MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

The Integrated Last-Mile Transportation Problem ILMTP

Raghunathan, A.U.; Bergman, D.; Hooker, J.; Serra, T.; Kobori, S.

TR2018-084 July 13, 2018

Abstract

Last-mile transportation (LMT) refers to any service that moves passengers from a hub of mass transportation (MT), such as air, boat, bus, or train, to destinations, such as a home or an office. In this paper, we introduce the problem of scheduling passengers jointly on MT and LMT services, with passengers sharing a car, van, or autonomous pod of limited capacity for LMT. Passenger itineraries are determined so as to minimize total transit time for all passengers, with each passenger arriving at the destination within a specified time window. The transit time includes the time spent traveling through both services and, possibly, waiting time for transferring between the services. We provide an integer linear programming (ILP) formulation for this problem. Since the ILMTP is NP-hard and problem instances of practical size are often difficult to solve, we study a restricted version where MT trips are uniform, all passengers have time windows of a common size, and LMT vehicles visit one destination per trip. We prove that there is an optimal solution that sorts and groups passengers by their deadlines and, based on this result, we propose a constructive grouping heuristic and local search operators to generate high-quality solutions. The resulting groups are optimally scheduled in a few seconds using another ILP formulation. Numerical results indicate that the solutions obtained by this heuristic are often close to optimal and that warm-starting the ILP solver with such solutions decreases the overall computational times significantly.

International Conference on Automated Planning and Scheduling (ICAPS)

The Integrated Last-Mile Transportation Problem (ILMTP)

Arvind U. Raghunathan
Mitsubishi Electric Research Labs,
Cambridge, MA, USA
raghunathan@merl.com

David Bergman
University of Connecticut,
Storrs, CT, USA
david.bergman@uconn.edu

John Hooker and **Thiago Serra**
Carnegie Mellon University,
Pittsburgh, PA, USA
{jh38,tserraaz}@andrew.cmu.edu

Shingo Kobori
Advanced Technology Center,
Mitsubishi Electric Corporation,
Hyogo, Japan

Abstract

Last-mile transportation (LMT) refers to any service that moves passengers from a hub of *mass transportation* (MT), such as air, boat, bus, or train, to destinations, such as a home or an office. In this paper, we introduce the problem of scheduling passengers jointly on MT and LMT services, with passengers sharing a car, van, or autonomous pod of limited capacity for LMT. Passenger itineraries are determined so as to minimize total transit time for all passengers, with each passenger arriving at the destination within a specified time window. The transit time includes the time spent traveling through both services and, possibly, waiting time for transferring between the services. We provide an integer linear programming (ILP) formulation for this problem. Since the ILMTP is NP-hard and problem instances of practical size are often difficult to solve, we study a restricted version where MT trips are uniform, all passengers have time windows of a common size, and LMT vehicles visit one destination per trip. We prove that there is an optimal solution that sorts and groups passengers by their deadlines and, based on this result, we propose a constructive grouping heuristic and local search operators to generate high-quality solutions. The resulting groups are optimally scheduled in a few seconds using another ILP formulation. Numerical results indicate that the solutions obtained by this heuristic are often close to optimal and that warm-starting the ILP solver with such solutions decreases the overall computational times significantly.

Introduction

Last-mile transportation (LMT) is defined as the service that delivers people from the hub of a mass transportation (MT) service to each passenger's final destination. The MT service can be one of air, boat, bus, or train. The LMT service can be facilitated by bike (Liu, Jiang, and Cheng 2012), car (Shaheen 2004; Thien 2013), autonomous pods (Shen, Zhang, and Zhao 2017) (see also http://www.ukautodrive.com/, http://www.innovaevcarshare.com/), or personal rapid transit systems. Though the term LMT has also been used for the movement of goods in supply chains, home-delivery systems, and telecommunications, we will restrict our attention in this paper exclusively to the transportation of people. A LMT service expands the access of MT to an area wider than that defined as "walking distance" of a transportation hub. Interest in the design and operation of LMT has grown tremendously in the past decade. This has been driven primarily by three factors (Wang 2017): (i) governmental push to reduce congestion and air pollution; (ii) increasing aging population in cities; and (iii) providing mobility for the differently abled and school children.

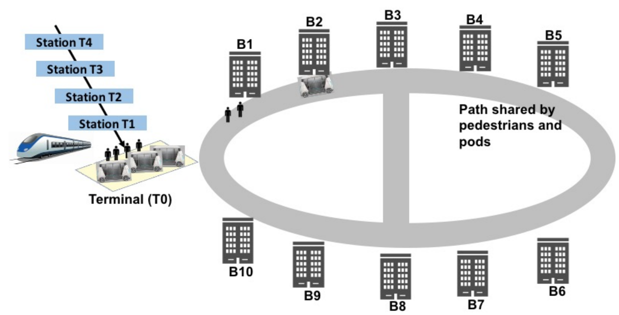

Figure 1: Schematic of the integrated LMT system (images licensed from shutterstock.com).

Figure 1 shows a typical scenario for the operation of the MT service in conjunction with the LMT. All passengers of the LMT service start their journey on the MT from one of the stations served by a train, and request automated transportation to the buildings within a time-window. The buildings B1-B10 can represent offices for different companies that are co-located in an industrial park, or residential buildings in a neighborhood. The buildings can be accessed by paths that are shared by both pedestrians and LMT vehicles. For convenience, we will refer to the vehicle providing LMT service as a *commuter vehicle* (CV). The CVs are typically parked at a terminal (T0) at which passengers arrive from trains and proceed to their respective destination buildings by sharing a ride in a CV. The LMT service may represent the morning commute to the office or the evening commute back to residences. Once all the passengers are delivered to their destinations, the CVs return back to the terminal for subsequent trips.

We envision an operational scenario such as in Figure 2, whereby the passengers indicate their origin station, destination building, and the desired time-window of arrival at the destination. This information is assumed to be available to the scheduler well in advance of scheduling decisions.

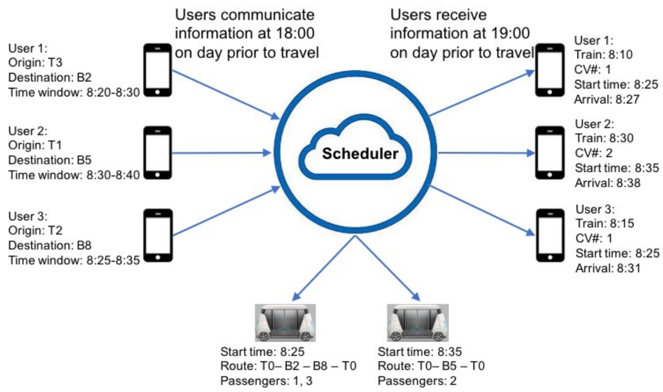

Figure 2: Schematic representation of the interaction between the passengers in the system, the scheduler and the CVs (images licensed from `shutterstock.com`)

For instance, consider the situation where the buildings represent offices and the passengers enter requests through a smartphone app on the evening of the previous day. Once all requests have been received, the scheduler determines for each user: (i) the train to board at their origin station; (ii) the CV to board at `T0`, (iii) the time the CV will depart from `T0`, and (iv) the time of arrival at the destination building. The scheduler also communicates to the different CVs the routes, start times, and list of passengers. The choice of route determines the times of arrival of passengers at their destination and the total time spent by the passengers in the CV.

Note that the first-mile operation, wherein the passengers first ride on CVs to reach a hub of a MT service, can be easily accommodated in an analogous manner. Furthermore, the requests need not be provided in advance and can instead be communicated to the scheduler in a *just-in-time* manner, but we do not address such online variants in this work.

