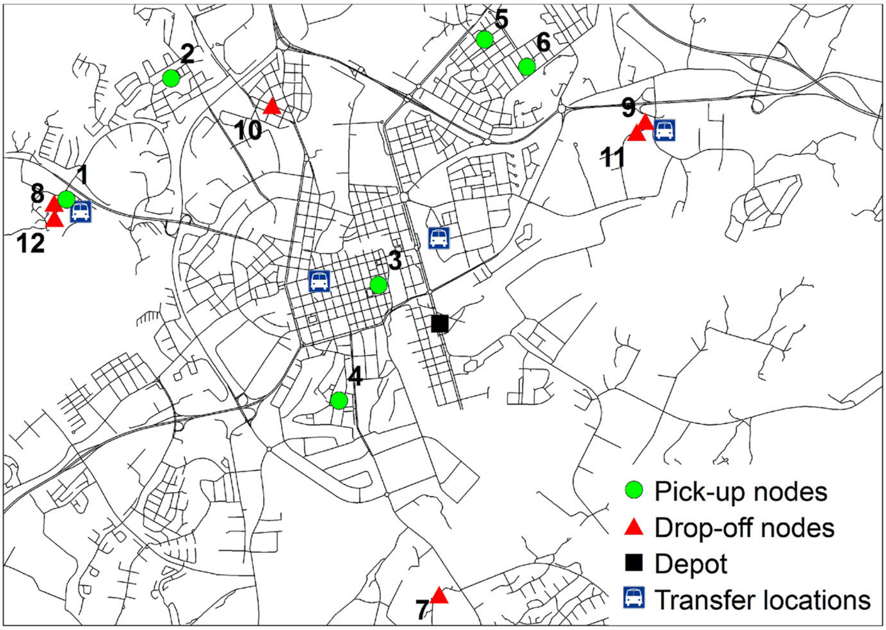

Fig. 4 The evaluation scenario with all included requests. The geography is from the town of Norrköping, Sweden

7.2 Computational results

Table 1 shows the computational time required to solve the evaluation scenario described in Sect. 7.1 to optimality.[1] Tests have been performed on both models and on both original and strengthened versions. Another comparison that can be made is the memory usage of the two models. This is summarized in Table 2.

The computational times quickly become very long, regardless of which model is used and whether or not it is strengthened. Note that the computational time and memory usage grows very quickly as a function of the number of included requests. Also, note that the memory needed for solving Model 2 is substantially less than the memory needed for Model 1.

Two values were used as cut-off points for breaking calculations before they have proved optimality of the solution. The calculations were interrupted if optimality was not proven when either (1) 48 h CPU time (~ 7 h wall-clock time) had passed, or (2) when the memory usage exceeded 24 GB. Both versions of Model 1 were aborted due to memory use before optimality was proven, for $\bar{r} = 4$ and 5, respectively. Both versions of Model 2 were aborted at $\bar{r} = 6$ due to the computational time cut-off.

[1] The computations were performed using CPLEX 12.5 on a virtual machine with four CPU's (Intel E5-2650) that have two cores each and 24 GB internal memory.

Table 1 Computational times (wall-clock time) needed to solve the evaluation scenario to optimality

Number of requests	Computational time			
	Model 1	Model 1—strengthened	Model 2	Model 2—strengthened
1	0.15 s	0.01 s	0.75 s	0.1 s
2	0.7 s	0.1 s	3.3 s	0.8 s
3	550 s	4 s	100 s	12 s
4	*	260 s	30 min	300 s
5		*	5 h	30 min
6			*	*

Cells with an asterisk were aborted before optimality could be proven

Table 2 The memory usage as a function of the number of included requests for both models and both original and strengthened versions

Number of requests	Memory usage (MB)			
	Model 1	Model 1—strengthened	Model 2	Model 2—strengthened
3	117	–	–	–
4	*	760	170	82
5		*	2000	345
6			*	*

Only cases where the memory usage exceeded 10 MB are included, this is indicated by a dash in the table. Cells with an asterisk were aborted before optimality could be proven

8 Conclusions and future research

In this paper we have introduced an extension of the integrated dial-a-ride problem that better reflects a real planning situation. The new version of the problem, called the integrated dial-a-ride problem with timetables (IDARP-TT) includes timetabled public transport, allows travel times to differ between different classes of demand responsive vehicles as well as the fixed route transport, and allows certain requests to begin or end their trips using the fixed route public transport. Two different arc-flow formulations which differ in the way they model transfer points are presented. Ways of strengthening the models through valid inequalities have also been presented. The two models are compared with respect to both the number of binary variables introduced and the computational time and memory usage needed to solve a specific evaluation scenario to optimality. Both models were proven to work as intended and provide the same solutions.

One conclusion that can be drawn from the comparison performed in this paper is that the choices made in the modeling stage can have a great impact on the difficulty of solving the problem. Both theoretical calculations and computational experiments indicate that Model 2 outperforms Model 1, but it is still clear that solving a MILP formulation of the IDARP-TT to optimality is extremely hard for real-worldsized instances. To solve larger problem instances a heuristic method will have to be developed. A heuristic solver is also necessary to be able to handle dynamic problem instances, which is a requisite feature in many real world applications.

