Simulation-based Design and Analysis of On-demand Mobility Services

BESTMILE SA
AVENUE DE RHODANIE 58
1007 LAUSANNE
SWITZERLAND
Abstract

On-demand mobility services – autonomous as well as human-driven – promise to transform transportation and urban living. Moving more people with fewer vehicles and fewer kilometers driven can reduce congestion and pollution, enable new levels of productivity otherwise lost in long commutes, and allow us to re-imagine cities around people instead of cars. Yet, designing such services involves an array of challenges: from demand data acquisition and analysis, to developing efficient matching and routing algorithms, to finally evaluating a variety of service configurations. This report illustrates the process of designing and analyzing various types of on-demand mobility services in the City of Chicago, Illinois. We study how different service designs respond to real travel demand patterns in terms of key performance indicators such as demand acceptance rate, excess ride time, deviation from desired pickup time, vehicle occupancy and various fleet efficiency metrics. Rather than giving a recipe, this type of analysis helps service providers and public transit authorities gain insights into the interplay of the different passenger- and fleet-related key performance indicators and the fundamental trade-offs between passenger service level and fleet efficiency. Furthermore, the effects on externalities associated with passenger transportation such as traffic volume, the need for parking space, and CO₂ emissions can be estimated, which is of interest to city planners and transportation authorities.

Keywords

service design, on-demand mobility, ridesharing, prebooking, instant booking, matching, routing, optimization, simulation, key performance indicators (KPIs), Chicago, demand data
## Contents

1 Introduction .............................................. 3  
2 State of the art ........................................... 5  
3 Modeling .................................................. 7  
   3.1 Service typology ...................................... 7  
   3.2 Demand matching model .............................. 8  
   3.3 Prebooking model ................................... 9  
   3.4 Traffic model ...................................... 10  
4 Methodology .............................................. 11  
   4.1 Simulation framework ............................... 11  
   4.2 Key performance indicators ........................ 12  
5 Demand data .............................................. 14  
6 Experimental design ..................................... 16  
   6.1 Service areas ....................................... 16  
   6.2 Service parameters ................................ 17  
   6.3 Demand patterns ................................... 18  
7 Simulation results ....................................... 19  
   7.1 Door-to-door Lincoln Park Micro-transit .......... 19  
   7.2 Station-based Lincoln Park Micro-transit .......... 22  
   7.3 O’Hare Shuttle ..................................... 25  
   7.4 City Ridesharing ................................... 28  
   7.5 Comparison of the service areas .................. 32  
8 Conclusions .............................................. 33  
   8.1 The framework ...................................... 33  
   8.2 The results ......................................... 34  
   8.3 The outlook ....................................... 35  
Bibliography .................................................. 38  
Author biographies ......................................... 39
Chapter 1

Introduction

On-demand mobility services – autonomous as well as human-driven – promise to transform transportation and urban living. Moving more people with fewer vehicles and kilometers driven can reduce congestion and pollution, enable new levels of productivity otherwise lost in long commutes, and allow us to re-imagine cities around people instead of cars. According to various studies, the travel demand in today’s cities can be served with only a small fraction of the currently used private cars, which most of the time stay parked. Depending on the service design, 9 to 15% of the current fleet size may be sufficient (Bischoff and Maciejewski, 2016), with more densely populated areas requiring fewer vehicles (Burghout et al., 2015) and more wide-spread cities requiring more (Burns et al., 2013). While such benefits are hailed by politicians, industry leaders and journalists, transportation planners and mobility service providers need to answer much more practical questions, such as how these services will work in their cities, with their infrastructure, population, travel demand, and traffic patterns. Ultimately, even more basic questions need to be addressed, such as how many and what type of vehicles are needed, and how the service design should be embarked on.

To answer these questions, we study a broad class of on-demand mobility services that can be summarized as follows. A traveler books a ride from a pickup point to a dropoff point in an area defined by the service provider. Both door-to-door and station-based services are considered. We differentiate between two types of booking, namely instant booking and prebooking. In the former, the traveler wants to be picked up as soon as possible, while in the latter he or she specifies a desired pickup time or a desired dropoff time in the future. Travelers are served by a fleet of vehicles, whether autonomous or human-driven, which may have different seating capacities, battery capacities or other characteristics. A deviation, positive or negative, from a passenger’s desired pickup or dropoff time is referred to as target time deviation. The service design specifies the maximum target time deviation, so that a booking is accepted whenever a vehicle is able to perform a pickup within this time window. The service design may allow for ridesharing as well. With ridesharing, passengers experience longer ride times due to detours. The service design specifies the maximum excess ride time in minutes, as percentage of the direct ride time, or as a combination of both. The same service design will perform differently for different demand distributions, ratios of prebooked versus instant demands, and how long in advance demands are prebooked.

In our study, we focus on the following service design questions:

1. What is the influence of the fleet size on the service’s key performance indicators (KPIs) and what fleet size is necessary to reach a given service level?

2. What is the benefit of ridesharing and how does it depend on the other design choices or the demand properties?

3. How do different service designs affect fleet efficiency and how does that reflect on passenger convenience or inconvenience in terms of target time deviation and excess ride time?

4. What is the value of prebooking on the acceptance rate and on the ability to plan better and provide a more efficient and cost-effective service?
5. What is the potential effect of the service design on externalities such as the reduction of traffic volume, use of parking space and CO₂ emissions?

This report illustrates the process of designing and analyzing on-demand mobility services in the City of Chicago, Illinois. We study how different service designs respond to a variety of travel demand patterns in terms of KPIs such as demand acceptance rate, target time deviation, excess ride time, and a host of fleet utilization and efficiency metrics. To this end, we present the first realistic simulation of on-demand mobility services with extensive KPI-based analysis within a well-defined service typology. We use sophisticated matching and vehicle routing algorithms that can be tailored to the preferences of service providers for improving fleet utilization, passenger service level, or a weighted combination of both. While demand and traffic are treated as exogenous parameters to the simulation, the applied matching and vehicle routing algorithms are commercially available as part of Bestmile’s Mobility Services Platform and used in various real-life operations. Thus, given the demand and traffic patterns, the simulated fleet behavior is realistic and not based on stylized assumptions and simplifications. As a consequence, the obtained insights and conclusions should be representative of real deployments. Yet, changes in travel demand or traffic patterns induced by the introduction of new mobility solutions are out of the scope of this work.

The remainder of this report is organized as follows. Chapter 2 gives an overview and categorization of the scientific literature on the simulation and analysis of on-demand mobility solutions involving both autonomous and human-driven vehicles. Chapter 3 proposes a comprehensive service typology wherein the services discussed here fit and introduces the core concepts, definitions and logic behind our matching, routing, prebooking and traffic models. This is followed by Chapter 4 which describes the structure and mechanics of the simulation framework and provides the list of KPIs we use. Chapter 5 presents the demand data we collected for the City of Chicago and explains how it was processed. Chapter 6 outlines the experimental design and Chapter 7 presents, analyzes and discusses the simulation results. Finally, Chapter 8 concludes with a summary of our findings and most important insights.
Chapter 2

State of the art

In recent years, the number of research articles treating on-demand mobility services, both autonomous and human-driven, has grown considerably. Fagnant and Kockelman (2014) is one of the first extensive simulation studies of autonomous on-demand mobility services in a grid-based city representative of Austin, Texas. The trip generation rates and patterns follow realistic travel profiles in terms of origins, destinations and desired pickup times. The service accepts instant bookings and does not allow for ridesharing. The service level is measured by the acceptance rate and the waiting time, where demands with waiting times of more than 30 minutes are considered rejected.

Fagnant and Kockelman (2014) test a range of scenarios by varying generation rates, trip distribution patterns, fleet size, and other parameters, and conclude that each autonomous vehicle can replace on average eleven vehicles currently in use, thus increasing fleet utilization rate and reducing CO$_2$ emissions and the need for parking space. Yet, due to deadheading, the travel distance for each trip is about 10% longer compared to conventional trips. This result is typical for non-ridesharing services and is supported by numerous studies (Pavone 2015; International Transport Forum 2015; Levin et al. 2017; Hörl et al. 2018). Ridesharing usually more than offsets this figure, leading to a higher acceptance rate and fleet utilization and lower total travel distances (Santi et al. 2014; International Transport Forum 2015; Levin et al. 2017). It is also important to highlight that the amount of deadheading as well as the benefits brought about by ridesharing strongly depend on the matching and routing algorithms used. Fagnant and Kockelman (2014), for instance, simply assign each demand to the closest available vehicle.

Table 2.1 classifies the relevant literature by a number of factors, including the source of the demand data, the road network used in the case study, the type of bookings allowed by the service design, the option of ridesharing, the service level metrics, and the type of matching and routing algorithms used. We observe that most studies use real data, albeit with various assumptions and simplifications, recognizing the importance of drawing conclusions and making recommendations that are practically relevant. The demand data mostly comes from origin-destination (OD) surveys or GPS traces, with taxi data usually falling in the latter category. In many cities, taxis are required to have a GPS device on board and to provide GPS traces and other information to the respective regulatory agency. This data is often made freely available to researchers and the general public. OD surveys contain aggregate information about the direction and volume of traffic flows among the different zones in a city or an urban area, and may have a coarse time resolution such as morning and afternoon peak and off-peak hours. This type of data is usually collected with questionnaires and interviews and can be used to generate trips following the underlying spatio-temporal OD distribution. Geographically, New York (especially the borough of Manhattan) and Singapore are over-represented in the case studies.