In this paper, we introduce the Integrated Last-Mile Transportation Problem (ILMTP). The ILMTP is defined as the problem of scheduling passengers jointly on the MT and LMT services so that the passengers reach their destinations within specified time-windows and the total transit time for all passengers is minimized. We assume that the information about passengers is available *in advance*. The transit time includes the time spent traveling in the transportation vehicles in addition to any time spent waiting for the LMT service. In determining the schedules on the LMT service, the problem also determines the set of groups that share a ride in a CV. Thus, the time spent by the passengers in the CV depends on the co-passengers. To the best of our knowledge, this general version has not been addressed in the literature.

We introduce a constructive heuristic and a local search method to group passengers for LMT trips, which based on this grouping are then optimally scheduled in a few seconds by an Integer Linear Programming (ILP) solver. These groups preserve the invariant that each group has a single destination and that the deadlines of the passengers across groups going to the same destination can be sorted in nondecreasing order. We show that there is always an optimal solution with such a structure if all passengers have time windows of the same size, all MT trips serve all stations with uniform trip times, and each LMT trip has a single destination. We present computational results restricted to solutions of this form and compare them with a general lower bound.

Related Work

The ILMTP can be broadly viewed as an instance of routing and scheduling with time-windows. We survey the relevant literature and describe the key differences with the problem studied in this paper.

Last-Mile Transportation Problem

The literature on last-mile transportation has been mostly focused on the LMT service, without much consideration to the MT system. Seminal work in this area dates back to the 1960s and has focused mostly on freight transportation (see Wang (2017) for a discussion). For passenger transportation, a number of case studies have analyzed the last-mile problem in different contexts, such as a bicycle-sharing program in Beijing (Liu, Jiang, and Cheng 2012). Wang (2017) is the first work to consider routing and scheduling in the LMT service. The paper also considered the minimization of total travel time and proposed a heuristic approach for constructing solutions. More recently, Mahéo, Kilby, and Hentenryck (2018) consider the design of a public transit system that includes multiple modes of transportation, however the authors did not consider scheduling aspects. The ILMTP is a strict generalization of Wang (2017), in that we consider time-windows for arrival and scheduling on the MT service. Furthermore, it also complements the work of Mahéo, Kilby, and Hentenryck (2018) by focusing exclusively on the scheduling aspects of multi-modal transportation.

Personal Rapid Transit

Personal Rapid Transit (PRT) has similarities to the last-mile problem and has attracted significant attention in the past decade. Research has been conducted on PRT system control frameworks (Anderson 1998), financial assessments (Bly and Teychenne 2005; Berger et al. 2011), performance approximations (Lees-Miller, Hammersley, and Davenport 2009; Lees-Miller, Hammersley, and Wilson 2010) and case studies (Mueller and Sgouridis 2011). However, none of these papers have addressed last-mile operational issues.

Demand Responsive Transit

A large body of research has been devoted to demand responsive transit (DRT), which is another type of on-demand service. Some papers focus on DRT concept discussions, practical implementation, and assessment of simulations in case studies (Brake, Nelson, and Wright 2004; Horn 2002b; Mageean and Nelson 2003; Palmer, Dessouky, and Abdelmaguid 2004; Quadrifoglio, Dessouky, and Ordóñez 2008). Models have been developed to assist in system design and regulation (e.g., Daganzo 1978, Diana, Dessouky, and Xia 2006, Wilson and Hendrickson 1980). Routing options in specific contexts have also been considered (Chevrier, Jourdan, and Dhaenens 2012; Horn 2002a)). The LMT can be viewed as a specific variant of a broadly defined DRT concept—namely, a demand responsive transportation system that addresses last-mile service requests with batch passenger demand and a shared passenger origin. Unlike most papers in the DRT literature, we also focus on routing and scheduling on the MT and LMT from an optimization and operational perspective.

Vehicle Routing & Dial-a-Ride Problems

Vehicle routing problems have long been studied, and they comprise a large body of literature. The vehicle routing problem with time windows (VRPTW) has been the subject of intensive study, with many heuristic and exact optimization approaches suggested in the literature. A thorough review of the VRPTW literature can be found in (Toth and Vigo 2014). The dial-a-ride problem (DARP) and related variations have also been extensively investigated (Cordeau and Laporte 2007; Jaw et al. 1986; Lei, Laporte, and Guo 2012). As argued by (Wang 2017), the VRPTW focuses on reducing operating costs while ILMTP aims to improve the level-of-service by minimizing total passenger transit time.

In summary, ILMTP has the following features that distinguishes it from previous studies in the literature: (1) joint scheduling of passengers on the MT and LMT services; (2) time-windows on arrival at destination; (3) common last-mile origin (which is also the vehicle depot); and (4) minimization of the total passenger transit time.

Problem Formulation

In this section we introduce notation relevant to ILMTP and present an optimization model. For ease of exposition, we will assume for the rest of the paper that the MT corresponds to trains that arrive at $\texttt{T0}$. The problem data, summarized in Table 1, can be divided into those associated with: (i) the MT service, (ii) the LMT service, and (iii) the passengers. The set $\mathcal{U}(s)$ represents the set of all possible MT service options that are available for passengers from station s to the *terminal* $\texttt{T0}$. The MT services that are identical in travel time to reach $\texttt{T0}$ but leave at different times are treated as distinct MT service options in $\mathcal{U}$. The set $\mathcal{U}(s)$ includes the possibility of passengers transferring between trains in order to travel from s to $\texttt{T0}$. A user that leaves from s on a MT service $u \in \mathcal{U}(s)$ has a travel time of $(t^{\texttt{T0}}(s) - t^{\text{start}}(u, s))$ on the MT service to reach $\texttt{T0}$. If a user does not leave on the CV until time t, then the user spends $(t - t^{\texttt{T0}}(u, s))$ time waiting at $\texttt{T0}$. For commuting in the CVs, the travel time for the passengers depends on: (i) the route choice on the CV, and (ii) the number of stops on the route prior to their destination. Without loss of generality, we abstract loading and unloading times as part of the pre-computed routes, since routes with stops that are not actually used are suboptimal.

The decision variables in the formulation are:

- $y_{p,u} \in \{0, 1\}$: is 1 if passenger p is assigned to train service $u \in \mathcal{U}(s(p))$;
- $z_{p,r,t} \in \{0, 1\}$: is 1 if passenger p is assigned to CV route $r \in \mathcal{R}$ (with $b(p) \in \mathcal{B}(r)$) and departs at time $t \in \mathcal{T}$;

$\mathcal{T}$	planning horizon for the problem
Mass Transportation (MT) Service	
$\mathcal{S}$	set of boarding stations for passengers
$\texttt{T0}$	the alighting station for passengers
$\mathcal{U}(s)$	the set of all MT services that serve s
$t^{\text{start}}(u, s)$	time that MT service $u \in \mathcal{U}(s)$ leaves s
$t^{\texttt{T0}}(u)$	time that MT service $u \in \mathcal{U}(s)$ reaches $\texttt{T0}$
Passengers	
$\mathcal{P}$	set of passengers
$s(p)$	origin station of $p \in \mathcal{P}$
$b(p)$	destination of $p \in \mathcal{P}$
$\texttt{rel}(p)$	earliest time of arrival for $p \in \mathcal{P}$
$\texttt{ded}(p)$	latest time of arrival for $p \in \mathcal{P}$
Last-Mile Transportation (LMT) Service	
$\mathcal{C}$	set of commuter vehicles
$CV^{\max}$	capacity of CV
$\mathcal{R}$	the set of routes for the CV
$\mathcal{B}(r)$	the set of stops served on LMT route $r \in \mathcal{R}$
$\tau^{\text{travel}}(r, b)$	travel time on CV to reach $b \in \mathcal{B}(r)$ from $\texttt{T0}$ on route r
$\tau^{\texttt{trip}}(r)$	round trip time for a CV to return to $\texttt{T0}$ on route r

Table 1: Problem data for the ILMTP.

