DEPLOYING MOBILITY-ON-DEMAND FOR ALL BY OPTIMIZING PARATRANSIT SERVICES*

Sophie Pavia
Vanderbilt University
sophie.r.pavia@vanderbilt.edu

David Rogers
Vanderbilt University
david.rogers@vanderbilt.edu

Amutheezan Sivagnanam
The Pennsylvania State University
amutheezan@psu.edu

Michael Wilbur
Vanderbilt University
michael.p.wilbur@vanderbilt.edu

Danushka Edirimanna
Cornell University
ke233@cornell.edu

Youngseo Kim
Cornell University
yk796@cornell.edu

Philip Pugliese
CARTA
philippugliese@gocarta.org

Samitha Samaranayake
Cornell University
samitha@cornell.edu

Aron Laszka
The Pennsylvania State University
laszka.aron@gmail.com

Ayan Mukhopadhyay
Vanderbilt University
ayan.mukhopadhyay@vanderbilt.edu

Abhishek Dubey
Vanderbilt University
abhishek.dubey@vanderbilt.edu

ABSTRACT

While on-demand ride-sharing services have become popular in recent years, traditional on-demand transit services cannot be used by everyone, e.g., people who use wheelchairs. Paratransit services, operated by public transit agencies, are a critical infrastructure that offers door-to-door transportation assistance for individuals who face challenges in using standard transit routes. However, with declining ridership and mounting financial pressure, public transit agencies in the USA struggle to operate existing services. We collaborate with a public transit agency from the southern USA, highlight the specific nuances of paratransit optimization, and present a vehicle routing problem formulation for optimizing paratransit. We validate our approach using real-world data from the transit agency, present results from an actual pilot deployment of the proposed approach in the city, and show how the proposed approach comprehensively outperforms existing approaches used by the transit agency. To the best of our knowledge, this work presents one of the first examples of using open-source algorithmic approaches for paratransit optimization.

1 Introduction

There are more than 7,000 public transit agencies in the USA (and many more private agencies), and together, they are responsible for serving 60 billion passenger miles each year. A well-functioning public transit system fosters the growth and expansion of businesses, distributes social and economic benefits, and links the capabilities of community members, thereby enhancing what they can accomplish as a society [Beyazit, 2011, Harvey, 2010, Federal Highway Administration, 2003]. Transit infrastructure is especially important for low-income communities and individuals with disabilities (or short-term issues) as they often do not own or are unable to use private vehicles and must rely on public transit for connecting to employment opportunities, education, healthcare, and other essential services [Federal Highway Administration, 2003]. However, many transit agencies struggle to meet their mission due to decreasing ridership and increasing operational costs. Indeed, existing public transit infrastructure often shows stark inequities;

*Accepted in the AI and Social Good Track at the International Joint Conference on Artificial Intelligence (IJCAI), 2024

Figure 1: (left) The van that CARTA uses for paratransit operations. (right) A driver helping a passenger with a wheelchair board the paratransit van. Paratransit services are compliant with the Americans with Disabilities Act (ADA) and offer competitive transit services to the section of the population who cannot avail standard transit routes.

e.g., low-wage workers, people with disabilities, and the elderly have poorer access to transit in the USA [Stacy et al., 2020, DeGood, 2011].

A particularly critical component of public transit service in USA is the *paratransit service*, which offers door-to-door transportation assistance for individuals who face challenges in using standard transit routes (e.g., individuals with physical disabilities). Paratransit services are directly tied to the United Nations Sustainable Development Goal (UN-SDG) 11, which focuses on sustainable cities and communities; indeed, indicator 11.2.1 of UN-SDG 11 specifically seeks to measure the "*proportion of the population that has convenient access to public transport, by sex, age, and persons with disabilities*" [United Nations, 2020]. Despite its importance, transit agencies typically face operational challenges as there are often limited resources (vehicles, operational budget, and personnel), and paratransit requests are compliant with the regulations of the Americans with Disabilities Act (ADA), i.e., requested pickup and dropoff times must be adhered to (modulo *some* caveats, e.g., the difference between the pickup and dropoff time must be greater than the minimum time for traveling between the locations).[2]

We directly work with a public transit agency in the southern USA, CARTA, Chattanooga Area Regional Transportation Agency, that exemplifies the transit challenges faced by mid-sized southern cities, where agencies have to balance the dichotomy between improving service coverage, improving ridership, and lowering operational costs [Shrikant, 2018]. CARTA spends more than USD $1.1 million annually on fuel while supporting a number of different transportation modalities, including a fixed-route service, demand-response service (using neighborhood shuttles), and paratransit service (we show a snapshot of the paratransit service in Figure 1). CARTA provides over 3 million passenger trips per year through these three services. However, operations are inefficient regarding productivity and energy consumption per passenger per mile. In particular, **the paratransit operations account for 22% of total service miles but less than 4% of passenger trips**. We work with CARTA to improve the efficiency of their paratransit fleet.