All articles, with the exception of Linares et al. (2017), Martinez and Viegas (2017), Santi et al. (2014) and Scheltes and de Almeida Correia (2017), consider instant bookings only. If ridesharing is not supported, demands are often matched to the closest available vehicle, which travels to the pickup point, performs the service, and becomes available once the passenger is dropped off. Waiting time (WT), the instant-booking counterpart of target time deviation, is the most frequently used service level metric. The analyses typically compare input parame-
Simulation-based design and analysis of on-demand mobility services

Table 2.1: Classification of on-demand mobility services literature

<table>
<thead>
<tr>
<th>Article</th>
<th>Demand data</th>
<th>Network</th>
<th>Booking</th>
<th>Ride-sharing</th>
<th>Service level</th>
<th>Matching and routing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azevedo et al. (2016)</td>
<td>OD survey, GPS data</td>
<td>Singapore</td>
<td>instant</td>
<td>no</td>
<td>WT, AR</td>
<td>closest vehicle</td>
</tr>
<tr>
<td>Basu et al. (2018)</td>
<td>synthetic</td>
<td>synthetic</td>
<td>instant</td>
<td>yes</td>
<td>WT, AR, ERT</td>
<td>heuristics</td>
</tr>
<tr>
<td>OD survey, taxi data</td>
<td>Berlin</td>
<td>instant</td>
<td>no</td>
<td>WT</td>
<td>closest vehicle</td>
<td></td>
</tr>
<tr>
<td>OD survey, taxi data</td>
<td>The Hague, core, Prague</td>
<td>instant</td>
<td>no</td>
<td>AR</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>OD survey</td>
<td>area in Melbourne</td>
<td>instant</td>
<td>yes</td>
<td>WT</td>
<td>unclear</td>
<td></td>
</tr>
<tr>
<td>OD survey</td>
<td>grid of Austin (TX)</td>
<td>instant</td>
<td>no</td>
<td>WT, AR</td>
<td>closest vehicle</td>
<td></td>
</tr>
<tr>
<td>OD survey</td>
<td>Zurich</td>
<td>instant</td>
<td>no</td>
<td>WT, price</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>OD survey</td>
<td>Lisbon</td>
<td>instant</td>
<td>yes</td>
<td>WT, ERT</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>OD survey, taxi data</td>
<td>Munich</td>
<td>instant</td>
<td>no</td>
<td>WT</td>
<td>closest vehicle</td>
<td></td>
</tr>
<tr>
<td>OD survey, taxi data</td>
<td>downtown Austin</td>
<td>instant</td>
<td>yes</td>
<td>WT, ERT</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>OD survey, taxi data</td>
<td>downtown Barcelona</td>
<td>both</td>
<td>yes</td>
<td>WT, ERT</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>OD survey, GPS data</td>
<td>Singapore</td>
<td>instant</td>
<td>uncertain</td>
<td>WT, AR</td>
<td>closest vehicle</td>
<td></td>
</tr>
<tr>
<td>OD survey, GPS data</td>
<td>Singapore</td>
<td>instant</td>
<td>uncertain</td>
<td>WT, AR</td>
<td>closest vehicle</td>
<td></td>
</tr>
<tr>
<td>OD survey</td>
<td>Lisbon</td>
<td>both</td>
<td>yes</td>
<td>WT, ERT</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>OD survey, taxi data</td>
<td>Manchester, Singapore</td>
<td>instant</td>
<td>no</td>
<td>WT</td>
<td>queuing theory</td>
<td></td>
</tr>
<tr>
<td>OD survey, taxi data</td>
<td>Manchester</td>
<td>both</td>
<td>yes</td>
<td>WT, ERT</td>
<td>graph theory</td>
<td></td>
</tr>
<tr>
<td>OD survey</td>
<td>Delft Zuid to TU Delft</td>
<td>both</td>
<td>no</td>
<td>WT, AR</td>
<td>closest vehicle</td>
<td></td>
</tr>
<tr>
<td>OD survey</td>
<td>New York</td>
<td>instant</td>
<td>no</td>
<td>WT, AR</td>
<td>heuristics</td>
<td></td>
</tr>
<tr>
<td>OD survey, taxi data</td>
<td>Singapore</td>
<td>instant</td>
<td>no</td>
<td>WT</td>
<td>queuing theory</td>
<td></td>
</tr>
<tr>
<td>OD survey, taxi data</td>
<td>grid of New Jersey</td>
<td>instant</td>
<td>yes</td>
<td>WT, ERT</td>
<td>heuristics</td>
<td></td>
</tr>
</tbody>
</table>

* Ann Arbor (MI), Manhattan, Babcock Ranch (FL)

WT: waiting time, AR: acceptance rate, ERT: excess ride time

In the literature, such as number of vehicles and their positions, the option of ridesharing, etc., in terms of their effect on waiting time. Some of the studies are explicitly interested in the acceptance rate as well, considering that if a demand cannot be served within the service requirements, e.g., a maximum waiting time limit, it is rejected. When ridesharing is allowed (Basu et al., 2018; Dia and Javanshour, 2017; International Transport Forum, 2015; Levin et al., 2017; Linares et al., 2017; Martínez and Viegas, 2017; Santi et al., 2014; Zachariah et al., 2013), the excess ride time (ERT) measures the additional time a passenger spends in the vehicle due to detours from the shortest path induced by picking up or dropping off other passengers. The ERT is normally constrained to an absolute or relative maximum set by the service provider, and its value is an important service level metric related to ridesharing.

Finally, matching and routing algorithms can largely be split into those assigning the closest available vehicle and those exploiting classical heuristic algorithms or modifications thereof. Several exceptions stand out, in particular Spieser et al. (2014) and Pavone (2015), who use queuing theory to analyze the conditions under which waiting times are uniformly bounded, and Santi et al. (2014), who apply a graph-theoretic approach to calculate the savings that can be obtained from sharing taxi rides in New York City subject to a maximum ERT of five minutes. The reduction of total travel distance is estimated at 40%. The rest of the studies use simulation on real, and occasionally synthetic, demand data to measure the performance of the respective service designs.

We can highlight a number of important takeaways from the above literature. Ignoring traffic information may significantly exaggerate the benefits of on-demand mobility services (Levin et al., 2017). If not designed properly, these services may lead to additional kilometers driven due to prepositioning or traveling empty to the pickup points. Hörl et al. (2018) note that the main objective of such systems should be the minimization of total empty pickup distance, or deadheading. On-demand mobility services have substantial positive externalities such as reduction of CO₂ emissions (International Transport Forum, 2015; Fagnant and Kockelman, 2014; Martínez and Viegas, 2017) and the use of parking space (Burns et al., 2013).
Chapter 3
Modeling

In the following sections, we present the main concepts, definitions and modeling elements used in the service design. Section 3.1 first develops a typology and classification scheme of mobility services and identifies the service types we treat in this document. Section 3.2 provides a high-level description of the optimization model with an emphasis on its objective function and constraints. Sections 3.3 and 3.4 outline our prebooking and traffic models, respectively.

3.1 Service typology

Table 3.1 presents a typology of mobility services by dimension (with aspect) and possible service types. We classify mobility services along five dimensions – access, routing, timing, sharing and booking – from the spatial, temporal and passenger points of view. Starting with access, we distinguish between station-based and coordinate-based services. Bus services are a typical example of station-based services, while taxi services usually operate as coordinate-based, or door-to-door, services. When it comes to routing, the most natural separation is between fixed-route services, such as bus services, and dynamic-route services, such as taxis, ridehailing and various other micro-transit services. Some service types may imply others. For example, fixed-route services are usually station-based. Dynamic-route services, on the other hand, can be station-based, coordinate-based or include a mixture of both.

In terms of timing, time-based services, whether schedule-based or frequency-based, are usu-

<table>
<thead>
<tr>
<th>Dimension (aspect)</th>
<th>Possible service types (synonyms in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access (spatial)</td>
<td>• station-based</td>
</tr>
<tr>
<td></td>
<td>• coordinate-based (door-to-door)</td>
</tr>
<tr>
<td>Routing (spatial)</td>
<td>• fixed-route</td>
</tr>
<tr>
<td></td>
<td>• dynamic-route</td>
</tr>
<tr>
<td>Timing (temporal)</td>
<td>• time-based</td>
</tr>
<tr>
<td></td>
<td>• schedule-based (timetable-based)</td>
</tr>
<tr>
<td></td>
<td>• frequency-based (headway-based)</td>
</tr>
<tr>
<td></td>
<td>• on-demand</td>
</tr>
<tr>
<td>Sharing (passenger)</td>
<td>• shared-ride (pooling)</td>
</tr>
<tr>
<td></td>
<td>• individual-ride (without pooling)</td>
</tr>
<tr>
<td>Booking (passenger)</td>
<td>• open-access</td>
</tr>
<tr>
<td></td>
<td>• reservation-based</td>
</tr>
<tr>
<td></td>
<td>• prebooking</td>
</tr>
<tr>
<td></td>
<td>• instant-booking</td>
</tr>
</tbody>
</table>
ally fixed-route services, while on-demand services follow dynamic routes that are re-optimized and evolve with each new demand acceptance. With regard to sharing, we have shared-ride (or pooled) services and individual-ride services. Taxi services, for instance, typically fall into the latter category. Finally, we differentiate between open-access and reservation-based services. Again, bus or train services are the most natural examples of open-access services, whereas most on-demand services are reservation-based. The latter category includes two subtypes – prebooking and instant-booking services. Prebooking services require that travelers book their ride in advance, specifying a desired pickup time at the origin or a desired dropoff time at the destination. In instant-booking services, the traveler wants to be picked up as soon as possible.