- $n_{\texttt{T0},t}$: number of CVs at $\texttt{T0}$ at time $t \in \mathcal{T}$;
- $n_{r,t} \in \{1, \ldots, |\mathcal{C}|\}$: number of CVs that are assigned to route r and depart at time $t \in \mathcal{T}$; and
- τ_p: transit time for passenger $p \in \mathcal{P}$.

We first describe the constraints in the problem that model operational requirements. The total transit time for each passenger $p \in \mathcal{P}$ is

$$\begin{aligned} \tau_p = &\sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} (t + \tau^{\text{travel}}(r, b(p))) \cdot z_{p,r,t} \\ &- \sum_{u \in \mathcal{U}(s(p))} t^{\text{start}}(u) \cdot y_{p,u}. \end{aligned} \tag{1a}$$

Next we model that each passenger $p \in \mathcal{P}$ is assigned to exactly one u on the MT service and has an unique CV route r and start time t:

$$\sum_{u \in \mathcal{U}(s(p))} y_{p,u} = 1, \ \sum_{r \in \mathcal{R}} \sum_{p \in \mathcal{P}} z_{p,r,t} = 1. \tag{1b}$$

The following constraints ensure that each passenger starts on a CV trip after arriving to $\texttt{T0}$:

$$\sum_{t \leq t^{\texttt{T0}}(u)} z_{p,r,t} \leq 1 - y_{p,u} \, \forall \, p \in \mathcal{P}, u \in \mathcal{U}(s(p)). \tag{1c}$$

The time-windows on the arrivals are imposed using

$$\texttt{rel}(p) \leq \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} (t + \tau^{\text{travel}}(r, b(p)) \cdot z_{p,r,t} \leq \texttt{ded}(p). \tag{1d}$$

The availability of CVs at $\texttt{T0}$ for transporting passengers are modeled using time-difference equations in (1e). This formulation was presented by Wang (2017) and is frequently used in modeling cumulative scheduling problems.

$$\begin{aligned} n_{\texttt{T0},t} = &n_{\texttt{T0},t-1} + \sum_{r \in \mathcal{R}} n_{r,t-\tau^{\texttt{trip}}(r)} - \sum_{r \in \mathcal{R}} n_{r,t} \, \forall \, t \in \mathcal{T} \\ n_{\texttt{T0},0} = &|\mathcal{C}|. \end{aligned} \tag{1e}$$

Constraint (1e) states that the number of CVs in T0 at t is the sum of three components: number of CVs in T0 at time $t-1$; number of CVs returning to T0 upon completion of trips started at time $t-\tau^{\texttt{trip}}(r)$ for each route r; and the negative of the number of CVs leaving T0 on trips at time t.

Finally, the capacity of the CVs are modeled using

$$\left.\begin{aligned} \sum_{p\in\mathcal{P}} z_{p,r,t} &\le CV^{\max}\cdot n_{r,t} \\ \sum_{p\in\mathcal{P}} z_{p,r,t} &\ge CV^{\max}\cdot (n_{r,t}-1) \end{aligned}\right\} \forall\, r\in\mathcal{R}, t\in\mathcal{T}. \tag{1f}$$

An optimization formulation modeling ILMTP is thus

$$\min \sum_{p\in\mathcal{P}} \tau_p \text{ s.t. (1a) -- (1f).} \tag{2}$$

To the best of our knowledge, this is the first optimization formulation for the ILMTP. The ILMTP is a generalization of the problem considered in (Wang 2017), which was shown to be NP-Hard. Hence, the ILMTP is NP-Hard.

Theoretical Results

In this section, we analyze the structure of optimal solutions to the ILMTP under the following assumptions:

Assumption 1. *The MT services in $u\in\mathcal{U}$ are identical and serve all boarding stations; i.e. $\mathcal{U}(s)=\mathcal{U}$ for all $s\in\mathcal{S}$ and $t^{T0}(u)-t^{\text{start}}(u,s) = t^{T0}(u')-t^{\text{start}}(u',s)$ for all $u,u'\in\mathcal{U}(s)$ and $s\in\mathcal{S}$.*

Assumption 2. *Each LMT trip serves a single destination.*

Assumption 3. *For all pairs of passengers $p_1,p_2\in\mathcal{P}$, $\texttt{ded}(p_1)-\texttt{rel}(p_1)=\texttt{ded}(p_2)-\texttt{rel}(p_2)$.*

Although Assumptions 1-3 are restrictive, the analysis in this section shows that there exists an optimal solution satisfying a particular ordering property. We exploit this in devising an heuristic, which we show through experimental evaluation is effective at identifying high-quality solutions.

Prior to presenting the technical results, additional notation is in order. We define a *group* to be a set of passengers that ride together in a CV on their LMT trip. Thus, a group consists of passengers having: the same start time on a common CV; cardinality less than the capacity of the CV; and a common destination (by Assumption 2). For any solution, let $G_1,\ldots,G_{\mathcal{L}}$ denote the partitioning of passengers into groups. Let $G(p_j)$ be the group that contains passengers p_j. We assume, without loss of generality, that the passengers are ordered as $p_1,\ldots,p_{|\mathcal{P}|}$ so that $i<j \iff \texttt{rel}(p_i)\le\texttt{rel}(p_j)$. Denote by $\mathrm{CVT}(G_\ell)$ the time that the CV for group G_ℓ reaches the destination building.

We begin the analysis with some preliminary results.

Lemma 1. *All passengers in a group have the same wait time between services in an optimal solution to the ILMTP.*

Proof. Due to Assumption 1, in any optimal solution, all passengers arrive with the latest MT service that arrives at T0 prior to their departure time on the CV. Hence, their wait times are the same. □

Following Lemma 1, let $\mathrm{TT0}(G_\ell)$ denote the time that G_ℓ reaches T0 on the MT in an optimal solution to ILMTP.

Lemma 2. *Let $G_1,\ldots,G_{\mathcal{L}}$ be the groupings in an optimal solution to an ILMTP instance. If passengers p_1,p_2 are such that $b(p_1)=b(p_2)$ and $G(p_1)\neq G(p_2)$, then exchanging passengers p_1 and p_2 between groups $G(p_1)$ and $G(p_2)$ (holding all else equal) does not affect the solution value.*

Proof. For each passenger p, let τ_p be the original total transit time and τ'_p be the resulting total transit time for that passenger. The only change to the objective function is the change in the total transit times for these two passengers. In particular, the resulting transit times for the passengers are $\tau'_{p_1}=\tau_{p_1}+\Delta_2-\Delta_1$ and $\tau'_{p_2}=\tau_{p_2}+\Delta_1-\Delta_2$ where $\Delta_i=\mathrm{CVT}(G(p_i))-\mathrm{TT0}(G(p_i))$. The net change in the objective function value is $\left(\tau'_{p_1}+\tau'_{p_2}\right)-\left(\tau_{p_1}+\tau_{p_2}\right)=0$. □

Lemma 2 states that we can exchange passengers between groups that are going to the same destination without affecting the objective. Note that Lemma 2 does not make any claim on the feasibility of the groupings after the exchange.

Lemma 3. *For any optimal solution and any group G_ℓ, let $p_{i_1},p_{i_2}\in G_\ell$, for $1\le i_1<i_2\le|\mathcal{P}|$. If $|G_\ell|<CV^{\max}$, then for any i with $i_1<i<i_2$, moving passenger p_i into group G_ℓ preserves feasibility.*

Proof. We need only show that $\texttt{rel}(p_i)\le\mathrm{CVT}(G_\ell)\le\texttt{ded}(p_i)$. The first inequality follows because $\texttt{rel}(p_i)\le\texttt{rel}(p_{i_2})$, by the assumption on the ordering of the passengers, and $\texttt{rel}(p_{i_2})\le\mathrm{CVT}(G_\ell)$, by the feasibility of the original solution. The second inequality follows because $\mathrm{CVT}(G_\ell)\le\texttt{ded}(p_{i_1})$, by the feasibility of the solution, and $\texttt{ded}(p_{i_1})\le\texttt{ded}(p_i)$, by the assumption on the ordering of the passengers, together with Assumption 3. □

We now state and prove, in Theorem 1, the main result of the section, which shows that an optimal solution exists where the passengers in a group have consecutive deadlines.