From a technical standpoint, the operational challenge of optimizing transit fleets entails solving a set of complex mathematical optimization and planning problems. At the core of transit optimization lies a vehicle routing problem (VRP). VRPs can be divided into two major categories: *offline VRPs* consider how to optimally allocate vehicles to a set of requests that is known completely apriori, whereas *online VRPs*, process requests as they arrive in real-time. Paratransit service, in practice, deals with both these problems. Paratransit trips are usually scheduled at least a day beforehand, i.e., before each day, agencies can strategically plan paratransit vehicle routes based on requested pickup and drop-off spots and designated pickup time frames. A small percentage of the requests arrive in real-time, and operators seek to accommodate these requests as the vehicles move around the city. Second, to ensure that the algorithms translate to deployment in practice, we must develop production-quality software and visualization interfaces for transit operators and drivers. Third, and arguably most importantly, deploying artificial intelligence and data-driven solutions in practice requires active and sustainable collaborations with transit agencies. These problems of prediction, optimization, planning, and software development must be solved together to translate algorithmic development into practice.

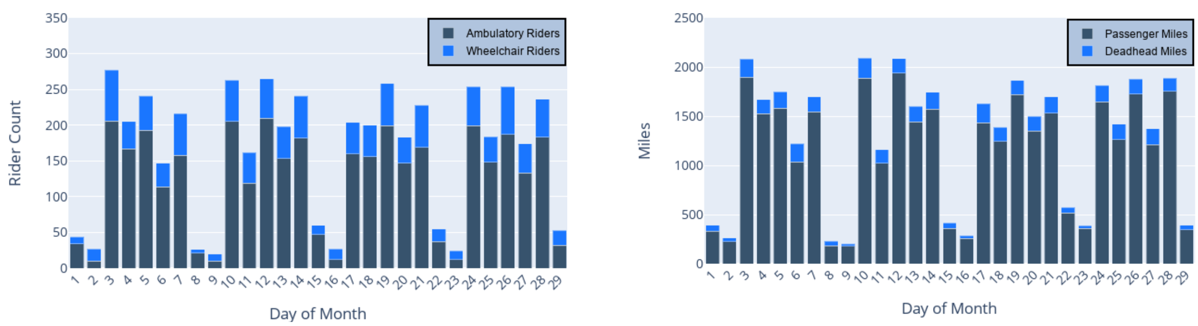
In this paper, we discuss how we develop a set of principled data-driven optimization modules for improving the efficiency of paratransit services in a mid-sized city in the southern USA. We describe our collaboration with CARTA in this paper , provide a background of the algorithmic approaches we developed, share an overview of the software

[2]Specifically, the ADA regulation states that transit agencies must provide service to "individuals with disabilities that is comparable to the level of service provided to individuals without disabilities who use the fixed route system."

tool for optimizing paratransit, present simulation results for validation on real-world data, and finally present results from the deployment of our toolchain in the city. We make the following contributions:

1. We describe the paratransit optimization problem from the perspective of a real-world transit agency and discuss both offline and online versions of the problem.
2. We describe how our formulation accommodates specific constraints faced by paratransit services (often not by traditional ride-sharing services). Then, we describe an offline and an online solution approach for solving the paratransit service problem. Our solution approaches are simple and can be implemented by using off-the-shelf VRP solvers, thereby facilitating easier deployment.
3. We present results from real-world data and a pilot study from a mid-sized southern city in the USA.

2 Background



(a) Paratransit ridership

(b) Paratransit deadhead miles

Figure 2: CARTA's paratransit metrics (a) Paratransit wheelchair and ambulatory riders per day in February 2020. (b) Deadhead and passenger miles per day in February 2020 for paratransit vehicles. Deadhead miles are defined as miles driven without passengers.

We begin by providing a brief overview of VRPs. Broadly, the VRP is a well-studied combinatorial optimization problem that deals with the optimal design of routes to be used by a fleet of vehicles to serve a set of customers [Dantzig and Ramser, 1959]. Such problems have been extensively modeled as mixed-integer linear programs (MILP), a class of optimization problems known to be NP-hard, i.e., they are computationally intractable to solve for large problem instances [Lenstra and Kan, 1981]. Often, VRPs are also modeled as planning problems that involve optimizing decisions over time under future uncertainty, e.g., optimizing the set of routes for a fleet of vehicles in real-time to serve a given set of customers while new customer orders arrive dynamically [Toth and Vigo, 2002, Pillac et al., 2013]. Such problems are often modeled as Markov decision processes (MDP) [Archetti et al., 2020, Wilbur et al., 2022a, Joe and Lau, 2020]. While both the ILP and the MDP settings for transportation optimization are computationally challenging to solve for large problem instances (e.g., optimizing city-wide transit), there are well-established heuristic approaches [Alonso-Mora et al., 2017, Kim et al., 2023, Wilbur et al., 2022a], that can solve these problems reasonably well (i.e., close to optimality). While these approaches can be used to model generic on-demand ride-sharing problems, the paratransit domain presents several unique challenges that we highlight below.

2.1 Paratransit Challenges

Paratransit services can resemble traditional on-demand ridepooling services in some ways (e.g., the arrival of real-time requests and ride-sharing); however, our collaboration with CARTA revealed several critical differences between traditional DVRPs and paratransit services. We point out some of these key differences below, which are especially important for deployment.

