Anticipatory Assignment of Passengers to Meeting Points for Taxi-Ridesharing

Peter Dieter[a,*], Miriam Stumpe[a], Marlin Wolf Ulmer[b], Guido Schryen[a]

[a] *Universität Paderborn, Warburger Str. 100, 33098 Paderborn, Germany*
[b] *Otto-von-Guericke-Universität Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany*

Abstract

Taxi-ridesharing systems are considered an important means for sustainable urban transport. Previous literature shows that introducing meeting points in ridesharing, where customers are picked up and dropped off, increases its performance. We consider an on-demand taxi-ridesharing system, where the focus lies on the anticipatory assignment of customers to meeting points. We model the problem as a sequential decision process with the objective to maximize the distance saved through sharing. We suggest an anticipatory solution method for the planning of trips which assigns passengers to meeting points. We evaluate the suggested method on instances arising from real-world data and show that it leads to a significant increase in saved distance, and consequently CO_2 emissions when compared to a benchmark. We analyze the problem's and method's parameters and show that anticipatory methods further leverage the economical and ecological advantages of ridesharing.

Keywords: Ridesharing, Meeting points, Sequential decision process

*Corresponding author.
Email addresses: peter.dieter@upb.de (Peter Dieter), miriam.stumpe@upb.de (Miriam Stumpe), marlin.ulmer@ovgu.de (Marlin Wolf Ulmer), guido.schryen@upb.de (Guido Schryen)

1. Introduction

The share of the urban population is projected to grow from 50% in 2018 to 68% in 2050 which equals an estimated growth of 2.5 billion urban dwellers (UN, 2019). This development is coupled with a growing demand for urban mobility services (Miskolczi et al., 2021). Most traditional urban mobility solutions center around two extreme paradigms: mass public transport (e.g., buses, trams, metro) and private transport (e.g. privately owned cars). Mass public transport is considered to result in low costs for the customer and low carbon emissions. However, the flexibility and comfort of these systems are rather low. Private transport, on the other hand, is costly for customers but offers great flexibility and high comfort.

Taxi-ridesharing systems are designed to combine the advantages of both these paradigms, i.e., high flexibility at low costs by consolidating multiple requests into one shared taxi trip, for all passengers. Next to cost advantages for individual customers, taxi-ridesharing systems significantly improve the efficiency of transportation systems, reducing traffic congestion, fuel consumption, and air pollution (Hosni et al., 2014; Agatz et al., 2012; Santi et al., 2014). For instance, in a case study in Shanghai, Yan et al. (2020) show that fuel consumption can be reduced by 22.8% when taxi rides are consolidated into shared trips. Furthermore, previous research has shown that the use of predefined meeting points, where customers are picked up and dropped off, can further decrease the environmental impact of taxi-ridesharing systems (Stiglic et al., 2015), by enabling more direct vehicle routes. Also, predefined meeting points increase the sharing potential by leveraging customers' spatial flexibility.

One of the main challenges to fully leveraging the benefits of taxi-ridesharing systems is dealing with their inherent dynamic nature. Since crucial information, such as customer requests, is not predetermined but revealed over time, the system needs to be able to react in real-time to incorporate new requests. However, previously developed methods in literature that account for this dynamism often lack anticipatory methods. We contribute to the literature by introducing an on-demand taxi-ridesharing problem with meeting points and presenting an effective anticipatory policy that assigns passengers to meeting points using a heuristic approach, as classical exact optimization methods cannot effectively address the considered sequential decision processes for realistically sized instances. The aim of our work is to investigate to what extent the anticipation of future customer requests further increases the effectiveness of taxi-ridesharing systems. Furthermore, we quantify the impact of the proposed system in terms of kilometers and CO_2 saved, and derive suggestions that enable practitioners to design more sustainable on-demand ridesharing systems.

Motivated by the practical example of Uber Express Pool (Stock, 2018), in our study, we consider an on-demand taxi-ridesharing system where customer requests are assigned to a pair of meeting points (i.e., a pick-up and drop-off point) with the aim of grouping multiple requests to one single shared trip. Customers are required to walk to/from these meeting points to profit from higher sharing potential which is coupled with lower costs for both customers and taxi providers as well as a lower environmental footprint. The proposed system can be classified as on-demand as there are no fixed schedules involved but a trip is only performed upon customer request. Furthermore, intermediate stops during a taxi trip are not considered as they can lead to significant delays for individual customers (due to detours and time to park) and consequently might lead to inconveniences for the customer (Barann et al., 2017) as well as for the driver (Stiglic et al., 2015). A case study on Uber Express Pool has shown that the cancellation rate of customers increases significantly when the time between request and matching increases (Farronato et al., 2018). Therefore, we propose an event-driven system where decisions need to be made immediately when a customer request enters the system, and consequently, customers are responded to immediately.