Theorem 1. *Suppose the passengers $\{p_1,\ldots,p_n\}$ have identical destinations and are ordered so that $i\le j\iff\texttt{rel}(p_i)\le\texttt{rel}(p_j)$. For any ILMTP instance there is an optimal solution with groupings $G_1,\ldots,G_{\mathcal{L}}$ for which if $p_j,p_{j+k}\in G_\ell$, for some $j\in\{1,\ldots,n-2\}$, $k\ge 2$ with $j+k\le n$, and $\ell\in\{1,\ldots,\mathcal{L}\}$, then $p_{j+1}\in G_\ell$.*

Proof. By way of contradiction, suppose there exists an instance for which there is no optimal solution satisfying the condition of the theorem. Consider the optimal solution for which the smallest index j that violates this condition is maximized. Let j^* be the smallest index in this solution for which there exists a k with $G(p_{j^*})=G(p_{j^*+k})$ and $G(p_{j^*+1})\neq G(p_{j^*})$. Let k^* be such an index k, $G_\ell=G(p_{j^*})$ and $G_{\ell'}=G(p_{j^*+1})$. We first show that $j\ge j^*+1$ for all $\{j\mid p_j\in G_{\ell'}\}$. Suppose not; let $\hat{j}=\arg\max\{j\mid p_j\in G_{\ell'},p_{j+1}\notin G_{\ell'},j<j^*+1\}$ (which will be non-empty by assumption). Then passenger indices j', $j'+k'$ with $j'=\hat{j},k'=j^*+1-\hat{j}$ satisfy $p_{j'},p_{j'+k'}\in G_{\ell'}$ and $p_{j'+1}\notin G_{\ell'}$ are a set of indices violating the claim of the

theorem, contradicting the minimality of j^*. Hence, (j^*+1) is the minimum index among all $p_j \in G_{\ell'}$.

In the remainder of the proof, we construct another solution in which the index j^* does not violate the claims of the theorem. This contradicts the maximality of j^* among all optimal solutions, thereby establishing the result.

Conditioning on the relative values of the departure times of the CVs for G_ℓ and $G_{\ell'}$, first consider the case where $\mathrm{CVT}(G_\ell) \leq \mathrm{CVT}(G_{\ell'})$. We claim that exchanging the group assignment of p_{j^*+1} and $p_{j^*+k^*}$ and holding all else equal results in another feasible solution. We need only show that (a) $\texttt{rel}(p_{j^*+1}) \leq \mathrm{CVT}(G_\ell) \leq \texttt{ded}(p_{j^*+1})$ and (b) $\texttt{rel}(p_{j^*+k^*}) \leq \mathrm{CVT}(G_{\ell'}) \leq \texttt{ded}(p_{j^*+k^*})$. (a) follows directly from Lemma 3 with $i_1 = j^*, i_2 = j^* + k^*$ and $i = j^* + 1$. The first inequality in (b) follows because $\texttt{rel}(p_{j^*+k^*}) \leq \mathrm{CVT}(G_\ell)$, by the feasibility of the original solution, and $\mathrm{CVT}(G_\ell) \leq \mathrm{CVT}(G_{\ell'})$, by assumption. The second inequality in (b) holds because $\texttt{ded}(p_{j^*+1}) \leq \mathrm{CVT}(G_{\ell'})$, by the feasibility of the original solution and $\texttt{ded}(p_{j^*+1}) \leq \texttt{ded}(p_{j^*+k^*})$, by the ordering of passengers. Furthermore, by Lemma 2 the objective function remains unchanged by this exchange, and is therefore optimal. If the resulting solution satisfies the claim of this theorem, then the claim holds. If not, then the claim of this theorem is violated for another $j > j^*$; contradicting the maximality of j^*.

We now consider the alternative case where $\mathrm{CVT}(G_\ell) > \mathrm{CVT}(G_{\ell'})$. The exchange from the previous case may not work because putting passenger $p_{j^*+k^*}$ into group $G_{\ell'}$ may not be feasible. Additionally, if there exist $k' > 0$ passengers in G_ℓ with indices lower than j^* then these passengers must be $p_{j^*-k'}, \ldots, p_{j^*-1}$. If not, it would contradict the assumption of j^* as the smallest index in the optimal solution violating the claim of this theorem. Define k' so that $p_{j^*-k'}$ is the minimum indexed passenger in group G_ℓ, which is 0 if p_{j^*} is the minimum indexed passenger.

Consider the following two-step exchange—for $i = 0, \ldots, k'$, move each passenger p_{j^*-i} from G_ℓ into $G_{\ell'}$. Then, move the $k'+1$ passengers with the highest indices in the resulting $G_{\ell'}$ into G_ℓ. The resulting groups have the same cardinality as they originally had, and so by Lemma 2 the objective values remain the same.

We now show that the resulting solution is valid and then show that the choice of optimal solution contradicts the maximality assumption on the selection of j^*, which concludes the proof. Any passenger $p_i \in G_\ell$ with index $i \leq j^*$ can be moved to $G_{\ell'}$ without violating p_i's arrival time window because $\texttt{rel}(p_i) \leq \texttt{rel}(p_{j^*+1}) \leq \mathrm{CVT}(G_{\ell'}) < \mathrm{CVT}(G_\ell)$ and $\texttt{ded}(p_i) \geq \mathrm{CVT}(G_\ell) > \mathrm{CVT}(G_{\ell'})$. Additionally, any passenger $p \in G_{\ell'}$ can be moved to G_ℓ without violating the arrival time windows because $\texttt{rel}(p) \leq \mathrm{CVT}(G_{\ell'}) < \mathrm{CVT}(G_\ell)$ and $\texttt{ded}(p) \geq \texttt{ded}(p_{j^*}) \geq \mathrm{CVT}(G_\ell)$. Hence, the resulting solution is also optimal. Finally, if the resulting solution violates the claim of this theorem, then the smallest index must be larger than j^*. This again contradicts the maximality of j^* among all optimal solutions, as assumed. □

Note that we can independently reorganize the set of passengers for each destination to be sorted as indicated in Theorem 1. The result can thus be extended to instances with multiple destinations, with the additional assumption that each CV is restricted to carry passengers going to a common building (Assumption 2).

Algorithm

We describe an algorithmic framework for constructing a heuristic solution and improving it with local search.

The heuristic construction comprises three phases: (i) sorting the passengers; (ii) grouping them; and (iii) assigning groups to vehicles and scheduling the LMT trips.

Sorting the Passengers

We define a bijection $\texttt{ordp} : \{1, \ldots, |\mathcal{P}|\} \to \mathcal{P}$ such that

$$\texttt{ded}\big(\texttt{ordp}(i)\big) \leq \texttt{ded}\big(\texttt{ordp}(j)\big), 1 \leq i < j \leq |\mathcal{P}|.$$

The intuition is that passengers with similar arrival deadlines can be grouped and served together.