Variable capacity constraints: Similar to traditional on-demand ride-pooling services, paratransit services have constraints on pickup times, dropoff times, and vehicle capacities. However, unlike most applications (or models) of VRPs, paratransit services pose the unique challenge of varying capacities due to the presence of both ambulatory and wheelchair-bound passengers. A wheelchair requires more space than a regular seat in a paratransit vehicle. In this setting, most agencies have flexible seating-enabled vehicles whose seats can be folded to accommodate additional wheelchair passengers at the expense of ambulatory passengers. Wheelchair passengers can also take longer to board and depart the vehicle, which must be considered as this constraint might potentially cause the vehicle to violate a future pickup or dropoff constraint.

Travel time constraint: In the case of our partner agency (and many other public transit agencies that operate paratransit services), regulations dictate that paratransit trips must not take longer than the longest comparable fixed-line route.

Strict time windows: The Federal Transit Administration presents guidelines about the maximum wait times for paratransit passengers. Consequently, transit agencies must add substantial slack to their schedules and limit trips to only a few passengers each, leading to low efficiency.

Guaranteed service: Additionally, not all requests are equal. Requirements associated with the Americans with Disabilities Act (ADA) mandate that ADA-eligible clients *must* be serviced when requesting trips within a certain radius of public fixed-line services (with some caveats on the agency being given enough time to plan for the trip). We show a summary of paratransit ridership and miles traveled by vehicles for CARTA in Figure 2a and 2b, respectively.

3 Model

The core decision-making and computational problem that transit agencies must tackle for operating paratransit services is a VRP, i.e., the agency must assign routes to vehicles that pick up and drop off passengers at designated locations and times. As we highlighted before, paratransit requests can either be made (at least) a day in advance or in real-time. While agencies are not obligated to serve requests made in real-time, most public transit agencies, including CARTA, strive to accommodate these requests. As a result, transit agencies typically require two variants of VRPs—an offline VRP model (and solver) that handles the day-ahead requests and an online VRP model (and solver) for handling requests made in real-time. While both offline and online VRPs have been studied extensively, they have not been studied in the context of paratransit services, which pose domain-specific challenges. A notable exception is a prior work by Sivagnanam et al. [2022], who study offline VRPs (specifically, a VRPPDTW) with online bookings in the context of paratransit services [Sivagnanam et al., 2022]. Our offline model is closely related to theirs with a simplified booking process. Note that *our focus on leveraging prior work and not seeking to present a novel problem definition is rooted in our goal of deployability*. Crucially, as we highlight later, our model can be directly solved by off-the-shelf solvers.

We begin by introducing our model for the offline VRP, which builds upon the classical VRPPDTW model with time windows [Toth and Vigo, 2002]. We use the same notation as Toth and Vigo [2002] to allow readers to cross-reference between our formulation and the formulation provided in prior work [Toth and Vigo, 2002]. We use a *request-based* model for the VRPPDTW, which deals with n requests, each with its pickup and delivery locations (or nodes), and associated time windows. Let i be a request in the network made up of two nodes, i and $n+i$, which correspond to the request's pickup and dropoff locations, respectively. Naturally, in this formulation, different nodes can be simultaneously at the same geographical location. Let P be the set of pickup nodes, where $P = \{1, \ldots, n\}$ and let D be the set of drop-off nodes, where $D = \{n+1, \ldots, 2n\}$ Let the set of all nodes be N, i.e., $N = P \cup D$. Let the number of ambulatory passengers and passengers who use wheelchairs for request i be denoted by d_{a_i} and d_{w_i}, respectively. We use ℓ_{a_i} and ℓ_{w_i} denote the incremental capacity affected by request i; therefore, $\ell_{a_i} = d_{a_i}$, $\ell_{w_i} = d_{w_i}$ and thus $\ell_{a_{n+1}} = -d_{a_i}$, $\ell_{w_{n+i}} = -d_{w_i}$

Next, we define the set of vehicles as K. Where each vehicle has a set of nodes it can service, $N_k = P_k \cup D_k$, where N_k, P_k, and D_k are subsets of their respective sets. Each vehicle k has its own network where we let $G_k = (V_k, A_k)$ be a directed graph. We set $V_k = N + k \cup \{o(k), d(k)\}$ to contains the set nodes of for vehicle k, including the origin, $o(k)$, and destination $d(k)$, which are its depots. A_k is the subset of $V_k \times V_k$, which is made up of all feasible arcs. Each vehicle k has a capacity C_{a_k} for ambulatory passengers and a capacity C_{w_k} for wheelchair passengers. Each vehicle also has a travel time τ_{ijk} and cost c_{ijk} between unique nodes $i, j \in V_k$. We make the assumption that vehicle k leaves its origin depot without any passengers at time $a_{o(k)} = b_{o(k)}$, each vehicle must have an admissible route which a corresponding feasible path from $o(k)$ to $d(k)$ within the network G_k, no node in G_k can be visited more than once, and a vehicle must visit a chosen node, $i \in N$, within the time window $[a_i, b_i]$ when the service s_i must begin.