Orchestrating the grouping of requests to trips between meeting points while considering the spatial and temporal restrictions of the individual customers is already a challenging task (Stiglic et al., 2015). The likelihood that two requests share the exact same origin and destination is very small. Thus, requesting customers have to be assigned to a convenient mutual pick-up point where the shared trip starts as well as to a drop-off point where the trip ends. Pick-up and drop-off points, both summarized under *meeting points*, connect multiple riders with one driver, as all customers of a shared trip are assigned to the same meeting point pair. Customers are required to walk to/from those meeting points, which should be in the close vicinity of the origin and destination for each customer. In practice, this task is further complicated by the fact that requests arrive dynamically over time and are unknown beforehand. Thus, when deciding about a customer's pair of meeting points, the other customers may not be known yet, but request a short time later. Thus, an *anticipatory* assignment of requesting customers to meeting points is required considering the stochastic arrivals of future requests. To the best of our knowledge and according to the review of Wang et al. (2022), the idea and evaluation of anticipatory meeting point assignments has not been studied in the literature yet.

When assigning requests to meeting points, two aspects play a significant role. Ideally, a customer should be sent to a "popular" meeting point where the likelihood of shared trips with future requests is high. At the same time, the system benefits from a flexible distribution of customers over meeting points to increase the probability that a future request can be matched. This can be achieved by a broad coverage of meeting points, i.e., by reducing the redundant overlap in the system. In this paper, we show how both aspects can be considered in a joint manner and how anticipatory deciding about meeting points can reduce taxi traffic significantly. To this end, we model the problem as a sequential decision process where over time requests occur and are assigned to pairs of pick-up and drop-off points. We present a policy that combines a lookahead on expected future demand to identify popular meeting points with a policy-function approximation (Powell, 2011) minimizing redundant overlap in the state to increase coverage. We test our policy on real-world taxi data from New York City and derive the following insights:

- Both parts, exploiting popularity and ensuring coverage perform very well by themselves, but work best when combined.

- In cases, where customers are rather flexible in their meeting points, ensuring coverage becomes significantly more valuable.

- Meeting points and trip sharing work particularly well in areas with high request density. In low-density areas, anticipatory meeting point design becomes crucial.

- Anticipatory assignments to meeting points and the resulting cost savings come at the expense of an increase in required customer walking, at least, in case a new trip is planned.

The paper is outlined as follows. In Section 2, the related literature is discussed. In Section 3, the problem is introduced and modeled. Our method is presented in Section 4. In Section 5, the design of experiments is defined followed by an evaluation of the policy in Section 6. The paper concludes with a summary and an outlook in Section 7.

2. Related work

Our work investigates the value of anticipatory policies for on-demand ridesharing with meeting points. In this paper, *ridesharing* means that multiple passengers share the same ride (in some literature also referred to as *ridepooling*). Work on this combination is rather limited, but there is work on the individual topics. In the following, we discuss the directly related work and give summaries about the individual components: meeting points, on-demand ridesharing and anticipatory methods in this context.

2.1. Meeting points in ridesharing

For a comprehensive review of literature on the optimization and categorization of ridesharing systems with walking paths and meeting points, we refer to Wang et al. (2022). Incorporating meeting points in ridesharing has its origins in carpooling, a form of ridesharing characterized by recurring trips of commuters who typically meet at park-and-ride locations and office buildings to form carpools (Kaan and Olinick, 2013). Stiglic et al. (2015) first apply the idea of meeting points to a system that provides automated matching between spontaneous drivers and passengers in an urban area. In their study, the authors consider a pre-defined set of meeting points and show that such alternative pick-up and drop-off points improve the performance of a ridesharing system in terms of matching rate and saved driving distance. Several similar studies further prove the benefits of meeting points for different types of ridesharing systems, such as *peer-to-peer ridesharing*, where private drivers with their own itineraries pick up and drop off passengers in between (e.g., Li et al., 2018; Zhao et al., 2018; Smet, 2021; Zuo et al., 2021), *semi-flexible transit services*, where schedules and stops of buses are flexibly adjusted to demand (e.g. Tong et al., 2017; Zheng et al., 2019) and *taxi-ridesharing*, where multiple customer requests are combined into one shared trip operated by drivers coordinated by a central organization (Ham, 2021; Aliari and Haghani, 2022). In the latter systems, the decision space consists of an assignment problem, in which groups of passengers are matched and assigned to a vehicle, and a balancing problem, which decides how to move idle vehicles (Tafreshian et al., 2020).