Grouping the Passengers

Algorithm 1 defines groups of passengers with common destination by traversing the set of passengers as sorted by $\texttt{ordp}$. For each passenger $p = \texttt{ordp}(i)$ that does not have a group yet, i.e., $\text{Grouped}_i = \texttt{False}$, the loop starting at line 2 creates a new group g. Then the loop starting at line 6 verifies if each of the next ungrouped passengers $q = \texttt{ordp}(j)$ for $j > i$ can be added to group g. Since the passengers are added as ordered by $\texttt{ordp}$, their deadlines are nondecreasing and the verification that the earliest arrival time for q is before the latest time for p in line 9 suffices to determine if the resulting group is feasible for a single destination.

input: Set of passengers $\mathcal{P}$ and their sorting $\texttt{ordp}$
output: Set of groups $\mathcal{G}$
1 Set: $\mathcal{G} \leftarrow \emptyset$ and $\text{Grouped}_i \leftarrow \texttt{False}$ for $i = 1, \ldots, |\mathcal{P}|$.
2 **for** $i \leftarrow 1, \ldots, |\mathcal{P}|$ **do**
3   **if** ***not*** *Grouped$_i$* **then**
4   Set $p \leftarrow \texttt{ordp}(i)$ and $g \leftarrow \{p\}$
5   Set $j \leftarrow i + 1$
6   **while** $j \leq |\mathcal{P}|$ ***and*** $|g| < CV^{\max}$ **do**
7   **if** ***not*** *Grouped$_j$* **then**
8   Set $q \leftarrow \texttt{ordp}(j)$
9   **if** $\texttt{rel}(q) < \texttt{ded}(p)$ ***and b(p) = b(q)*** **then**
10   Set $g \leftarrow g \cup \{q\}$ and $\text{Grouped}_j \leftarrow \texttt{True}$
11   **end**
12   **end**
13   Set $j \leftarrow j + 1$
14   **end**
15   Set $\mathcal{G} \leftarrow \mathcal{G} \cup \{g\}$
16   **end**
17 **end**

Algorithm 1: Groups the passengers by destination

Optimal Scheduling of the Groups

We now present an ILP formulation to obtain an optimal schedule of groups on MT and LMT for a given grouping of passengers for the LMT trips. For ease of exposition, we begin by describing some sets that can be pre-computed for a fixed grouping.

The set of feasible start times on CVs for a group $g \in \mathcal{G}$ using route r can be computed as

$$\mathcal{T}(g,r) := \left\{ t \;\middle|\; \begin{array}{r} (t + \tau^{\text{travel}}(r, b(p))) \in [\mathtt{rel}(p), \mathtt{ded}(p)] \\ \forall\, p \in g \end{array} \right\}.$$

We use binary variables $x_{g,r,t} \in \{0,1\}$ to represent the start time t for group g on a CV route r, i.e.

$$x_{g,r,t} = \begin{cases} 1 \text{ if group } g \text{ uses route } r \text{ and starts at } t \\ 0 \text{ otherwise} \end{cases}$$
$$\forall\, g \in \mathcal{G}, r \in \mathcal{R}(g), t \in \mathcal{T}(g,r),$$

where $\mathcal{R}(g)$ is the set of CV routes that stop at destinations of passengers in g; i.e.

$$\mathcal{R}(g) := \{ r \in \mathcal{R} \mid b(p) \in \mathcal{B}(r) \,\forall\, p \in g \}.$$

To model the use of CVs by different groups, we compute a set defining the groups, routes, and start times directly affecting the number of CVs used at each time instant t as

$$\mathcal{V}(t) := \left\{ (g, r, t') \mid t \in [t', t' + \tau^{\mathtt{trip}}(g,r)] \right\}.$$

Each of these is a subset of the indices of decision variables and its use will be clear in the upcoming formulation.

The exact ILP modeling the optimal scheduling of passengers on CVs for LMT service is as follows:

$$\min \sum_{g \in \mathcal{G}} \sum_{r \in \mathcal{R}(g)} \sum_{t \in \mathcal{T}(g,r)} \alpha_{g,r,t} \cdot x_{g,r,t} \tag{3a}$$

$$\text{s.t.} \sum_{r \in \mathcal{R}(g)} \sum_{t \in \mathcal{T}(g,r)} x_{g,r,t} = 1 \,\forall\, g \in \mathcal{G} \tag{3b}$$

$$\sum_{(g,r,t') \in \mathcal{V}(t)} x_{g,r,t'} \leq |\mathcal{C}| \,\forall\, t \in \mathcal{T}. \tag{3c}$$

The objective coefficient $\alpha_{g,r,t}$ is defined as

$$\alpha_{g,r,t} := \sum_{p \in g} \left(t + \tau^{\text{travel}}(r, b(p)) - t^{\text{start}}(u^{min}(t, s(p))) \right),$$
$$\text{where } u^{\min}(t, s(p)) := \arg \min_{u \in \mathcal{U}(s(p)) : t^{\mathtt{T0}}(u) \leq t} (t - t^{\text{start}}(u)).$$

Constraint (3b) imposes that each group is assigned to exactly one CV route and start time. Constraint (3c) ensures that the number of groups that are on the LMT service at any time is less than or equal to the total number of CVs.

For a given optimal solution $x^*_{g,r,t}$ to (3), we can obtain an optimal assignment of routes $r^* : \mathcal{G} \to \mathcal{R}$ and start times $t^* : \mathcal{G} \to \mathcal{T}$ for groups on CVs as follows:

$$r^*(g) = r \iff x^*_{g,r,t} = 1 \text{ and } t^*(g) = t \iff x^*_{g,r,t} = 1.$$

The optimal MT services $u^* : \mathcal{P} \to \mathcal{U}$ are then obtained as

$$u^*(p) = u^{\min}(t^*(g), s(p),) \text{ where } g : p \in g.$$

Formulation (3) does not explicitly assign a CV to any of the groups. The CV assignments can be determined once the solution is given by Algorithm 2. The algorithm makes use of a priority queue $\mathtt{veh\_queue}$ consisting of (v, t) pairs, where v is index of the CV and t is time that the CV is available at $\mathtt{T0}$ for servicing a group. The pairs with lower t have

input: Set of vehicles $\mathcal{C}$, group sorting $\mathtt{ord}$, route choice $r^*(g)$
output: Mapping $\mathtt{veh} : \mathcal{G} \to \mathcal{C}$

$\mathtt{veh\_queue} \leftarrow \emptyset$
for $v \in \mathcal{C}$ **do**
    $\mathtt{veh\_queue}$.push$\big((v, 0)\big)$
end
for $i \leftarrow 1, \ldots, |\mathcal{G}|$ **do**
    $(v, t) \leftarrow \mathtt{veh\_queue}$.pop()
    $g \leftarrow \mathtt{ord}(i)$
    $\mathtt{veh}(g) \leftarrow v$
    $\mathtt{veh\_queue}$.push$\big((v, t + \tau^{\mathtt{trip}}(g, r^*(g)) + 1)\big)$
end

Algorithm 2: Assigns vehicles to scheduled groups

higher priority in the queue. Let $\mathtt{ord} : \mathcal{G} \to \{1, \ldots, |\mathcal{G}|\}$ be a mapping that sorts $\mathcal{G}$ in ascending order of $t^*(g)$, i.e.

$$\mathtt{ord}(g) \leq \mathtt{ord}(g') \iff t^*(g) \leq t^*(g').$$

Algorithm 2 returns a mapping $\mathtt{veh} : \mathcal{G} \to \mathcal{C}$ such that $\mathtt{veh}(g)$ is the vehicle assigned to service the group g.

Some comments regarding formulation (3) are in order. In contrast to formulation (2), the time indices in formulation (3) for each group are restricted to the intersection of possible departures for all passengers in the group, hence making the formulation more tractable. Furthermore, formulation (3) also avoids the combinatorial explosion in the number of routes, which takes all possible subsets of $CV^{\max}$ stops. It is possible to define smaller formulations in both cases by modeling time with continuous variables instead of indexing binary variables, as discussed by Floudas and Lin (2004). However, Balas (1985) has shown that the linear relaxation of such formulations is the weakest possible, thereby pushing the solution process towards a time-consuming enumeration of alternatives that does not truly benefit from the ILP techniques that mathematical optimization solvers exploit.

Lower Bounds

This section describes lower bounds on MT and LMT trip times per passenger as well as on the sum of wait and LMT trip times per passenger and per group of passengers. We use these bounds to evaluate the quality of solutions and also to direct the local search methods as described next.