3.1 Solution Space

We extend a standard vehicle routing problem with pickup and delivery and time windows [Toth and Vigo, 2002] with additional capacity constraints. Our decision variables are as follows: binary flow variables $x_{ijk} \in \{0, 1\}$ that indicate if arc $(i, j) \in A_k$ is used by vehicle k, time variables $t_{ik} \in \mathbb{R}_{\geq 0}$ that indicate when vehicle k starts the service at node $i \in V_k$, and two load variables $y_{a_{ik}} \in \mathbb{R}_{\geq 0}$, for ambulatory passengers and $y_{w_{ik}} \in \mathbb{R}_{\geq 0}$ for wheelchair passengers. Both give the load of vehicle k once the service at node $i \in V_k$ has been completed. We minimize the total travel cost under the following linear objective function 1:

$$\min \sum_{k\in K} \sum_{(i,j)\in A_k} c_{ijk} x_{ijk} \tag{1}$$

Subject to the following sets of constraints M:

$$\begin{aligned}
&\sum_{k\in K} \sum_{j\in N_k \cup \{d(k)\}} x_{ijk} x_a = 1, && \forall i \in P && \text{(2a)}\\
&\sum_{j\in N_k} x_{ijk} - \sum_{j\in N_k} x_{j,n+i,k} = 0, && \forall k \in K, i \in P_k && \text{(2b)}\\
&\sum_{j\in P_k \cup \{d(k)\}} x_{o(k),jk} = 1, && \forall k \in K && \text{(2c)}\\
&\sum_{i\in N_k \cup \{o(k)\}} x_{ijk} - && &&\\
&\sum_{i\in N_k \cup \{d(k)\}} x_{jik} = 0, && \forall k \in K, j \in N_k && \text{(2d)}\\
&\sum_{i\in D_k \cup \{o(k)\}} x_{i,d(k),k} = 1, && \forall k \in K && \text{(2e)}\\
&x_{ijk}(t_{ik} + s_i + \tau_{ijk} - t_{jk}) \leq 0, && \forall k \in K, (i,j) \in A_k && \text{(2f)}\\
&a_i \leq t_{ik} \leq b_i, && \forall k \in K, i \in V_k && \text{(2g)}\\
&t_{ik} + \tau_{i,n+i,k} \leq t_{n+i,k}, && \forall k \in K, i \in P_k && \text{(2h)}\\
&x_{ijk}(y_{a_{ik}} + \ell_{a_j} - y_{a_{jk}}) = 0, && \forall k \in K, (i,j) \in A_k && \text{(2i)}\\
&\ell_{a_i} \leq y_{a_{ik}} \leq C_{a_k}, && \forall k \in K, i \in P_k && \text{(2j)}\\
&0 \leq y_{a_{n+i,k}} \leq C_{a_k} - \ell_{a_i}, && \forall k \in K, n+i \in D_k && \text{(2k)}\\
&x_{ijk}(y_{w_{ik}} + \ell_{w_j} - y_{w_{jk}}) = 0, && \forall k \in K, (i,j) \in A_k && \text{(2l)}\\
&\ell_{w_i} \leq y_{w_{ik}} \leq C_{w_k}, && \forall k \in K, i \in P_k && \text{(2m)}\\
&0 \leq y_{w_{n+i,k}} \leq C_{w_k} - \ell_{w_i}, && \forall k \in K, n+i \in D_k && \text{(2n)}\\
&y_{a_{o(k),k}} = 0, && \forall k \in K && \text{(2o)}\\
&y_{w_{o(k),k}} = 0, && \forall k \in K && \text{(2p)}\\
&x_{ijk} \in \{0,1\}, && \forall k \in K, (i,j) \in A_k &&\\
&t_{ik} \in \mathbb{R}_{\geq 0}, && \forall k \in K, i \in V_k &&\\
&y_{a_{ik}} \in \mathbb{R}_{\geq 0}, && \forall k \in K, i \in V_k &&\\
&y_{w_{ik}} \in \mathbb{R}_{\geq 0}, && \forall k \in K, i \in V_k &&
\end{aligned}$$

where constraints (2a) enforce that each request is served exactly once, and Constraints (2b) require that each request is served by the same vehicle. Constraints (2c-2e) ensure that vehicle k starts from its origin depot $o(k)$ and ends its route at its destination depot $d(k)$, forming a multicommodity flow structure. Constraints (2f) handle the compatibility requirements between routes and schedules, while constraints 2g implement the time windows for each vehicle and their nodes/stops. Constraints (2h) ensure that each request is picked up before it is dropped off. Constraints (2i and 2l) handle the compatibility requirements between routes and vehicle loads. Next, constraints (2j–2k and 2m–2n) enforce capacity intervals at pickup and delivery stops/nodes, for each vehicle. Lastly, Constraints (2o–2p) ensure that the starting load of each vehicle is 0. Note that under this implementation, a route's duration can only be at most $b_{d(k)} - a_{o(k)}$. A vehicle is also allowed to wait before it visits a node under Constraints (2f-2g). Finally, the arrival time at an arbitrary node j can be calculated as:

$$x_{ijk} = 1 \Rightarrow t_{jk} = \max\left(a_j, t_{ik} + s_i + \tau_{ijk}\right) \qquad (i,j) \in A_k$$

For the online VRPPDTW, the same set of constraints apply. However, the requests arrive one at a time and must be accommodated into existing schedules. A summary of the notation can be found in Table 1 below.