In addition to service-related aspects, the design is characterized by its operational aspects as well, in particular regarding the vehicles. Those are presented in Table 3.2. Steering-wise, vehicles can be autonomous or human-driven, and they can be powered by electric, hybrid or combustion engines. The infrastructure can be road- or rail-based. Using these classification schemes, the mobility services that we treat in this document are dynamic-route, on-demand, reservation-based services operating on road-based infrastructure. We do not specify an access attribute as we consider both station-based and coordinate-based services. Likewise, we analyze both shared-ride and individual-ride services, as well as both prebooking and instant-booking reservation-based services, and compare their performance in a series of experiments. From an operational point of view, our service design is independent of the vehicle steering and propulsion mechanisms.

3.2 Demand matching model

Our demand-to-vehicle matching logic relies on bi-objective optimization. The primary objective is the maximization of the number of accepted demands, while the secondary objective is the minimization of the weighted combination of three efficiency and service-level metrics, namely:

1. vehicle movement distance (VMD),
2. excess ride time (ERT), and
3. target time deviation (TTD).

The ERT is a concept that only appears in shared-ride services and, as defined before, is the additional time a passenger spends in the vehicle due to detours from the shortest path. The TTD is the earliness or tardiness with respect to the desired target time. In particular, when requesting a ride, a traveler specifies either a desired pickup time or a desired dropoff time, collectively referred to as target time. Any deviation, positive or negative, from the target time is referred to as target time deviation.

While the VMD and TTD terms are always present, the ERT term only appears when ridesharing is allowed. Minimizing the VMD aims at improving the operational efficiency of the fleet. For the same number of accepted demands, we want to use fewer vehicles that drive shorter distances. However, minimizing the distance driven naturally deteriorates the service level by entailing higher ERT and TTD for some or all passengers. Choosing an appropriate weight for
Simulation-based design and analysis of on-demand mobility services

Figure 3.1: Prebooking distribution for $\delta = 60$ minutes

each of these three terms in the secondary objective allows the mobility service provider to find their preferred balance between the fleet’s efficiency and cost effectiveness on the one hand and the resulting service level on the other.

Let $V$ denote the vehicles in the fleet and $A$ the accepted demands. The secondary objective can then be stated as

$$
\text{minimize } z = \omega_{VMD} \sum_{v \in V} VMD_v + \omega_{ERT} \sum_{a \in A} ERT_a + \omega_{TTD} \sum_{a \in A} |TTD_a|, \quad (3.1)
$$

where $\omega_{VMD}$, $\omega_{ERT}$ and $\omega_{TTD}$ are the weights on the VMD, ERT and TTD, respectively. Expression (3.1) sums over the vehicle movement distances of all vehicles and the excess ride times and target time deviations of all passengers, multiplying each term by its respective weight. We take the absolute value of the target time deviation as the latter may be positive or negative.

For all accepted demands, the matching and the vehicle routes should satisfy a number of hard service-level and operational constraints:

$$
|TTD_a| \leq \text{maxTTD} \quad \forall a \in A \quad (3.2)
$$

$$
ERT_a \leq \text{maxERT} \quad \forall a \in A \quad (3.3)
$$

$$
n_{(v,i)} \leq \text{capacity}_v \quad \forall v \in V, \forall i \in P_v \quad (3.4)
$$

$$
\text{shiftStart}_v \leq t_{(v,i)} \leq \text{shiftEnd}_v \quad \forall v \in V, \forall i \in I_v \cup D_v \quad (3.5)
$$

$$
\epsilon_{(v,i)} \geq \text{minEnergy}_v, \quad \forall v \in V, \forall i \in D_v \quad (3.6)
$$

Constraints (3.2) limit, for each accepted demand, the deviation of the actual pickup or dropoff time from the desired one. The absolute value bounds thisearliness or tardiness within a time window defined by the maximum target time deviation. In a similar fashion, constraints (3.3) enforce a maximum excess ride time that should not be exceeded when ridesharing is allowed. Constraints (3.4) ensure that the occupancy $n_{(v,i)}$ of vehicle $v$ when leaving pickup point $i$ in the set of pickup points $P_v$ on its route does not exceed its capacity. Let $t_{(v,i)}$ denote the time when vehicle $v$ leaves point $i$ on its route, which can either be a dropoff point or belong to the set of initial positions $I_v$. Then, constraints (3.5) establish that vehicle $v$ can only leave point $i$ after the start of its shift and before the end of its shift. Finally, constraints (3.6) impose a minimum threshold on the energy or fuel level $\epsilon_{(v,i)}$ that each vehicle $v$ must have at dropoff point $i$. Specifically, a vehicle must have sufficient energy or fuel to drive to the nearest gas station or charging facility after each planned dropoff.

3.3 Prebooking model

As illustrated in Table 3.1, booking is one of the two dimensions of the passenger aspect in the service typology. As explained in Section 3.1, we treat reservation-based systems that can support both instant booking and prebooking. Given the absence of real-world data on prebooking,
we develop a simple model for this purpose. The prebooking time is defined as the time between
booking a ride and the specified desired pickup time. Let \( \delta \) denote an average prebooking time,
for instance 60 minutes. The individual prebooking times in a given demand sample are then
randomly drawn from an exponential distribution with a rate \( \lambda = 1/\delta \). Figure 3.1 shows such a
prebooking distribution for \( \delta = 60 \) minutes. While the average prebooking time is 60 minutes,
the distribution has a strong positive skew, with most prebooking times in the order of minutes
and a relatively small proportion of long prebooking times. This model is thought to capture a
realistic behavior.

### 3.4 Traffic model

Besides the demand and booking process, traffic is the other major exogenous factor that affects
the performance of mobility services. In order to reflect realistic traffic conditions, we consider
travel times between any two points on the road network as being time-dependent. We express
the time-dependent travel time \( \tau_{o,d,t} \) between an origin location \( o \) and a destination location \( d \)
when starting at time \( t \) as a piecewise linear function defined as

\[
\tau_{o,d,t} = f_{o,d}g_{o,d,t},
\]

where \( f_{o,d} \) is the free-flow travel time between origin and destination and \( g_{o,d,t} \) is a time
dilation factor. The representation of \( g_{o,d,t} \) relies on a suitable discretization of space and time.
Chapter 4
Methodology

We use a simulation-based approach to evaluate the performance of any number of potential service designs defined by their parameter configurations. Section 4.1 describes the simulation framework, and Section 4.2 lists the KPIs for assessing the quality of the tested service designs.

4.1 Simulation framework

Figure 4.1 below is a schematic representation of our simulation framework. At the center is the simulator, which reads the service and fleet design parameters, applies them on the demand process using the matching and routing algorithms, and moves the vehicles on the road network. The logs it generates are used for computing the KPIs, which allow us to judge the resulting quality of service and fleet efficiency. The service design itself comprises the specification of the weights in objective function (3.1), the service area and the service level parameters of (1) maximum excess ride time, (2) maximum target time deviation, and (3) whether ridesharing is enabled or disabled. The fleet design defines each vehicle’s (1) capacity, (2) shift hours, and (3) initial position at the start of the shift. The demand process mimics real-life demand arrivals ordered chronologically by booking time. In addition to (1) the booking time, each demand arrives with (2) a target time (desired pickup time or desired dropoff time), (3) a load, i.e., the number of passengers in the booking, and (4) origin and destination coordinates.

The simulator communicates with two principal actors – the matching and routing algorithms and the map service. The matching and routing algorithms implement the bi-objective optimization model described in Section 3.2. That is, they seek to accept as many demands as possible while minimizing the objective function (3.1). The map service provides time-dependent travel times and distances allowing the simulator to track vehicle positions at all times. The same map service is also used by the matching and routing algorithms. For the purpose of this

Figure 4.1: Simulation framework
study, it relies on OpenStreetMap\(^1\) and a time-dependent adaptation of the OSRM\(^2\) shortest path computation engine, such that travel times and distances depend on the hour of the day as described in Section 3.4. The output of the simulator is a set of KPIs measuring the performance of the service design along the passenger-related and fleet-efficiency dimensions.

The simulation logic is event-based with each time-stamped event triggering an action. Examples of such events are bookings, pickups and dropoffs. The simulation starts with the first demand arrival, which triggers a call with a ride request to the matching and routing engine. The matching and routing engine returns a vehicle mission to the simulator, which then moves the vehicle towards the pickup point until the next event arrives. At that point, the simulator estimates the exact positions of all vehicles using the map service. If the new event happens to be another booking, the simulator issues another call to the matching and routing engine, which returns updated vehicle missions. The simulator continues to iterate through the demand arrivals, and the simulation ends when the last demand has been dropped off. At this point, the vehicle logs are processed and the KPIs presented next are calculated.