One important branch of future research is the quality of service of demand responsive transport, such as e.g. Paquette et al. (2009) and Knutsson (2003). Recently, a literature review covering psychological, operational, and policy perspectives on commuters' willingness to use an integrated public transport system was presented in Chowdhury and Ceder (2016). To our knowledge, no studies have been performed on the level of service impacts of offering integrated trips to special transport service users.

Another important point is the extended understanding of the problem structure that the modeling presented in this paper has given. This can be of use in developing future solution methods. Future research could include studies of stronger valid inequalities, better suited branching techniques, and different modeling techniques.

Acknowledgements This research is part of a project financed by the Swedish Governmental Agency for Innovation Systems (VINNOVA). The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

Aldaihani M, Dessouky MM (2003) Hybrid scheduling methods for paratransit operations. Comput Ind Eng 45(1):75–96

Chowdhury S, Ceder A (2016) Users willingness to ride an integrated public-transport service: a literature review. Transp Policy 48:183–195

Cordeau JF (2006) A branch-and-cut algorithm for the dial-a-ride problem. Oper Res 54(3):573–586

Cordeau JF, Laporte G (2007) The dial-a-ride problem: models and algorithms. Ann Oper Res 153(1):29–46

Cortés CE, Matamala M, Contardo C (2010) The pickup and delivery problem with transfers: formulation and a branch-and-cut solution method. Eur J Oper Res 200(3):711–724

Crainic T, Malucelli F, Nonato M (2001) Flexible many-to-few + few-to-many = an almost personalized transit system. In: Proceedings of the triennial symposium on transportation analysis: TRISTAN IV, pp 435–440

Deleplanque S, Quilliot A (2013) Transfers in the on-demand transportation: the DARPT Dial-a-Ride Problem with transfers allowed. In: Multidisciplinary international scheduling conference: theory and applications (MISTA), pp 185–205

Edwards D, Elangovan A, Watkins K (2012) Reaching low-density urban areas with the network-inspired transportation system. In: 15th international IEEE conference on intelligent transportation systems (ITSC), pp 826–831

Edwards D, Trivedi A, Elangovan A, Dickerson S (2011) The network-inspired transportation system: A hierarchical approach to integrated transit. In: 14th IEEE conference on intelligent transportation systems (ITSC), pp 1507–1512

Ghilas V, Demir E, Woensel TV (2016) An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows and scheduled lines. Comput Oper Res 72:12–30

Häll CH, Lundgren JT, Värbrand P (2008) Evaluation of an integrated public transport system: a simulation approach. Arch Transp 20(1–2):29–46

Häll CH, Andersson H, Lundgren JT, Värbrand P (2009) The integrated dial-a-ride problem. Public Transp 1(1):39–54

Hickman M, Blume K (2001) Modeling cost and passenger level of service for integrated transit service. In: Voß S, Daduna JR (eds) Computer-aided scheduling of public transport, vol 505., Lecture Notes in Economics and Mathematical SystemsSpringer, Berlin, pp 233–251

Horn ME (2002) Multi-modal and demand-responsive passenger transport systems: a modelling framework with embedded control systems. Transp Res Part A Policy Pract 36(2):167–188

Horn MET (2004) Procedures for planning multi-leg journeys with fixed-route and demand-responsive passenger transport services. Transp Res Part C Emerg Technol 12(1):33–55

Knutsson S (2003) Valuing rider quality in Swedish special transport services—new findings. J Public Transp 6(3):4

Liaw CF, White C III, Bander J (1996) A decision support system for the bimodal dial-a-ride problem. IEEE Trans Syst Man Cybern Part A Syst Hum 26(5):552–565

Li H, Lim A (2003) A metaheuristic for the pickup and delivery problem with time windows. Int J Artif Intell Tools 12(02):173–186

Masson R, Lehuédé F, Péton O (2014) The dial-a-ride problem with transfers. Comput Oper Res 41:12–23

Paquette J, Cordeau JF, Laporte G (2009) Quality of service in dial-a-ride operations. Comput Ind Eng 56(4):1721–1734

Potter B (1976) Ann Arbor's dispatching system. Transportation Research Record (608)

Rais A, Alvelos F, Carvalho MS (2014) New mixed integer-programming model for the pickup-and-delivery problem with transshipment. Eur J Oper Res 235(3):530–539

Ronald N, Thompson R, Winter S (2015) Simulating demand-responsive transportation: a review of agent-based approaches. Transp Rev 35(4):404–421

Stålhane M, Andersson H, Christiansen M, Fagerholt K (2014) Vendor managed inventory in tramp shipping. Omega 47:60–72

The Swedish Bus and Coach Federation (2015) Statistik om bussbranschen 2015. Tech. rep, Stockholm

Toth P, Vigo D (2014) Vehicle routing: problems, methods, and applications, vol 18. SIAM, Philadelphia

Uchimura K, Takahashi H, Saitoh T (2002) Demand responsive services in hierarchical public transportation system. IEEE Trans Veh Technol 51(4):760–766

Wilson N, Weissberg R, Hauser J (1976) Advanced dial-a-ride algorithms research project: final report. Research report, Massachusetts Institute of Technology, Department of Materials Science and Engineering