Symbol	Description
n	number of transportation requests;
N	the set of nodes;
P	the set of pick-up nodes;
D	the set of drop-off nodes;
ℓ_{a_i}	number of ambulatory passengers at node i;
ℓ_{w_i}	number of wheelchair passengers at node i;
C_{a_k}	ambulatory capacity of a vehicle;
C_{w_k}	wheelchair capacity of a vehicle;
K	the set of available vehicles;
τ_{ijk}	travel time between distinct nodes;
c_{ijk}	cost between distinct nodes;
s_i	service time at node i;

Table 1: Notation Table

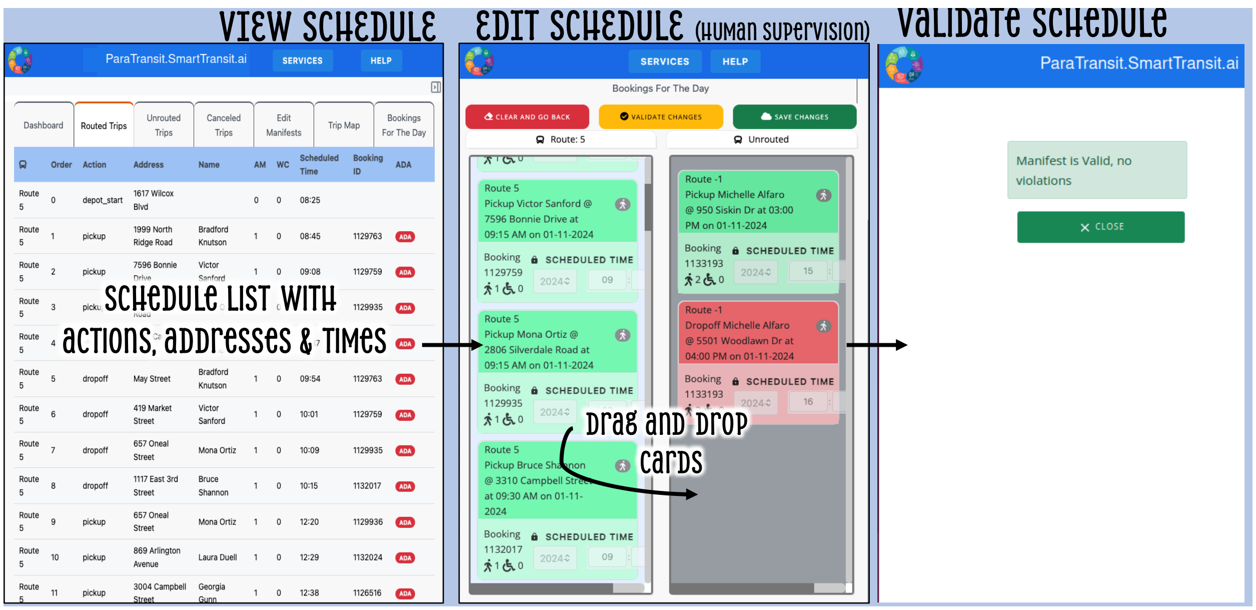


Figure 3: Supervisors can edit the schedule and override the machine-generated schedule.

3.2 Solution Approaches

To solve the offline VRP, we can directly use an off-the-shelf VRP solver. Specifically, we use the guided local search from Google OR-Tools VRPTW solver. For the online problem, we use an insertion heuristic [Wilbur et al., 2022b] that seeks to insert a request into an existing schedule without violating any constraints. While an insertion heuristic is not optimal (as it does not alter the order of existing requests that a vehicle has committed to), it often works fairly well in practice. To quantify the insertion quality, we choose the insertion position that minimizes the VMT/PMT ratio for the resulting vehicle schedule, where VMT and PMT refer to vehicle and passenger miles traveled, respectively. Specifically, VMT represents the total vehicle miles traveled by a vehicle over a day, including the miles when the passenger(s) are on board and when a vehicle travels without any passengers on board; PMT, on the other hand, is the summation of the shortest path between origin and destination for each OD pair in the dataset for a specific day. Therefore, PMT represents the vehicle miles required if each passenger drove directly between their corresponding origin and destination. Using VMT/PMT as the criteria for insertion penalizes longer vehicle miles traveled compared to the shortest route for each passenger.

3.2.1 Human in the Loop Solution

While algorithmic solutions can maximize desired objectives, e.g., service rate, such approaches often do not consider the domain knowledge known by the public transit operators. During our discussions with CARTA, we found that considering such domain knowledge is imperative for the smooth operation of real-world transit fleets. For example, a transit supervisor might know that a physically disabled passenger trusts a specific driver over many years and feels comfortable boarding a vehicle with a wheelchair with help from the said driver. Such tacit knowledge is often critical in practice, and we hypothesize that algorithmic solutions deployed for real-world transit fleets (and particularly paratransit services) must enable human supervisors to modify solutions generated by the algorithm. To enable such modification, we provide ways for CARTA to edit the generated solutions while also suggesting to them the feasibility of potential for constraint violations. We also allow them to override a solution if they so choose. Figure 3 shows the online screens of the deployment system that show the routed trips and the ability to switch the trips between a route and another bucket, which can be another route or the unrouted category. In our developed software system, edits require a simple drag and drop validated through a constraint checker that facilitates easy and seamless operation for the end user. For example, the online interface does not allow a client's drop-off card to be moved before the pick-up card. Also, the schedule validation generates warnings based on potential travel time violations or if the pick-up and drop-off times are outside the windows. Once saved, the schedule is communicated to the corresponding driver in real-time (if it is a same-day change).