4.2 Key performance indicators

We use the simulation experiments to test service design performance, where the latter is evaluated using a set of key performance indicators (KPIs). The KPIs are defined to reflect both service level and system efficiency in order to provide a global overview of how the system performs as well as insights into what can be adjusted to move it to the target performance level. We consider the following KPIs.

- **Acceptance rate**: The percentage of accepted demands out of all demands.
- **Average excess ride time**: The average excess ride time experienced by all passengers. This is an indicator of passenger inconvenience due to ridesharing. Its value is zero if ridesharing is not enabled.
- **Average relative excess ride time**: The average of the excess ride times as percentage of direct ride times. This indicator provides a better understanding of how excess ride time is experienced. For example, an excess ride time of 10 minutes may be acceptable for a direct ride time of 30 minutes, but perhaps less so for a direct ride time of 5 minutes.
- **Average target time deviation**: The average deviation over all passengers of their actual pickup or dropoff time compared to their desired one. This is another indicator of passenger inconvenience.
- **Total vehicle movement distance**: The total distance driven by all vehicles in the fleet.
- **Total vehicle transportation distance**: The total distance driven by all vehicles in the fleet while with passengers on board. This distance is always shorter than the total vehicle movement distance.
- **Vehicle transportation distance ratio**: The ratio of the total vehicle transportation distance to the total vehicle movement distance. This ratio is an indicator of fleet utilization. The closer it is to one, the more efficient the fleet utilization.
- **Total effective vehicle transportation distance**: The sum of the direct ride distances of accepted demands, where a direct ride distance is the shortest distance between a demand’s origin and destination point. For individual-ride services, this metric is equal to the total vehicle transportation distance, while for shared-ride services it is longer.
- **Effective vehicle transportation distance ratio**: The ratio of the total effective vehicle transportation distance to the total vehicle movement distance. This is a measure of the savings in distance driven obtained from ridesharing. Moreover, it is a conservative estimate since it underestimates the amount of deadheading that would have been done with no ridesharing.

\(^{1}\)URL: https://www.openstreetmap.org
\(^{2}\)URL: http://project-osrm.org
• *Average vehicle occupancy:* The average number of passengers in the vehicle while in revenue service or, in other words, total passenger-kilometers over total vehicle transportation distance.
Chapter 5

Demand data

For our case study, we selected the City of Chicago, Illinois. With a population of 2.7 million within its administrative limits, it is the third most populous city proper in the United States. Yet, as revealed in Chapter 2, it has largely been neglected in the literature on simulation of on-demand mobility services, despite its prominence and the data availability.

The demand data comes from taxi trips reported to the City of Chicago in its role as a regulatory agency (City of Chicago 2017). The database contains over 100 million records from January 2013 to August 2017. Each record represents a taxi ride and includes information about which taxi provided the service, the trip’s start and end time, the travel time and distance, and the starting and ending community areas. For high-volume origin-destination pairs, the starting and ending census tracts are also reported. Chicago’s 77 community areas and approximately 800 census tracts are shown in Figures 5.1a and 5.1b, respectively. Community areas are larger and contain multiple census tracts. Exact street-level pickup and dropoff addresses in the database have been masked for privacy reasons. The same holds for the exact pickup and dropoff times.

![Chicago community area boundaries](image1.png)  ![Chicago census tract boundaries](image2.png)

Figure 5.1: Map of Chicago administrative boundaries

3Sources: City of Chicago (2018) for the community area boundaries, United States Census Bureau (2018) for the census tract boundaries, and Google Maps for the map layer.
Simulation-based design and analysis of on-demand mobility services

which are rounded to the nearest 15 minutes. The data has been preprocessed to ensure basic data quality and the City believes that most trips are reported (City of Chicago, 2017).

For our study, we extracted the records corresponding to March 1st, 2017, to simulate a typical day. After removing records with missing community area fields, we end up with 31,254 records. A quarter of these do not report the pickup and dropoff census tracts. According to City of Chicago (2017), the census tracts are only reported if both the starting and ending census tract of a trip appears in at least three trips in the relevant 15-minute time period. If missing, we use the the census tracts that contain the centroids of the pickup and dropoff community areas. To obtain a realistic time granularity, we generate for each demand a desired pickup time by sampling uniformly in the relevant 15-minute time interval. Spatial granularity is achieved by uniform random sampling inside each trip’s pickup and dropoff census tract to obtain the pickup and dropoff coordinates. The resulting pickup and dropoff points are presented in Figures 5.2a and 5.2b respectively. Sometimes, the census tract polygons protrude into Lake Michigan, hence the points that appear in the water. These do not pose a problem as they are snapped to the nearest points in the road network. Finally, because the data does not reveal the number of passengers in each trip, we assumed that all trips are single-person trips.

(a) Pickup points

(b) Dropoff points

Figure 5.2: Pickup and dropoff points of all demands

4Sources: Produced with ggmap (Kahle and Wickham, 2013) using Google Maps for the map layer.
Chapter 6
Experimental design

We simulate three distinct on-demand mobility service scenarios in Chicago. The differences in their geographic extents and their travel demand volumes, patterns and spatial distributions create an ideal testing ground for the versatility and robustness of our matching algorithms and for demonstrating the simulation-based service design and analysis approach.

6.1 Service areas

Lincoln Park Micro-transit: The first case study envisions an on-demand mobility service within the adjacent community areas of Near North Side and Lincoln Park. These two communities are located in north central Chicago and contain a mixture of business and residential districts as well as entertainment and tourist attractions such as the Lincoln Park Zoo, the Magnificent Mile, and the Navy Pier. The extent of this area, which measures approximately four by eight kilometers, is depicted in red in Figure 6.1a. For March 1st, 2017, our dataset has 4458 taxi trips entirely contained inside this area. This number excludes trips that are reported to have...
Simulation-based design and analysis of on-demand mobility services

started and ended within the same census tract as we consider such trips unreasonably short. The moderate size of this area makes it a good candidate for phasing in on-demand micro-transit. Given the compactness of this service area, we also explore how the KPIs react when the door-to-door service is replaced by a station-based service with 40, 80 and 250 stations, resulting in an average walking distance of 300, 200 and 100 meters per passenger, respectively.

**O’Hare Shuttle:** The second service area covers the demands to and from Chicago’s main airport, O’Hare International, shown in red in Figure 6.1. The airport comprises the major part of the O’Hare community area. For March 1st, 2017, our dataset has 2633 taxi trips to and from the airport. Given the strong directionality of the travel demand in this service area, we would like to examine the effect of ridesharing on the number of vehicles needed to serve this demand. Considering the long travel distances from the city to the airport, we would also expect that a service that requires prebooking will perform much better in terms of its acceptance rate and the various vehicle-level KPIs. Globally, due to much longer travel distances, we would expect to see a higher number of vehicles needed and a more pronounced benefit of ridesharing here compared to the Lincoln Park Micro-transit.

**City Ridesharing:** The final case study includes the whole city of Chicago, within the administrative limits shown in Figure 6.1. For March 1st, 2017, our dataset has 31,254 trips performed by 2711 taxis. We consider a city ridesharing service that aims to cover 50% of this number, or 15,627 demands, obtained by a uniform random sampling from the original trips. We would expect to significantly reduce the number of vehicles needed, with the remark that we do not know how long the shifts of these taxis are, how they compete, or what they do when not serving customers. Moreover, we would like to examine the synergy effects arising from a denser vehicle coverage of the city, suspecting that a linear increase in the number of demands would require a less-than-linear increase in the number of vehicles needed to maintain the same service level.

### 6.2 Service parameters

We test a range of service parameters in each of the areas described in Section 6.1 in order to examine how the KPIs respond to various service designs. The purpose of this sensitivity analysis is not necessarily the choice of the best service design. This choice is the responsibility of the service provider. Our goal is rather to demonstrate the trade-offs between the fleet size and its cost effectiveness, and the acceptance rate and passenger-related KPIs, such as target time deviation and excess ride time in the case of ridesharing. These trade-off are specific to the service area and its demand pattern.

We consider six-seater vehicles with shifts covering the whole 24-hour period of March 1st, 2017, with a two-hour margin on both sides. For the O’Hare Shuttle case study, all vehicles are positioned at the airport at the beginning of the shifts as virtually all travel demand in the first hours of the day is from the airport to the city. For the Lincoln Park Micro-transit, there is an equal number of vehicles positioned on West Chicago Avenue and West Armitage Avenue at the beginning of the shifts.

In terms of the variable vehicle parameters, we test the service with

- 10, 20, 30, 40, 50 and 60 vehicles for the Lincoln Park Micro-transit,
- 25, 50, 75 and 90 vehicles for the O’Hare Shuttle, and
- 60, 80, 100, 120, 140 and 160 vehicles for the City Ridesharing.

For each fleet size, we examine the effects of

- enabling and disabling ridesharing, and
- varying the maximum pickup time deviation between five and ten minutes.