The minimum MT trip time per passenger is defined as

$$\text{MMT}(p) := \min \left\{ t^{\mathtt{T0}}(u) - t^{\text{start}}\big(u, s(p)\big) \mid u \in \mathcal{U}\big(s(p)\big) \right\}.$$

In turn, the minimum LMT trip time is given by

$$\text{MLM}(p) := \min \left\{ \tau^{\text{travel}}\big(r, b(p)\big) \mid r \in \mathcal{R},\; b(p) \in \mathcal{B}(r) \right\}.$$

For convenience, we define a function corresponding to the minimum waiting time if a passenger coming from station s leaves the terminal in the time window $[t_a, t_b]$:

$$\text{MWT}(s, t_a, t_b) := \min \left\{ t_a - t^{\mathtt{T0}}(u) \mid u \in \mathcal{U}(s), t^{\mathtt{T0}}(u) \leq t_b \right\}.$$

This waiting time is implied by the set of MT trips that can bring a passenger from s to the terminal within $[t_a, t_b]$.

In turn, the time window may be restricted according to the passengers that are grouped together.

For each passenger p, we can compute the minimum waiting time if the passenger takes the shortest travel time as

$$\text{MPWT}(p) := \text{MWT}\begin{pmatrix} s(p), \texttt{rel}(p) - \text{MLM}(p), \\ \texttt{ded}(p) - \text{MLM}(p) \end{pmatrix}.$$

Note that this is not a bound on the waiting time, since passenger p could wait less if taking a longer LMT trip. However, trading waiting for trip time would never lead to a smaller sum of both, and thus the combined bound is valid. We can therefore obtain a lower bound **LB** that combines the trip times and the wait times for each passenger as

$$\textbf{LB} := \sum_{p \in \mathcal{P}} \text{MMT}(p) + \text{MLM}(p) + \text{MPWT}(p). \tag{4}$$

If the number of CVs is large enough to define groups where no passenger delays another and all groups can be scheduled at the most convenient time for its passengers, then **LB** corresponds to the value of the optimal solution. When this is not the case, we need to account for the tighter departure time windows defined by each group.

For a group of passengers g, the minimum wait time of each passenger $p \in g$ if taking a trip of shortest time is

$$\text{MGWT}(g,p) := \text{MWT}\begin{pmatrix} s(p), \max\limits_{q \in g}(\texttt{rel}(q) - \text{MMT}(q)), \\ \min\limits_{q \in g}(\texttt{ded}(q) - \text{MMT}(q)) \end{pmatrix}.$$

Whenever $\text{MGWT}(g,p) > \text{MPWT}(p)$ for some $p \in g$, then some passenger in $g \setminus \{p\}$ is delaying p. While the former expressions define a lower bound per passenger, the latter is used to decide which groups to modify by local search.

Break-and-Shift Local Search

Algorithm 1 aims to create as few groups as possible, each with many passengers. This can increase total waiting time because the CV fleet may be underutilized. Hence, we define a local search method to iteratively modify these groups by breaking them and swapping passengers.

We define a local search to address this, which loops between two operations: (i) breaking groups where one passenger delays another; and (ii) shifting passengers to groups with earlier departure times when this is feasible and no delay is implied. These operations are performed on sets of groups that arrive through MT trips at the same time to `T0` according to an optimal solution of (3). If the LMT trips take less than the inter-arrival time of MT, operation (i) can result in as many groups as the number of CVs ($|\mathcal{C}|$). The fragmented groups resulting from operation (i) can be reorganized differently by operation (ii), whereby some groups might vanish. In such a case, we are able to repeat a loop with both operations once more. Hence, we define a set of MT arrival times as

$$T := \left\{ t^{\text{T0}}(u) \mid u \in \mathcal{U}(s), s \in \mathcal{S} \right\}$$

and the groups corresponding to each $t \in T$ is

$$G_t := \left\{ g \in \mathcal{G} \mid t = \max\left\{ t^{\text{T0}}(u) \mid u \in \mathcal{U}, t^{\text{T0}}(u) \leq t^*(g) \right\} \right\}.$$

For convenience, we denote the set of passengers on each group g as $g := \{g(1), \ldots, g(|g|)\}$, with $\texttt{ded}(g(i)) \leq \texttt{ded}(g(j))$ for $i \leq j$. Furthermore, let G^b denote the set of groups with all passengers heading to b, and let $G^b := \left\{ g_1^b, \ldots, g_{|G^b|}^b \right\}$ with $t^*(g_i^b) \leq t^*(g_j^b)$ for $i \leq j$. Algorithm 3 describes operation (i), which breaks down some groups in G_t. Staring with a set H_t corresponding to G_t, the procedure loops over each group $g \in G_t$, attempting to break them to increase the objective value of the solution. If the total number of groups is still smaller than the number of CVs and the earliest passenger $g(1)$ has to wait more with the group than alone, the group is broken before the first passenger $g(i)$ that causes the delay, thereby replacing g with two groups in H_t. Algorithm 4 describes operation (ii), which moves passengers between consecutive groups. In this case, H^b begins empty, we refer to the next group to be added to H^b as h, and we loop on G^b to construct this set H^b. We define the first group in G^b as the initial incumbent group h and, while possible, we keep adding passengers to it by looping on the subsequent groups. We either empty these groups on h and proceed to the next one, or else fill h to capacity. In the latter case, we add h to H^b and reassign h as h', which consists of the remaining passengers from the last group of G^b that we iterated on.

In order to avoid infeasibility when breaking and reorganizing the groups, one can solve (3) after each modification. For efficiency, we have found that it was sufficient to apply both operators until either there was no change in the groups or the groups resulted in an infeasible schedule.

input: A subset of groups $G_t \in \mathcal{G}$
output: A fragmented set H_t with same passengers

Set $H_t \leftarrow G_t$
foreach $g \in G_t$ **do**
  if $|H_t| < |\mathcal{C}|$ & $MGWT(g, g(1)) > MPWT(g(1))$ **then**
    for $i \leftarrow 2, \ldots, |g|$ **do**
      if $MPWT(g(i)) > MPWT(g(1))$ **then**
        Set $g' \leftarrow \{g(1), \ldots, g(i-1)\}$
        Set $g'' \leftarrow \{g(i), \ldots, g(|g|)\}$
        Set $H_t \leftarrow (H_t \setminus \{g\}) \cup \{g', g''\}$
        break
      end
    end
    if $|H_t| = |\mathcal{C}|$ **then**
      break
    end
  end
end

Algorithm 3: Breaks groups with co-passenger delays

Experiments

We present numerical experiments in the setting of Figure 1. The passengers originate from a set of 4 stations and desire to reach one of the buildings in B1-B10 within a specified time-window. $\mathcal{U}$ consists of MT services that reach `T0` every 15 minutes and the travel time between stations is 5 minutes. We assume that the CVs are all parked at `T0` and return back to `T0` after dropping off all passengers. The CVs take 1 minute to go between `T0`-B1, `T0`-B10, and all pairs of build-

input: A subset of groups $G^b \in \mathcal{G}$
output: A reorganized set H^b with same passengers

Set $H^b \leftarrow \emptyset, h \leftarrow g_1^b$
for $i \leftarrow 2, \ldots, |G^b|$ **do**
    if $|h| = CV^{\max}$ **then**
        Set $H^b \leftarrow H^b \cup \{h\}, h \leftarrow g_i^b$
    end
    if $|h| < CV^{\max}$ **then**
        for $j \leftarrow 1, \ldots, \min\{CV^{\max} - |h|, |g_i^b|\}$ **do**
            Set $p \leftarrow g_i^b(j)$
            if $\mathtt{rel}(p) \leq \mathtt{ded}(h(1))$ *&* $\mathit{MGWT}(h \cup \{p\}, h(1)) = \mathit{MGWT}(h, h(1))$ **then**
                Set $h \leftarrow h \cup \{p\}$
            end
        end
        $h' \leftarrow g_i^b \setminus h$
        if $|h'| > 0$ **then**
            Set $H^b \leftarrow H^b \cup \{h\}, h \leftarrow h'$
        end
    end
end
$H^b \leftarrow H^b \cup \{h\}$

Algorithm 4: Reorganizes passenger groups

ings that are adjacent on the shaded track, except for B5-B6 which takes 2 minutes. The CVs are restricted to move along the shaded region shown in Figure 1. We also assume that a CV spends 0.5 minutes at a building where they drop passengers. As a result, the time that a passenger reaches the destination depends on co-passengers in the CV that have prior destinations. The modeling of drop-off time is important in applications where the capacity of $CV^{\max}$ is small, typically ≤ 5. Inspired by the application to corporate-campus settings, we use a small number of destinations. This may not be the case in all applications, where one might expect the number of destinations for the LMT service to be in the same order as the number of passengers, thus making the problem harder to solve. In those cases, however, we believe that the suggestion by Mahéo, Kilby, and Hentenryck (2018) to treat last-mile stops as aggregations of several passenger destinations, such as bus stops, is a reasonable compromise. Hence, the chosen number of destinations is of minor importance.