4 Experiments

4.1 Dataset and Experimental Setup

We conduct our validation in two phases. First, we validate the algorithmic approaches using real-world paratransit data collected by CARTA. Then, we validate the proposed approach through a deployment exercise. Upon acceptance, we will release our code and a sample of simulated data.

4.1.1 Real-world paratransit data

We run experiments for all solvers with real-world paratransit data from CARTA from August 13th, 2023. The paratransit fleet consisted of 15 vehicles available with schedules staggered between morning and afternoon shifts according to CARTA's driver and vehicle availability and a total of 125 passengers. Each request in the dataset consisted of a geo-coordinate (longitude, latitude) for the pick-up location and drop-off location, as well as the requested pickup time. It also contained the capacity type for each request, with the option of ambulatory or wheelchair.

In addition to the set of requests, we also use CARTA's scheduled *manifest* and driver schedules for August 13th, 2023, where a manifest refers to an ordered list of locations that a vehicle will visit. Each entry in the manifest is marked with an action, pick-up or drop-off, and estimated arrival time. The manifest and driver schedules allow us to calculate CARTA's metrics for comparison against our solvers. CARTA uses a black-box software product to generate the manifests. We instantiate constraints for the VRP by using the same real-world constraint that CARTA follows: 1) the time window, i.e., the amount of time before he requested pickup time that a request can be picked up, is set to 15 minutes; 2) The vehicle capacity is set to maximum 8 ambulatory passengers and 2 wheelchair passengers (based on CARTA's paratransit vans); 3) the dwell time, i.e., the estimated duration that a vehicle will stay at a location to pick-up/drop-off a passenger is set to 5 minutes; and 4) any trip request that can not be serviced within the time windows is dropped, i.e., not considered.

Solver	VMT	VDM	VMT/PMT	Shared Rate	Passengers Served	Violations
Offline Solver	**990**	**252**	**1.03**	**85%**	125	0
PTA - Offline	1278	538	1.31	53%	125	12
Online Solver, 90%/10% split	1145	301	1.19	83%	125	0
Online Solver, 80%/20% split	1157	370	1.21	79%	125	0
Online Solver, 70%/30% split	1220	454	1.27	65%	125	0
Online Solver, 60%/40% split	1172	384	1.35	76%	115	0
Online Solver, 50%/50% split	1142	437	1.39	59%	110	0

Table 2: Solver Metrics recorded for paratransit data on compared with PTA's real schedule. VMT: Vehicle Miles Travelled, VDM: Vehicle Detour Miles, VMT/PMT: Vehicle Miles Travelled to Passenger Miles Travelled, Shared Rate: percentage of trip requests that shared their trip with another passenger, Violations: number of time window constraint violations. **We observe that the proposed offline solver serves all passengers without any violations, with significantly fewer detour miles and a significantly higher shared rate, thereby reducing the total number of miles the vehicles have to drive.**

4.1.2 Road network and travel time matrix

We use OpenStreetMap (OSM) [OpenStreetMap contributors, 2017] for the road network construction and OSMNX [Boeing, 2017] to generate a routing graph of the road network with travel time for edge weights. The travel time matrix is computed offline by calculating the shortest paths between all pairs in the network, with free flow speed as the edge weight.

4.2 Offline Solver

Given a set of requests for August 13, 2023, we generate a manifest using our offline VRP solver. The objective of the solver is to minimize vehicle miles traveled (VMT) with an additional large penalty term for dropping a trip request. Table 2 shows the manifest metrics for multiple solvers. We observe CARTA operations average a VMT/PMT ratio of 1.31; **the offline solver reduces VMT/PMT by 21.3% while servicing the same number of requests with the same vehicle configurations**. It is also important to note that the offline solver is more efficient with a higher shared rate than CARTA, meaning **85% of vehicle trips have more than one passenger on board** for a given time. CARTA's manifest also contained 12-time window violations, meaning passenger requests were scheduled to be served outside of the 15-minute time window constraint, while the proposed offline solver did not violate any time constraints.

4.3 Online Solver

To evaluate the online solver, we simulate real-time requests from the real-world data collected from CARTA. We randomly split the set of requests into two parts and treat one part as the set of day-ahead requests and another as the set of real-time requests. Our experiment design mimics CARTA's operational constraints, where a set of requests are available a day in advance, and at times, real-time requests must be accommodated while serving the day-ahead requests. For example, a 70-30 split denotes that 70% of requests are available a day in advance (to be solved by the offline solver), and 30% of requests arrive in real-time as same-day requests. In this case, our online solver is called for each real-time request as they arrive. The request arrival time, defined as the time before the requested pickup time that the request is available to the system, is set to 30 minutes.