Regarding the last point, all demands in the experiments are specified in terms of desired pickup time. Hence, in the analysis in Chapter 7, we refer to average or maximum pickup time deviation as opposed to the more general terms of average or maximum target time deviation. When ridesharing is enabled, the maximum excess ride time
• for the Lincoln Park Micro-transit service area is set to 50% of the direct ride time plus two minutes,
• for the O’Hare Shuttle service area is set to 50% of the direct ride time plus ten minutes in view of its much longer average ride times, while
• for the City Ridesharing service area both values are tested.

Finally, we apply the same weight in objective function (3.1) to the vehicle movement distance, the excess ride time and the deviation from the desired pickup time to simulate a balanced setup.

6.3 Demand patterns

To examine how the service designs perform under different travel demand characteristics, each of the service parameter configurations described in Section 6.2 is tested for a number of different demand patterns. To accomplish this, the pickup time in the taxi data is taken as the desired pickup time. We emulate different prebooking distributions by

• varying the ratio of prebooked demands as 0, 50, and 100%, and
• varying the average prebooking time (i.e., the time between booking and desired pickup time) as 60 and 120 minutes.

As described in Section 3.3, the individual prebooking times themselves are drawn from an exponential distribution with a rate \( \lambda \) equal to the inverse of the average prebooking time.
Chapter 7

Simulation results

In this chapter we describe, analyze and discuss the results of the simulation experiments. Each of the three service area scenarios – the Lincoln Park Micro-transit, door-to-door and station-based, the O’Hare Shuttle and the City Ridesharing – is discussed in a separate section, after which some general topics comparing the scenarios are addressed. For each scenario, the results are presented along the main KPIs, starting with passenger-related KPIs – acceptance rate, excess ride time and pickup time deviation – followed by vehicle-related and fleet-efficiency KPIs – vehicle movement distance, transportation distance ratio, effective transportation distance ratio and occupancy. In each case, we examine how the interplay of the various service design parameters and the demand model properties affect the KPIs.

7.1 Door-to-door Lincoln Park Micro-transit

The Lincoln Park Micro-transit covers a compact service area. In addition, the demand data reveals that the average direct ride time here is in the order of five minutes. These two factors have a strong and favorable impact on the number of vehicles needed to serve the area. The sections below provide a detailed analysis for the service designs presented in Section 6.2 as applied to a door-to-door service. The service designs are tested for 0, 50 and 100% prebooked demands with an average prebooking time of 120 minutes. The maximum excess ride time is set to 50% of the direct ride time plus two minutes.

7.1.1 Acceptance rate

Figure 7.1: Acceptance rate for door-to-door Lincoln Park Micro-transit

Figure 7.1 shows the acceptance rate as a function of the fleet size for different prebooking ratios. The extremes of the vertical bars at each data point indicate the values for a maximum pickup time deviation of five and ten minutes, with the marker showing the mean of the two values. Unsurprisingly, the higher acceptance rates always correspond to a maximum pickup time deviation of ten minutes.
We observe that ridesharing improves the acceptance rates by roughly 10 to 15 percentage points for small fleet sizes. These improvements fade as the fleet is expanded, in particular as the acceptance rate approaches 100% and half or more of the rides are prebooked. Prebooking proves to have a positive effect on the acceptance rate. When all rides are prebooked, a fleet of 20 vehicles can already achieve an acceptance rate close to 90% with ridesharing and above 70% without ridesharing.

Furthermore, prebooking reduces the sensitivity of the acceptance rate to the maximum pickup time deviation. Without prebooking, the range of the acceptance rate with respect to the maximum pickup time deviation is often more than 20 percentage points, as indicated by the long vertical bars. This range narrows down significantly as the prebooking ratio goes up. When all rides are prebooked, the maximum pickup time deviation hardly matters anymore. This would suggest that in order to achieve high acceptance rates, either the pickup time window must be wide enough or ride requests must be known with sufficient lead time so as to guarantee a good service with a small fleet.

7.1.2 Excess ride time

The benefits of ridesharing in terms of higher acceptance rates come at the expense of excess ride times for the passengers. The maximum excess ride time is specified as an input parameter and guarantees a minimum service level. The actual excess ride times are summarized in Figure 7.2 and appear to be relatively stable across the different prebooking configurations, with values around two minutes. Without prebooking, excess ride times exhibit a slight decrease for larger fleet sizes, explained by the reduced number of shared rides. With prebooked rides, we observe the opposite effect. The higher excess ride times here are the traded for shorter vehicle movement distances – a consequence of the flexibility afforded by a larger fleet and sufficient lead times to make better plans.

7.1.3 Pickup time deviation

Figure 7.3: Average pickup time deviation for door-to-door Lincoln Park Micro-transit
Figure 7.3 plots the average pickup time deviation. It improves with an increasing fleet size due to the greater choice of vehicles to match a pickup at the desired time. This effect is less pronounced with ridesharing because, again, slightly higher pickup time deviations are traded for shorter vehicle movement distances. Interestingly, without prebooking, ridesharing results in consistently lower pickup time deviations. The narrow pickup time window of five minutes corresponds to better values as it already restricts the maximum allowed pickup time deviation for any given ride. With an increasing prebooking ratio, the pickup time deviation improves when ridesharing is not allowed, while the improvement is less marked with ridesharing. At the extreme, for a prebooking ratio of 100%, the pickup time deviation without ridesharing is consistently better.

7.1.4 Vehicle movement distance

![Vehicle movement distance](image)

The trade-off just discussed in Section 7.1.3 is clearly visible in Figure 7.4. The total movement distance is substantially shorter with ridesharing despite the higher demand acceptance rates. This efficiency gain can be attributed to the actual ridesharing effect, but also to the cutback on deadheading. With ridesharing, the slight contraction of total movement distance observed when the fleet is expanded beyond 30 or 40 vehicles is explained by the added flexibility in constructing more efficient trips when more vehicles are available.

7.1.5 Vehicle transportation distance ratio

![Vehicle transportation distance ratio](image)

Figure 7.5 shows the vehicle transportation distance ratio, which is the distance driven in revenue service, i.e., the vehicle transportation distance, divided by the vehicle movement distance. In this KPI, the difference to 1 on the vertical axis represents the fraction of the movement distance spent deadheading. Without ridesharing, we observe ratios in the range of 60 to 75%, with larger fleet sizes alleviating the problem of deadheading thanks to the denser vehicle distribution over the service area. Ridesharing consistently pushes this KPI to values in the range of 75 to 85%.
Yet, some of these gains are undoubtedly at the expense of excess ride distances experienced by passengers. The higher acceptance rates with prebooking, coupled with longer vehicle movement distances, lead to lower vehicle transportation distance ratios.

### 7.1.6 Effective vehicle transportation distance ratio

To correct for excess ride distances, Figure 7.6 shows the effective vehicle transportation distance ratio, which is the sum of the direct ride distances of all accepted demands divided by the total vehicle movement distance. Without ridesharing, the values remain around 0.6 to 0.7, with slightly better values for larger fleets. With ridesharing, they grow more visibly from slightly below 0.9 to almost 1.2, in correlation with the prebooking ratio, the fleet size and the maximum pickup time deviation. In effect, each vehicle-kilometer driven is able to deliver up to 1.2 kilometers of effective (direct) transportation distance. That is, a shared-ride service can be almost twice as efficient in moving passengers than the alternative.

### 7.1.7 Occupancy

Figure 7.7 plots the average vehicle occupancy during revenue service. With ridesharing, it ranges between 1.6 and 2 passengers. As already observed in Section 7.1.2, the occupancy slightly deteriorates when expanding the fleet size with no prebooking while it shows a moderate improvement when at least half of the rides are prebooked. This again shows that prebooking allows for better planning and fleet utilization.

### 7.2 Station-based Lincoln Park Micro-transit

The small size of the Lincoln Park area makes it suitable for a station-based service where walking distances to and from the stations can be kept reasonably low. Traveling among a
Simulation-based design and analysis of on-demand mobility services

limited number of stations can save vehicle kilometers. Moreover, a station-based design should be more advantageous for pooling as passengers gather at the same spots. Effectively, higher fleet efficiency here is exchanged for the inconvenience of walking.

To compare station-based and door-to-door designs, we cluster the pickup and dropoff points so that the average walking distance – origin to pickup station plus dropoff station to destination – per passenger is 100, 200 or 300 meters. These values yield configurations with 250, 80 and 40 stations displayed in Figures 7.8a, 7.8b and 7.8c, respectively. Not all KPIs are significantly affected by the shift to a station-based service. The sections below examine the effect on the acceptance rate and the fleet efficiency metrics.

7.2.1 Acceptance rate

Figure 7.9 suggests that with ridesharing, the acceptance rate of a 10-vehicle fleet can be improved by up to ten percentage points by moving to an 80-station design, and by a similar amount by further reducing the number of stations to 40. Interestingly, replacing the door-to-door service with 250 stations has no visible effect on the acceptance rate. Without ridesharing, the station-based design does not appear to significantly improve the acceptance rate.

7.2.2 Vehicle movement distance

Figure 7.10 demonstrates that there are savings in vehicle movement distance both with and without ridesharing. The amount of savings correlates positively with the fleet size and the number of stations. Yet, with ridesharing, the savings are larger nominally, much larger in

Figure 7.8: Station-based service designs: the station locations are marked in orange; the origin and destination locations of all demands in blue.
relative terms, and clearly seen already with fleets of 20 or 30 vehicles. Replacing the door-to-door service with 80 stations results in vehicle movement distance savings of up to 1000 kilometers, or more than 10%, for larger fleets.