In the experiments performed, we restrict the passengers to be grouped according to their destination. The set of possible CV round-trip movements considered are (in the sequence of building visits)

$$\{\mathtt{T0}, B1, B2, \ldots, B9, B10, \mathtt{T0}\},\ \{\mathtt{T0}, B10, B9, \ldots, B2, B1, \mathtt{T0}\}$$
$$\{\mathtt{T0}, B1, B2, B3, B8, B9, B10, \mathtt{T0}\},\ \{\mathtt{T0}, B10, B9, B8, B3, B2, B1, \mathtt{T0}\}$$
$$\{\mathtt{T0}, B1, \ldots, B3, B8, B7, \ldots, B3, B8, \ldots, B10, \mathtt{T0}\}$$
$$\{\mathtt{T0}, B10, \ldots, B8, B3, B4, \ldots, B8, B3, \ldots, B1, \mathtt{T0}\}.$$

On each of the CV round-trip movements, a stop at the first occurrence of the building is considered a route r for the CV. Thus, $\mathcal{R}$ consists of 52 (= 10 + 10 + 6 + 6 + 10 + 10) routes. From an optimality perspective, we expect the CVs to typically use only the shortest route to the destination. However, in a few instances the passengers with earliest deadlines are assigned to longer routes so that they arrive to their destinations at the start of the time-windows. Removing this option can result in infeasibility.

To test the algorithms described, we generated 10 scenarios consisting of 600 passengers. The desired earliest time ($\mathtt{rel}(p)$) of arrival for the passengers is assumed to be uniformly distributed over 1 hour, in increments of 30 seconds. For each passenger, the origin and destination station are drawn uniformly and independently at random from the four stations and 10 buildings, as shown in Figure 1. We assume that the length of time windows ($\mathtt{ded}(p) - \mathtt{rel}(p)$) are identical for all the passengers; the values are set by assigning, if t' is the requested arrival time for a passenger, $\mathtt{rel}(p) = t' - K/2$ and $\mathtt{ded}(p) = t' + K/2$, for a fixed $K > 0$. We test the impact of K and the number of CVs by varying $K \in \{5, 10, 20\}$ and $|\mathcal{C}| \in \{30, 40, 50, 60\}$, resulting in 12 different configurations, and thus 120 different instances. All experiments were run on a machine with an Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz and 32 GB RAM. All algorithms were implemented in `Python 2.7.6` and the ILPs are solved using `Gurobi 7.5.1`.

Table 2 presents a summary of the results. The first two columns report the size of the time windows and the number of CVs, respectively. Column $\text{UB}^{\text{red}} = (UB^0 - UB^H)/(UB^0 - LB^*) \times 100$ is the percent decrease in the optimality gap between the initial solution obtained using the heuristic (UB^0) and the final (UB^H) solution resulting from the local search procedure, over the best known lower bound (LB^*) obtained by `Gurobi` solving model (2) with a time limit of 10 minutes, averaged over all instances in that configuration. A 100% reduction means the entire optimality gap is closed. It is clear that the heuristic is able to obtain optimal solutions on problems where the number of CVs and time windows are not very constrained.

The fifth and sixth columns in Table 2 show the average percent optimality gap at termination over the instances where solutions are found when solving model (2) using `Gurobi`: (i) without the heuristic solution (MIP) and (ii) with the heuristic solution (MIP+H) as an initial solution using the `MIPStart` feature. In both approaches, the time limit is set to 10 minutes. The optimality gap is computed as $(UB^* - LB^*)/LB^* \times 100$ where UB^* is the best feasible solution and LB^* is the best lower bound obtained by the approach (MIP or MIP+H). In all cases, the gaps are computed after subtracting the constant MT travel times from each station of origin to `T0`. In the cases that not all runs found feasible solutions, next to each average there is the number of cases where at least one feasible solution is found. The last two columns in Table 2 show the number of instances that were solved to optimality. From the results in Table 2, it is clear that MIP+H outperforms MIP in terms of the number of problems solved to optimality and average optimality gap closed.

We also investigate the effect on the solution quality due to the imposition of single destination per CV in Assumption 2. An upper bound to the optimal gap with respect to the general case is computed as $\frac{(UB(MIP+H) - \mathbf{LB})}{\mathbf{LB}}$, where $UB(MIP + H)$ is the best solution obtained from our algorithm and **LB** is the lower bound in (4). Note that the

Table 2: Summary of results averaging 10 instances per configuration of time window length K and flee size $|\mathcal{C}|$

K	$\|\mathcal{C}\|$	UB$^{\text{red}}$ (%)	t^{H}	Avg. Gap (%) MIP		Avg. Gap (%) MIP+H	# solved MIP	# solved MIP+H
5	30	2.9	0.99	0.0	(1)	0.1	1	4
5	40	14.3	3.73	0.0	(7)	0.0	7	10
5	50	31.4	6.38	0.0	(8)	0.0	8	10
5	60	60.2	15.25	0.0		0.0	10	10
10	30	6.5	3.65	–	(0)	2.0	0	0
10	40	34.0	12.31	0.3	(6)	0.2	2	4
10	50	76.1	25.97	0.0		0.0	10	10
10	60	96.0	32.23	0.0		0.0	10	10
20	30	62.4	7.99	0.0		0.0	10	10
20	40	100.0	8.7	0.0		0.0	10	10
20	50	100.0	7.62	0.0		0.0	10	10
20	60	100.0	7.54	0.0		0.0	10	10

computation of **LB** in (4) is independent of the number of stops per CV. Table 3 summarizes the gaps for the solution obtained using MIP+H. These results indicate that Assumption 2 is not very restrictive on the instances tested.

Table 3: Upper bound on the optimal gap of the solutions obtained under Assumption 2 with respect to the general case.

K	$\|\mathcal{C}\|$	Avg. Gap %	Min. Gap %	Max. Gap %
5	30	6.2	5.6	6.7
5	40	2.7	2.3	3.2
5	50	1.3	1.0	1.6
5	60	0.5	0.4	0.8
10	30	12.3	10.7	14.32
10	40	2.9	2.5	3.5
10	50	0.7	0.4	0.9
10	60	0.1	0.0	0.2
20	30	0.0	0.0	0.0
20	40	0.0	0.0	0.0
20	50	0.0	0.0	0.0
20	60	0.0	0.0	0.0

Figure 3 depicts (left) a scatter plot with a point per instance and (right) a cumulative distribution plot of performance, comparing the solution times for MIP and MIP+H. The scatter plot contains a point per instance where the coordinates of the point correspond to the time taken to solve the instance with and without the heuristic. The plots are sized according to $|\mathcal{C}|$ and colored according to K. The cumulative distribution plot of performance provides another depiction of the data, showing the number of instances that are solved at each point of time in the solution horizon for both algorithmic configurations.