We present the results of the online solver in Table 2. We observe that even when 70% of the requests are known in advance, the proposed approach outperforms CARTA's approach even if it knows all requests in advance. The performance gain is captured in lower VMT/PMT ratios, lower VMT, lower VDM, and a higher shared rate. We also observe that if a significant proportion of the trip requests are made in real-time (e.g., more than 40% of requests), then the proposed online solver fails to accommodate all requests while maintaining all paratransit constraints; e.g., in such a situation, 115 out of 125 passengers are served. In practice, the current operating conditions of CARTA resemble a 90-10 split, which the online solver can easily accommodate.

Pilot	Solver	VMT	VDM	VMT/PMT	Shared Rate	Passengers Served
Pilot 1	PTA	1531	601	1.41	61%	159
	Offline Optimizer	**1175**	**300**	**1.07**	**86%**	**159**
Pilot 2	PTA	1269	517	1.27	68%	129
	Offline Optimizer	**1061**	**281**	**1.06**	**84%**	**129**

Table 3: Pilot Metrics recorded for system testing for both deployments. VMT: Vehicle Miles Travelled, VDM: Vehicle Detour Miles, VMT/PMT: Vehicle Miles Travelled to Passenger Miles Travelled, Shared Rate: percentage of trip requests that shared their trip with another passenger. **We observe that the proposed offline solver serves all passengers without any violations, with significantly fewer detour miles and a significantly higher shared rate, thereby reducing the total number of miles the vehicles have to drive.**

5 Paratransit Pilot

Having validated the proposed approach in simulation by using real-world data from CARTA, we conducted a pilot exercise. Before describing the results, we highlight that conducting a pilot exercise that deploys an algorithmic approach is extremely difficult for public transit agencies, particularly in the context of paratransit service, as regulations mandate performance and service standards. We ran a pilot in August 2023 with CARTA's paratransit team with the offline approach. The goal of our pilot was the following: first, we aimed to evaluate the integration of our offline VRP optimizer with the paratransit service of CARTA. Second, we wanted to evaluate the algorithm during real-time operations. This goal involved equipping drivers with a tablet mounted in the vehicle that contained their schedule and

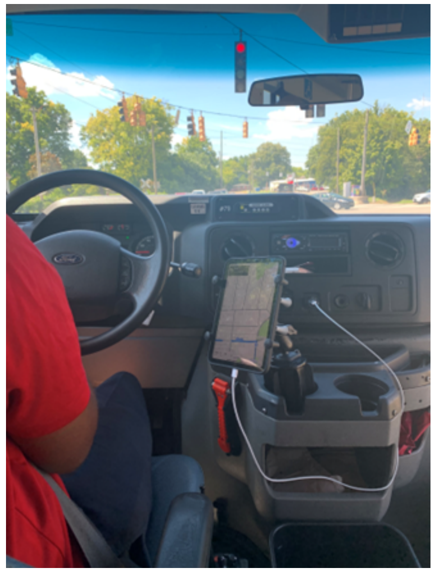


Figure 4: CARTA driver using the proposed algorithm to follow the manifest during a real-world deployment.

monitoring operations through a real-time web interface with members of CARTA's operations team (Fig. 4). Third, we wanted to gain feedback from schedulers and drivers on system usability and identify possible improvements going forward.

5.1 Pilot Design and Setup

We selected August 3, 2023, and August 10, 2023, for the pilot. We generated new manifests based on the proposed approach, which were actually deployed by CARTA. To compare the efficacy of the proposed approach, CARTA also generated manifests using their existing black-box approach. We exported the trip requests, driver schedules, vehicles, and scheduled manifests from CARTA's existing system. On both days, there were 15 vehicles available, with schedules staggered between morning and afternoon shifts according to CARTA's driver and vehicle availability. There were a total of 159 passengers and 129 passengers on August 3rd and 10th, respectively.

For the deployment, CARTA followed strict time window constraints for two types of passenger requests. CARTA mandated that *pickup-constrained* requests must be picked up within a 15-minute window before or after the requested pickup time, and the passenger must be dropped off within an hour of the requested pickup time at their destination. Dropoff-constrained requests represent situations where a passenger must be dropped off before the requested dropoff time and must be picked up no earlier than one hour before the appointment (e.g., these constraints often represent medical appointments in CARTA's data). Additionally, each vehicle had two capacity constraints—no more than 8 ambulatory passengers and 2 wheelchair passengers could be on a vehicle at a given time.

5.2 Evaluation of Offline VRP Solver

We present the key metrics related to the performance of CARTA's original schedule compared to the schedule generated by the offline VRP solver in Table 3. As shown, **the proposed approach reduced VMT by 356 miles on August 3, 2023, and by 236 miles on August 10, 2023**. There was a **24% and 17% improvement in VMT/PMT over CARTA's initial paratransit schedule** for August 3, 2023, and August 10, 2023, respectively. The efficiency gain correlates with the finding that our implementation had a much higher shared Rate, which is the percentage of passengers who shared their trip with at least one other passenger compared to CARTA's schedule.