7.2.3 Vehicle transportation distance ratio

Because the station-based design does not lead to higher acceptance rates for the individual ride service, even minor savings in total movement distance translate to tangible cutbacks on deadheading. Figure 7.10 shows that, as expected, the individual-ride services boast similar, or slightly higher, improvements in the vehicle transportation distance ratio. These improvements seem to be rather uniform across different fleet sizes.

7.2.4 Effective vehicle transportation distance ratio

Figure 7.10: Vehicle movement distance for station-based Lincoln Park Micro-transit
However, the impact on fleet utilization is much more clear-cut when looking at the effective vehicle transportation distance ratios in Figure 7.12. Here, the station-based design leads to significant efficiency improvements only with ridesharing, and only when the number of stations is small, i.e., 40 or 80. That is to say, replacing a door-to-door service with a dense network of stations increases passenger inconvenience while bringing little benefit to the service provider. The efficiency improvements also correlate with the prebooking ratio, confirming the importance of advance information in making better decisions.

7.2.5 Occupancy

The previous conclusion is also confirmed by Figure 7.13, which shows that a service with 40 or 80 stations is indeed able to boost vehicle occupancy, while a service with 250 stations has almost no effect. The benefit of prebooking is clearly visible as well.

7.3 O’Hare Shuttle

The sections below present the results for the O’Hare Shuttle service in the same fashion and sequence as those for the Lincoln Park Micro-transit. Again, the service designs presented in Section 6.2 are applied to a door-to-door service and tested for 0, 50 and 100% prebooked demands with an average prebooking time of 120 minutes. Given the much longer travel times in this service area, the maximum excess ride time is raised to 50% of the direct ride time plus ten minutes.

7.3.1 Acceptance rate

Figure 7.14 plots the acceptance rate as a function of the fleet size for different prebooking ratios. Due to the long travel times to and from the airport, ridesharing has a strong positive
impact, much more pronounced compared to the Lincoln Park Micro-transit. Without ridesharing, acceptance rates of 50% are hardly reachable with a fleet of 60 vehicles, even with a high prebooking ratio. When enabled, ridesharing makes it possible to achieve an acceptance rate of 80% already with a fleet of 45 vehicles and moderate prebooking. When all rides are prebooked, even a 90% acceptance rate can be ensured with only 45 vehicles.

7.3.2 Excess ride time

Figure 7.15: Average excess ride time for O'Hare Shuttle

Figure 7.15 reveals that the excess ride times of the ridesharing services strongly depend on the prebooking ratio. They range from seven minutes without prebooking to roughly 12 minutes when all rides are prebooked. This should be seen in conjunction with the higher acceptance rates, and the benefits visible in the other KPIs, that a higher prebooking ratio brings about thanks to the ability to plan better with advance booking information. Nonetheless, the ability to serve significantly more passengers to and from the airport, a ride of approximately 40 minutes, entails only moderate excess ride times.

7.3.3 Pickup time deviation

Figure 7.16: Average pickup time deviation for O’Hare Shuttle

As demonstrated by Figure 7.16 and similar to the Lincoln Park Micro-transit, the long vertical bars indicate that the average pickup time deviation strongly depends on the width of the pickup time window as defined by the maximum pickup time deviation of five or ten minutes. The pickup time deviation is quite stable for the pooled service across different fleet sizes and prebooking ratios, ranging from 2.5 minutes for the narrow pickup time window to around five minutes for the wide pickup time window. Without ridesharing, the pickup time deviation depends on the prebooking ratio. It ranges from three to seven minutes without prebooking and narrows down to a range of one to three minutes for a medium to high prebooking ratio.
7.3.4 Vehicle movement distance

The total vehicle movement distance should be seen in conjunction with the acceptance rate. As clearly shown in Figure 7.17, the vehicle movement distance without ridesharing is positively correlated both with the fleet size and the prebooking ratio. Contrarily, with ridesharing, the vehicle movement distance is almost constant, which would suggest that the additionally accepted demands are handled by vehicle trips that are happening anyway and underlines the benefit of sharing rides with the same vehicle. The slight increase is mainly due to the extra movements when picking up or dropping off the additional passengers. Without ridesharing, the increase of the vehicle movement distance is essentially proportional to the increase in the acceptance rate with slight savings due to reduced deadheading with larger fleets.

7.3.5 Vehicle transportation distance ratio

Due to the long travel distances in this service area, the proportion of deadheading is relatively low, resulting in high vehicle transportation ratios as illustrated in Figure 7.18. Again, ridesharing leads to higher ratios, but the difference is very small when there is no prebooking and grows with the prebooking ratio. Similar to the Lincoln Park Micro-transit, the actual ratios themselves diminish with more prebooking. As more demands are accepted due to advance information, they tend to be spread out, thus causing more deadheading. With ridesharing, this effect is mitigated at the expense of excess ride distances for some passengers.

7.3.6 Effective vehicle transportation distance ratio

The plots of the effective vehicle transportation distance ratio in Figure 7.19 clearly demonstrate the advantage of ridesharing in performing more effective transportation work per vehicle-kilometer. The ratios are between three and four even without prebooking. As the prebooking ratio grows, the effective vehicle transportation distance also exhibits a small increase, but more importantly it shows a major reduction of variability with respect to the maximum pickup time deviation. At low prebooking ratios, widening the pickup time windows affords more flexibility.
Simulation-based design and analysis of on-demand mobility services

in constructing more efficient vehicle trips. When the prebooking ratio increases, this flexibility
is not so crucial anymore.

7.3.7 Occupancy

As the high effective vehicle transportation distance ratio would suggest, ridesharing leads to
a high average occupancy as well. Figure 7.20 indicates that the average occupancy is around
four passengers without prebooking and reaches up to five when all rides are prebooked. This
implies that the vehicle capacity constraint of six seats may be binding sufficiently often for
higher capacities to be considered in this scenario.

7.4 City Ridesharing

To complement the previous analysis, this section offers a slightly different perspective. All of
the experiments are carried out for a door-to-door service with a prebooking ratio of 50%, an
average prebooking time of 60 minutes, and a maximum pickup time deviation of ten minutes.
What varies instead is the maximum excess ride time. For the Lincoln Park Micro-transit, we
fixed that to 50% of the direct ride time (DRT) plus two minutes. For the O’Hare Shuttle,
we raised it to 50% of DRT plus ten minutes, in view of the longer travel times. Since the
City Ridesharing service area includes a mixture of the two, it is instructive to test both these
maximum ERT constraints and examine their impact on the KPIs.

7.4.1 Acceptance rate

Figure 7.21 shows how the acceptance rate develops with increasing the fleet size for the tight
(left) and the relaxed (right) maximum ERT constraint. Relaxing the maximum ERT helps
increase the acceptance rate by about 5 percentage points for smaller fleets but this effect almost
disappears as more vehicles are added. When the tight maximum ERT is applied, 140 vehicles
would be needed to reach an acceptance rate of over 90%. With the relaxed maximum ERT, this
is already possible with 120 vehicles. In both cases, a fleet of 160 vehicles accepts roughly 95% of
the demands, with each vehicle serving on average 93 demands per day. The marginal acceptance
rate shows slightly diminishing returns, which set in earlier with ridesharing due to the higher
acceptance rate reached. For a fleet of 160 vehicles, the demands that could not be accepted are
typically long rides, including to and from Chicago’s two airports - O’Hare and Midway. As a
consequence, we observe a weaker saturation effect compared to the Lincoln Park Micro-transit
where rides are short. At the other extreme, having only long rides with strong directionality,
the O’Hare Shuttle scenario exhibits no saturation effect. When it comes to ridesharing per se,
it is instrumental in boosting the acceptance rate by 20 to 25 percentage points for small fleets
and by 5 to 15 percentage points for large fleets.

7.4.2 Excess ride time

Figure 7.22 illustrates the effect of the maximum ERT constraint on the actual ERT. With the
tight maximum ERT, the average excess ride time is just over three minutes, or roughly 45%
of the DRT, and very stable. When the maximum ERT is relaxed, the average excess ride time
grows to 6-6.5 minutes, or 110-120% of the DRT, exhibiting a slight decrease when more vehicles
are added to the fleet. This difference is considerable compared to the minor improvements in
the acceptance rate. Not to mention that the high relative ERTs under the relaxed maximum
ERT constraint may be unacceptable to many passengers. And while the acceptance rate is
almost unaffected, higher ERTs, as we will see below, are compensated by major savings in
vehicle movement distance and deadheading.

7.4.3 Pickup time deviation

Figure 7.23 shows an interesting pattern. For a fleet of 60 vehicles, the pickup time deviation
with and without ridesharing is essentially the same, equal to approximately 5.6 minutes. When
ridesharing is enabled, adding more vehicles generally helps reduce the pickup time deviation,
with the exception of a slight initial increase under the tight maximum ERT. This is in line with
Simulation-based design and analysis of on-demand mobility services

Figure 7.23: Average pickup time deviation for City Ridesharing

what we observe in the Lincoln Park Micro-transit as well – more vehicles allow greater flexibility for passengers to be picked up closer to their desired pickup times. Without ridesharing, the pickup time deviation grows for a fleet of up to 100 vehicles. The denser vehicle distribution allows for higher acceptance rates but at the expense of pickup time deviation. As soon as a certain saturation level is reached, the additional vehicles help reduce the pickup time deviation and accept longer rides. Relaxing the maximum ERT constraint leads to an improvement of 10 to 15 seconds in the pickup time deviation for the pooled service. Generally speaking, the nominal differences are small, with the pickup time deviation staying in the range of five to six minutes.