Conclusions & Future Work

This paper addresses the ILMTP, focusing on a special case of this problem where each CV makes a stop at just one building per trip. Based on this setting, we prove a structured result regarding the form of optimal solutions that results in insights used to efficiently solve the problem. Our experiments indicate that the optimal solutions to this case are not far from a lower bound for the general problem, suggesting that (a) starting from this solution for a future study is promising, and (b) real-world systems implementing this solution may already achieve near optimal schedules.

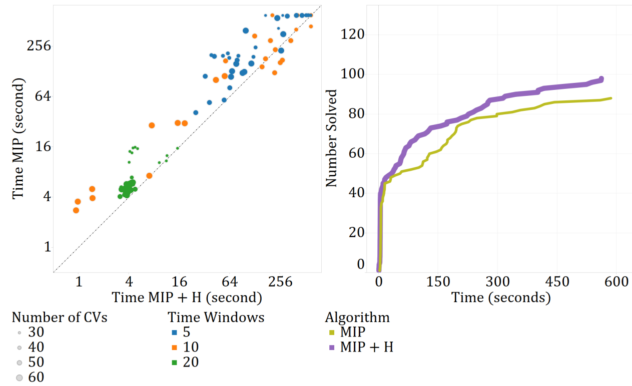

Figure 3: Comparison of model 2 with and without the heuristic. (left) A scatter plot with a point were run. (right) A cumulative distribution plot of performance indicating the number of instances solved by second.

We generalize and make improvements to a recent MIP formulation for the general problem suggested by Wang (2017), and use a constructive heuristic and local search procedure for identifying an initial high-quality solution that, when used as the starting solution for a commercial MIP solver, often closing the entire optimality gap. The results obtained with and without the heuristic also indicate a significant statistical improvement in solution times. Applying the (paired) Wilcoxon Signed-Rank Test (a non-parametric statistical hypothesis test which assesses whether population mean ranks differ in matched samples) to test whether the run times of the algorithms are different results in a p-value of $1.267 \cdot 10^{-10}$, indicating a strong statistical difference in the run times. The instances appear to increase in difficulty as K and $|\mathcal{C}|$ decrease, which can be attributed to the problem becoming more constrained.

As the design of automated transportation systems becomes a reality, it is critical that models and algorithms such as the ones developed in this paper are developed to ensure the system used are run efficiently. One critical extension will be to the case when CVs can stop at more than one building. The local search algorithm developed can be modified to accept solutions of this form, which may reduce the total waiting times of customers. Additionally, we plan to design problem-specific solution algorithms that can exploit results like the one proven in Theorem 1.

Another avenue for extending this work would be to exploit its complementarity with the study by Mahéo, Kilby, and Hentenryck (2018), where more consideration is given to designing the MT service. We hope that a unified framework can be developed in future work.

References

Anderson, J. E. 1998. Control of personal rapid transit systems. *J. Adv. Transportation* 32(1):57–74.

Balas, E. 1985. On the facial structure of scheduling polyhedra. *Mathematical Programming Study* 24:179–218.

Berger, T.; Sallez, Y.; Raileanu, S.; Tahon, C.; Trentesaux, D.; and Borangiu, T. 2011. Personal rapid transit in an open-control framework. *Comput. Indust. Engrg.* 61(2):300–312.

Bly, P. H., and Teychenne, R. 2005. Three financial and socio-economic assessments of a personal rapid transit system. In *Proc. 10th Internat. Conf. Automated People Movers*, 1–16.

Brake, J.; Nelson, J. D.; and Wright, S. 2004. Demand responsive transport: Towards the emergence of a new market segment. *J. Transport Geography* 12(4):323–337.

Chevrier, R.; Jourdan, A. L. L.; and Dhaenens, C. 2012. Solving a dial-a-ride problem with a hybrid evolutionary multi-objective approach: Application to demand responsive transport. *Appl. Soft Comput.* 12(4):1247–1258.

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

Daganzo, C. F. 1978. An approximate analytic model of many-to-many demand responsive transportation systems. *Transportation Res.* 12(5):325–333.

Diana, M.; Dessouky, M. M.; and Xia, N. 2006. A model for the fleet sizing of demand responsive transportation services with time windows. *Transportation Res. Part B: Methodological* 40(8):651–666.

Floudas, C. A., and Lin, X. 2004. Continuous-time versus discrete-time approaches for scheduling of chemical processes: a review. *Computers & Chemical Engineering* 28(11):2109–2129.

Horn, M. E. 2002a. Fleet scheduling and dispatching for demand responsive passenger services. *Transportation Res. Part C: Emerging Tech.* 10(1):35–63.

Horn, M. E. 2002b. Multi-modal and demand-responsive passenger transport systems: A modelling framework with embedded control systems. *Transportation Res. Part A: Policy Practice* 36(2):167–188.

Jaw, J. J.; Odoni, A. R.; Psaraftis, H. N.; and Wilson, N. H. 1986. A heuristic algorithm for the multi-vehicle advance request dial-a-ride problem with time windows. *Transportation Res. Part B: Methodological* 20(3):243–257.

Lees-Miller, J. D.; Hammersley, J. C.; and Davenport, N. 2009. Ride sharing in personal rapid transit capacity planning. In *12th Internat. Conf. Automated People Movers*, 321–332.

Lees-Miller, J. D.; Hammersley, J. C.; and Wilson, R. E. 2010. Theoretical maximum capacity as benchmark for empty vehicle redistribution in personal rapid transit. *Transportation Res. Record: J. Transportation Res. Board* 2146(1):76–83.

Lei, H.; Laporte, G.; and Guo, B. 2012. Districting for routing with stochastic customers. *Eur. J. Transportation Logist.* 1(1?2):67–85.

Liu, Z.; Jiang, X.; and Cheng, W. 2012. Solving in the last mile problem: Ensure the success of public bicycle system in beijing. *Procedia Soc. Behav. Sci.* 43:73–78.

Mageean, J., and Nelson, J. D. 2003. The evaluation of demand responsive transport services in europe. *J.Transport Geography* 11(4):255–270.

Mahéo, A.; Kilby, P.; and Hentenryck, P. V. 2018. Benders decomposition for the design of a hub and shuttle public transit system. *Transportation Science*.

Mueller, K., and Sgouridis, S. P. 2011. Simulation-based analysis of personal rapid transit systems: Service and energy performance assessment of the masdar city prt case. *J. Adv. Transportation* 45(4):252–270.

Palmer, K.; Dessouky, M.; and Abdelmaguid, T. 2004. Impacts of management practices and advanced technologies on demand responsive transit systems. *Transportation Res. Part A: Policy Practice* 38(7):495–509.

Quadrifoglio, L.; Dessouky, M. M.; and Ordóñez, F. 2008. A simulation study of demand responsive transit system design. *Transportation Res. Part A: Policy Practice* 42(4):718?737.

Shaheen, S. 2004. U.S. carsharing & station car policy considerations: Monitoring growth, trends & overall impacts. Technical report, Institute of transportation studies, Working paper series, Institute of Transportation Studies, UC Davis.

Shen, Y.; Zhang, H.; and Zhao, J. 2017. Simulating the First Mile Service to Access Train Stations by Shared Autonomous Vehicle. In *Transportation Research Board 96th Annual Meeting*.

Thien, N. D. 2013. *Fair cost sharing auction mechanisms in last mile ridesharing*. Ph.D. Dissertation, Singapore Management University.

Toth, P., and Vigo, D. 2014. *Vehicle Routing: Problems, Methods and Applications*. SIAM, second edition.

Wang, H. 2017. Routing and Scheduling for a Last-Mile Transportation Problem. *Transportation Science* 1–17 (forthcoming).

Wilson, N. H., and Hendrickson, C. 1980. Performance models of flexibly routed transportation services. *Transportation Res. Part B: Methodological* 14(1):67–78.