5.3 Post-Deployment Feedback

After the pilot, we conducted a qualitative survey of drivers to evaluate their experience with the proposed system. Recurring themes from the feedback were, *User-Friendly*, *Efficient*, and *Easy to use* with one driver noting that "*It was great to have something like that to rely on when you need it [the application] instead of calling it [a problem] in to dispatch.*"

We also conducted interviews with paratransit operators at CARTA. When interviewing the operators, they explained that without the proposed offline optimizer, **it would take approximately six hours in a ten-hour day to generate a schedule for a day, compared to the proposed offline optimized generating manifests 'within minutes'**. When asked about positive attributes of the application, one operator said when it comes to tweaking schedules after generation, to account for domain knowledge and expertise, it took "*...about 50% less than what we would have to do, which is very good to me.*"

6 Conclusion

In collaboration with CARTA, we present data-driven optimization modules for paratransit services. We evaluate the offline and online algorithms on real-world paratransit data. The successful validation of the proposed approaches resulted in a pilot deployment, which showed that our approach comprehensively outperforms existing approaches followed by CARTA. Our software module and the model can be configured easily with agency-specific constraints and allow for transit operator intervention with violation checking. Crucially, our technological intervention for CARTA only amplifies the existing paratransit initiative rather than creating a new one, thereby resulting in a framework that is more likely to reach and maintain deployment [Toyama, 2015].

Acknowledgements

This material is based on work sponsored by the National Science Foundation under grant CNS-1952011 and the Department of Energy under Award Number DE-EE0009212.

References

Eda Beyazit. Evaluating social justice in transport: lessons to be learned from the capability approach. *Transport Reviews*, 31(1):117–134, 2011.

David Harvey. *Social justice and the city*, volume 1. University of Georgia press, 2010.

Federal Highway Administration. Status of the nation's highways, bridges, and transit: 2002 conditions and performance report. https://www.fhwa.dot.gov/policy/2002cpr/, 2003.

Christina Plerhoples Stacy, Yipeng Su, Eleanor Noble, Alena Stern, Kristin Blagg, Macy Rainer, and Richard Ezike. Access to opportunity through equitable transportation: Lessons from four metropolitan regions. Technical report, Urban Institute, 2020.

Kevin DeGood. Aging in place, stuck without options: Fixing the mobility crisis threatening the baby boom generation. Technical report, Transportation for America, 2011.

United Nations. Sustainable development goals. https://www.un.org/sustainabledevelopment/health/, 2020.

Aditi Shrikant. Why US public transportation is so bad — and why Americans don't care. https://www.vox.com/the-goods/2018/9/26/17903146/mass-transit-public-transit-rail-subway-bus-car, 2018.

G. B. Dantzig and J. H. Ramser. The truck dispatching problem. *Management Science*, 6(1):80–91, 1959. doi: 10.1287/mnsc.6.1.80. URL https://doi.org/10.1287/mnsc.6.1.80.

Jan K Lenstra and AHG Rinnooy Kan. Complexity of vehicle routing and scheduling problems. *Networks*, 11(2): 221–227, 1981.

Paolo Toth and Daniele Vigo. *The vehicle routing problem*. Society for Industrial and Applied Mathematics, 2002.

Victor Pillac, Michel Gendreau, Christelle Guéret, and Andrés L Medaglia. A review of dynamic vehicle routing problems. *European Journal of Operational Research*, 225(1):1–11, 2013.

Claudia Archetti, Dominique Feillet, Andrea Mor, and Maria Grazia Speranza. Dynamic traveling salesman problem with stochastic release dates. *European Journal of Operational Research*, 280(3):832–844, 2020.

Michael Wilbur, Salah Kadir, Youngseo Kim, Geoffrey Pettet, Ayan Mukhopadhyay, Philip Pugliese, Samitha Samaranayake, Aron Laszka, and Abhishek Dubey. An online approach to solve the dynamic vehicle routing problem with stochastic trip requests for paratransit services. In *ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, April 2022a.

Waldy Joe and Hoong Chuin Lau. Deep reinforcement learning approach to solve dynamic vehicle routing problem with stochastic customers. In *Proceedings of the international Conference on Automated Planning and Scheduling*, volume 30, pages 394–402, 2020.

Javier Alonso-Mora, Samitha Samaranayake, Alex Wallar, Emilio Frazzoli, and Daniela Rus. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 114 (3):462–467, 2017.

Youngseo Kim, Danushka Edirimanna, Michael Wilbur, Philip Pugliese, Aron Laszka, Abhishek Dubey, and Samitha Samaranayake. Rolling horizon based temporal decomposition for the offline pickup and delivery problem with time windows. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI-23)*, 2023.

Amutheezan Sivagnanam, Salah Uddin Kadir, Ayan Mukhopadhyay, Philip Pugliese, Abhishek Dubey, Samitha Samaranayake, and Aron Laszka. Offline vehicle routing problem with online bookings: A novel problem formulation with applications to paratransit. In *31st International Joint Conference on Artificial Intelligence (IJCAI)*, July 2022.

Michael Wilbur, Salah Uddin Kadir, Youngseo Kim, Geoffrey Pettet, Ayan Mukhopadhyay, Philip Pugliese, Samitha Samaranayake, Aron Laszka, and Abhishek Dubey. An online approach to solve the dynamic vehicle routing problem with stochastic trip requests for paratransit services. In *2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)*, pages 147–158. IEEE, 2022b.

OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2017.

Geoff Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126–139, 2017.

Kentaro Toyama. *Geek Heresy: Rescuing Social Change from the Cult of Technology*. PublicAffairs, 2015.