7.4.4 Vehicle movement distance

Figure 7.24: Vehicle movement distance for City Ridesharing

As illustrated in Figure 7.24, the vehicle movement distance grows almost proportionally with the acceptance rate when there is no ridesharing and exceeds 80,000 kilometers for an acceptance rate of under 90%. The distance savings with ridesharing are impressive – from 5000 kilometers for a fleet of 60 vehicles to 30,000 kilometers for a fleet of 160 vehicles, and with major improvements in the acceptance rate at the same time. If the maximum ERT is relaxed, the savings approach 40,000 kilometers, or almost 50% of the vehicle movement distance without ridesharing. The 5000 kilometers of additional savings resulting from the relaxation of the maximum ERT should be seen in conjunction with the doubled values of the average ERT experienced by the passengers.

7.4.5 Vehicle transportation distance ratio

As shown in Figure 7.25, relaxing the ERT constraint pushes the vehicle transportation distance ratio from around 85 to above 90%. The advantage of a larger fleet able to cover the service area more efficiently with less deadheading is visible in both cases.
7.4.6 Effective vehicle transportation distance ratio

Figure 7.26: Effective vehicle transportation distance ratio for City Ridesharing

Figure 7.26 presents the effective vehicle transportation distance ratio, which exhibits a mild improvement as more vehicles are added to the fleet. This effect is more noticeable under the relaxed maximum ERT. The maximum ERT relaxation itself produces a more tangible effect, lifting the ratio of the pooled service from around 1.5 to 1.6-1.8. Globally, the pooled service is more than twice as effective in transporting passengers per vehicle-kilometer as the non-pooled service.

7.4.7 Occupancy

A similar effect is observed for the average vehicle occupancy. Figure 7.27 shows that it is rather stable across different fleet sizes but improves significantly when the maximum ERT constraint is relaxed, jumping from 2.3 to almost 3 passengers. Again, we observe that improved fleet
efficiency, this time brought about by a laxer maximum ERT rule, comes at the expense of passenger inconvenience in terms of a much higher average ERT. It is up to the service provider to determine the level of service to offer, to price differently the two service levels, or to explore the middle ground between these two extremes.

7.5 Comparison of the service areas

Since the City Ridesharing scenario is essentially a mixture of the Lincoln Park Micro-transit and the O’Hare Shuttle, it is expected that its KPIs somehow reflect this. Although the many additional demands throughout the city are different, they can be thought of as falling in-between these two extremes of very long and very short rides. The mixture effect is evident in both the passenger-related KPIs, such as the excess ride time, and some of the fleet-related KPIs, including the transportation distance ratio, the effective transportation distance ratio and the occupancy.

This is reflected in the acceptance rate as well. While we see a relatively fast supply saturation with diminishing returns in the Lincoln Park Micro-transit and to a certain degree in the City Ridesharing, this is not the case for the O’Hare Shuttle, which is dominated by much longer rides with strong directionality over an extensive service area. Indeed, the City Ridesharing area includes some of these demands due to the 50% sampling rate, but the rest of the demands are mostly concentrated in the central Chicago communities.

On the other hand, the City Ridesharing example shows that a large service area can be served much more efficiently. While a fleet of 90 vehicles is needed to approach an acceptance rate of 90% for the O’Hare Shuttle under a prebooking ratio of 50%, for City Ridesharing 120 vehicles achieve an acceptance rate of over 90% under the same conditions while serving six times as many demands. Even if we consider all 31,254 demands originally present in our data for March 1st, 2017, with ridesharing a fleet of 200 vehicles achieves an acceptance rate of around 90% with only moderate prebooking. Once again, this underscores the strong synergy effects arising from the coordinated and optimized operation of large fleets over extensive service areas.
Chapter 8

Conclusions

On-demand mobility services promise to change transportation as we know it, but their full integration into urban dynamics and lifestyle is yet to be seen. Large-scale deployments, the ideal way to evaluate the potential of such services, are difficult due to operational and design-related uncertainties. Questions regarding travel demand patterns, integration of traffic information, service level requirements, fleet composition, the benefits of ridesharing, and how these factors mix together and influence one another, and the overall system's efficiency and reliability as a result, are just as pressing, and perhaps even more so, than the vehicles' technical characteristics and their ability to eventually drive themselves.

8.1 The framework

In an effort to bridge the gap between theory and practice, this report proposes a comprehensive typology of shared mobility services and identifies the types of services we address here, namely dynamic-route, on-demand, reservation-based services operating on road infrastructure. Routes are dynamic, as opposed to fixed, because bookings arrive continuously and are matched to vehicles, which are then routed in real time using advanced algorithms. Our custom-built matching and routing algorithms balance between improving system efficiency and fleet utilization on the one hand, and passenger convenience in terms of minimum pickup delays and excess ride times on the other. The balance between these two inherently conflicting objectives is expressed by a set of parameters that service providers can tune to their preferences. Reservations refer to the fact that rides are booked through an application, either requesting an immediate service – instant booking – or a service in the future – prebooking. Furthermore, the services we consider can operate both in a station-based network and as door-to-door services.

Service design includes the setting of the high-level objectives of cost efficiency versus passenger convenience, or the sought balance thereof. It also involves the determination of the fleet composition and the service level parameters such as pickup time windows, ridesharing versus no ridesharing, maximum excess ride time, etc. Adding to this complexity is the fact that the same service design may perform very differently under different travel demand patterns. To measure this, we formulate a set of key performance indicators (KPIs) that evaluate both passenger- and fleet-related performance metrics, such as acceptance rate, pickup earliness or delay, deadheading, vehicle occupancy, and others. Deployments of shared mobility services are expensive, and choosing the wrong service design may have serious financial implications.

This work presents the first realistic simulation framework of on-demand mobility services with extensive KPI-based analysis within a well-defined service typology. It illustrates the process of designing and analyzing multiple service types in Chicago, Illinois. We demonstrate how designs respond to changes in the travel demand patterns, how this is reflected in the KPIs, and what fundamental trade-offs appear between passenger- and fleet-related KPIs. The matching and routing algorithms used in the simulation framework are commercially available as part of Bestmile’s Mobility Services Platform and currently used in various real-life operations. Therefore, given the demand and traffic patterns, the simulated fleet behavior is realistic and not based on stylized assumptions and simplifications.

© Bestmile 2018
8.2 The results

Coming back to the fundamental research questions we started with, we observe that the fleet size has a strong influence on many of the KPIs. Unsurprisingly, acceptance rates improve as more vehicles are added to the fleet. Yet, diminishing returns may quickly settle in, especially when the service area is compact.

A case in point is the Lincoln Park Micro-transit. If we take the moderate prebooking ratio of 50%, increasing the fleet size from 10 to 20 vehicles boosts the average acceptance rate from 40 to 70% without ridesharing, and from just over 50 to almost 85% with ridesharing. Another 40 vehicles would be needed, both with and without ridesharing, to bring the acceptance rate close to 100%. The rationale here is that the additional vehicles are used to improve the service level of the accepted demands by reducing the pickup time deviation and excess ride time values, which may be too high when the fleet is small. This analysis allows service providers to choose their preferred configuration. The same saturation effect does not occur in the O'Hare Shuttle service area, or at least its rate is much slower. Here, each 25 additional vehicles improve the acceptance rate by 7-8% without ridesharing, and by about 2% with ridesharing. The lower value in the latter case is due to the fact that the acceptance rates with ridesharing are already much higher, reaching almost 100% when the fleet is 90 vehicles strong and all demands are prebooked.

The different response of the acceptance rate to the fleet size in the Lincoln Park Micro-transit and the O'Hare Shuttle is undoubtedly caused by their very different geographies and demand patterns. While the Lincoln Park Micro-transit covers a small service area with short ride times, which is conducive to fast supply saturation and offers only a limited potential for pooling rides, the O'Hare Shuttle serves an extensive area where passengers experience long rides. The synergy effect brought about by a dense vehicle coverage is also clearly seen when comparing the number of vehicles needed to reach the same acceptance rate for the O'Hare Shuttle and the City Ridesharing. While the latter sees an acceptance rate in excess of 90% with only 120 to 140 vehicles, the former requires at best 45, even when all demands are prebooked, despite the six-time lower total number of demands. The explanation is that while the City Ridesharing scenario includes approximately half of the 2633 demands to and from O'Hare due to the 50% sampling rate, the rest of the demands are mostly concentrated in the city core, which creates the synergy effect just discussed.

Ridesharing proves to be crucial for increasing both the acceptance rate and the fleet efficiency. For the O'Hare Shuttle, ridesharing boosts the acceptance rate by roughly 40 percentage points for a fleet of 45 vehicles, and 30 percentage points for a fleet of 90 vehicles. When all demands are prebooked, ridesharing makes the difference between 67 and 97% acceptance rate, while keeping the excess ride time in the order of 12 minutes, or approximately 30 to 35% of the direct ride time. The pickup time deviation is slightly more than a minute longer when compared to an individual-ride service. At the same time, the efficiency impacts are considerable. While the vehicle movement distance remains almost constant at around 20,000 kilometers with ridesharing regardless of fleet size, without ridesharing it seems almost perfectly correlated with the number of accepted demands and exceeds 60,000 kilometers for a mere 67% acceptance rate. Not surprisingly, ridesharing leads to an effective transportation distance ratio of 3.5, reflected by an average occupancy of four to five passengers. The effect of ridesharing is less pronounced for the Lincoln Park Micro-transit. Here, ridesharing leads to an improvement of 10 to 15 percentage points in the acceptance rate with small fleet sizes but this difference quickly disappears as we add more vehicles. The average excess ride time of two minutes translates to rides that are 50% longer. Nonetheless, the vehicle movement distance for a fleet of 40 vehicles is already 30% shorter. The effect on the transportation distance ratio and the effective transportation distance ratio is milder compared to the O'Hare Shuttle, the average occupancy being just under two passengers. The shorter rides in the Lincoln Park Micro-transit area make ridesharing less advantageous. Yet, as we have seen, shifting to a station-based service may help improve the ridesharing potential at the expense of walking. The City Ridesharing service area, having common features with both of the above, falls in the middle, with ridesharing resulting in an effective transportation distance ratio of 1.5 to 1.8, an average vehicle occupancy of 2.3 to 3 passengers, and savings in vehicle movement distance similar or slightly higher, in relative terms, to those in the Lincoln Park Micro-transit.
Undoubtedly, ridesharing affects passenger convenience due to excess ride times and other more intangible factors such as loss of privacy. One would also expect that it deteriorates pickup time deviation as well. Yet, our results consistently show that in the prevalence of instant bookings, or for lower values of the average prebooking time as in the City Ridesharing, ridesharing improves pickup time deviation. It appears that in the lack of advance booking information, pooling allows passengers to be picked up closer to their desired times with the same fleet size as compared to a service without pooling. This relationship is reversed with higher prebooking ratios thanks to the ability to better exploit the pickup time window for improving fleet efficiency when ridesharing is enabled. While the excess ride time and the pickup time deviation are the main concerns of passengers, potential travelers may be discouraged from using a service if they perceive a low acceptance rate. The fleet size and the ratio of prebooked demands emerge as the main influences on the acceptance rate. What is more, prebooking also acts to reduce the variability of the acceptance rate and the rest of the KPIs for the same fleet size. The longer in advance rides are booked, the easier it is to plan them with better service levels and shorter additional distances driven.

Finally, the analysis of the simulation results provides important insights into the fundamental trade-offs between the passenger- and fleet-related KPIs. This is most clearly seen in the relationship between fleet efficiency, as measured by the vehicle movement distance, the vehicle transportation distance ratio and the effective vehicle transportation distance ratio on the one hand, and the excess ride time on the other. In all three service areas, higher excess ride times correspond to lower vehicle movement distances and higher ratios – indicators of a more efficiently utilized fleet. The relationship between fleet efficiency and the pickup time deviation appears to be more complex. While for higher prebooking ratios we observe the expected effect, i.e., better fleet efficiency comes at the expense of higher pickup time deviations, for lower prebooking ratios it is the inverse. As observed earlier, in the absence of advance booking information, the lower acceptance rate comes with the flexibility to better satisfy the desired pickup times when we allow ridesharing. Our ability to test such a wide variety of service designs on different demand patterns is the key to observing these important relationships that may often appear unintuitive at first sight.

8.3 The outlook

The mobility service design has various stakeholders. The goal of the service providers is to maximize profit or minimize cost. The passengers, on the other hand, shop around for the best offer to satisfy their travel demand and service level preferences. The public transit authorities are direct stakeholders as well. They may commission new mobility services as part of the general public transportation service portfolio and in order to improve it. Service design helps them estimate what is possible and what the consequences would be, even if these services are later subcontracted to private providers. Last but not least, there are important externalities of the service design that make the general public a stakeholder as well.

By reducing the number of vehicles on the road and kilometers driven, smart micro-transit solutions can be instrumental in reducing green house gas emissions in densely populated urban areas (Leach, 2016). The taxi demand in Chicago on March 1st, 2017, was satisfied by 2711 vehicles that covered a reported 127,341 kilometers in transportation distance. With ridesharing, our results suggest that this value can be cut by almost half, while the number of vehicles needed by a factor of 10. The census tract spatial granularity in the original taxi data precludes the computation of deadheading, even if the trips of the same taxi were to be optimally chained. Still, we believe this value to be much higher compared to the one we obtain with the ridesharing service. Although each of the six-seater vehicles we consider may have a higher carbon footprint compared to a regular taxi, the difference factor is lower than 1.5 (Martinez and Viegas, 2017). Therefore, even with the current technologies and combustion-engine vehicles, we can still expect a substantial reduction of CO\textsubscript{2} and other green house gas emissions.

Fewer vehicles on the road also means faster and smoother travel, less time spent in congestion, and, as a result, additional savings in CO\textsubscript{2} emissions. Areas now used for parking can be freed up and converted to other more productive uses, for example towards solving the rising housing demand and property prices that many of our cities are currently facing. Indeed, fewer
vehicles will need fewer parking spots, also because they will be on the road transporting passengers most of the time. Embracing ridesharing services in favor of the perceived convenience and security of owning your personal vehicle can only happen when the convenience, security and reliability of these services begin to approach those of the private car. For this to happen, these services need to be designed towards the preferences of the local populations and in line with the expectations of the service providers and transit authorities. Simulation-based design and analysis is the first and necessary step in this direction.
Bibliography


© Bestmile 2018
Simulation-based design and analysis of on-demand mobility services


Author biographies

Samira Ehsani  received her M.Sc. in statistics (2011) from University Putra Malaysia. In her thesis, she focused on modeling repairable systems with covariates and interval failure data. She was a research assistant at the University of Malaya, Malaysia from 2014 to 2015 and focused on time series of over-dispersed and zero-inflated count data. She worked as a data analyst and quality control expert in the industry for four years. Samira joined Bestmile in November 2016 as a statistician and focuses on data analysis, modeling and identifying opportunities in large, rich data sets in transportation, traffic flow, demand and weather, and creating and implementing data-driven strategies.

Rafael Guglielmetti  received his M.Sc. in pure mathematics (2012) from the École Polytechnique Fédérale de Lausanne, and his Ph.D. from the University of Fribourg, Switzerland, in 2017. In his thesis, he worked on the computational aspects of geometry and algebraic number theory and published two open-source softwares and several articles. He joined Bestmile in June 2017 as an Algorithm Researcher / Software Engineer and works on an online solver for the dial-a-ride problem (DARP) that construct vehicle plans and schedules in real time. His research interests include exact approaches and heuristics for the DARP as well as algorithmic and implementation-related questions.

Marco Laumanns  received his Diploma in computer science (1999) from the University of Dortmund, and his Ph.D. in information technology and electrical engineering (2003), as well as a Habilitation in operations research (2010), from the ETH Zurich. From 2010 to 2017, he was a research staff member and project leader for transportation and operations research at IBM Research, where he worked on various research, development, and consulting engagements with large public transportation operators and agencies. He is a renowned expert in optimization and transportation, has co-authored over 100 scientific articles with over 17,000 citations, and serves on the executive committee of the Swiss Operations Research Society and the editorial board of the EURO Journal on Transportation and Logistics. Marco is responsible for Bestmile’s scientific and technological direction and for leading a research agenda covering data science, machine learning, simulation, optimization and control techniques for transportation system design and management.

Iliya Markov  received his M.Sc. in operations research from the University of Edinburgh (2010), and his Ph.D. in the field of stochastic vehicle and inventory routing problems from the École Polytechnique Fédérale de Lausanne (2017). During his Ph.D., he also spent time as a visiting scholar at the CIRRELT center for research in transportation in Montréal, Canada. He has published in Transportation Research Part B and C, and has presented at numerous conferences in Europe and worldwide. He is a member of EURO and its VeRoLog working group, as well as the Swiss Operations Research Society. Iliya joined Bestmile in November 2017 as an Algorithm Researcher / Software Engineer focusing on exact and heuristic approaches for various classes or routing and scheduling problems. His research interests include multi-period stochastic routing problems and the development of meta- and matheuristics.

Ravin de Souza  received his M.Sc. in computer science (1999) from Duke University in North Carolina and his Ph.D. from the École Polytechnique Fédérale de Lausanne, Switzerland and Instituto Superior Técnico, Lisbon, in 2016. In his thesis, he focused on transferring grasp
strategies from humans to robots for task-related operation in real-world environments. He has published articles in robotics journals and built sensing apparatus and software to study human grasp behavior. Prior to obtaining his Ph.D., Ravin has held positions in industry working on operating system design and parallel / distributed computing frameworks. Ravin joined Bestmile in November 2016 as a Machine Learning Engineer and focuses on the data analysis, modeling and implementation aspects of building and deploying predictive models for real-time decision making.