Similarly to our approach, locations for meeting points are determined a priori in the above mentioned papers (e.g. equidistantly distributed or at intersections of the underlying road network), and the presented algorithms focus on the assignment of requests to these meeting points. Other approaches use customer locations as meeting points and require all customers of a shared trip to have the same pick-up and drop-off point (Barann et al., 2017; Qian et al., 2017). The latter ensures a particularly high service level for customers, due to the avoidance of detours (Barann et al., 2017) and consequently, we adopt this feature to our approach and do not allow any intermediate stops and detours.

The aforementioned literature exclusively considers static settings, i.e., all requests are known in advance. Only a few approaches account for dynamically arriving requests in a ridesharing system with meeting points, either through agent-based simulation models (like Engelhardt and Bogenberger, 2021; Lotze et al., 2022) or centralized algorithms (like Lyu et al., 2019; Fielbaum et al., 2021; Tafreshian et al., 2020). The only one of these studies that anticipates future customer requests is Tafreshian et al. (2021).

2.2. Dynamism in ridesharing

Since, as shown in the previous section, only a few studies on meeting points pay attention to the dynamics typical for on-demand ridesharing, in this section, we explicitly address studies that focus on this key characteristic. The dynamic nature, i.e., continuously arriving transportation requests over time, necessitates algorithms capable of dealing with incomplete information, implying (i) short run times for real-size instances to enable regular re-optimization and/or (ii) anticipation of future requests that are not known beforehand but affect the overall solution quality. While the literature on the matching of customers and assignment to vehicles focuses on the scalability of algorithms, anticipatory methods are prominent in the literature on rebalancing of idle vehicles, which is reflected in the current and the following section. For a more extensive review of optimization for dynamic ridesharing and the need for agile algorithms, we refer to Agatz et al. (2012) and Martins et al. (2021).

Traditionally, researchers focus on a rolling horizon approach (also referred to as *batch mode*) when it comes to dynamic planning problems. Here, a static problem consisting of all known information within a planning horizon (batch) is resolved at certain time intervals. This is done to gather as much information as possible before a decision is made. As one of the first, Agatz et al. (2011) apply this approach to the bilateral matching of private drivers and passengers. Prominent examples of matching algorithms for ridesharing of multiple passengers are the approaches of Alonso-Mora et al. (2017), Wang et al. (2018) and Simonetto et al. (2019). Fielbaum et al. (2021) modify the method proposed in Alonso-Mora et al. (2017) by introducing meeting points and some heuristics to optimize the pick-up and drop-off points of each passenger.

Event-driven approaches are other means of dealing with constantly arriving information, for example, by performing a planning operation each time a new customer request is received. Such algorithms have the advantage of providing customers with an immediate response. For example, the assignment of passengers and vehicles is accelerated by algorithms based on a spatio-temporal index of taxis (Ma et al., 2013), a smart data structure (Schreieck et al., 2016), or a customer's time-expanded feasible network (Masoud and Jayakrishnan, 2017). Moreover, agent-based simulations are a popular means to investigate the performance of ridesharing systems Fagnant and Kockelman (2018); Lokhandwala and Cai (2018); Vosooghi et al. (2019). Engelhardt and Bogenberger (2021) and Lotze et al. (2022) use agent-based simulation models to study the merits of meeting points in particular.

2.3. Anticipatory methods in ridesharing

The anticipation of future events is considered in only a few studies on ridesharing systems, where information on future demand is incorporated into rolling-horizon approaches (e.g. Riley et al., 2020; Fielbaum et al., 2022) or event-driven approaches (e.g. Van Engelen et al., 2018; Wang et al., 2020). Some studies exclusively address either the assignment problem or the rebalancing problem. On the one hand, Qin et al. (2020) use reinforcement learning to adaptively adjust the time interval of a rolling horizon approach for matching participants in peer-to-peer ridesharing, which can be extended to a setting with a ride-pooling component. Matching of multiple passenger requests and vehicles through an integrated decomposition and approximate dynamic programming approach is presented by Yu and Shen (2019). On the other hand, Huang and Peng (2018) and Lin et al. (2018) propose methods for the rebalancing of idle vehicles based on demand predictions or reinforcement learning, respectively. Vehicle dispatching is another popular application of anticipatory methods in ridesharing, as shown by insertion algorithms based on lookaheads (Wei et al., 2017) or prediction of future demands Van Engelen et al. (2018); Wang et al. (2020). Another example is the approach of Riley et al. (2020) that integrates a machine-learning model to predict zone-to-zone demand over time, and a model predictive control optimization to relocate idle vehicles. Moreover, Haliem et al. (2021) use a deep reinforcement learning approach for joint matching, pricing, and dispatching. Fielbaum et al. (2022) build on the approach of Alonso-Mora et al. (2017) and develop anticipatory routing methods, which are characterized by the fact that no historical data is required. These works show that the incorporation of mechanisms for anticipatory decision-making is necessary to realize the greater potential of ridesharing approaches.

Our approach shares most features with the one by Tafreshian et al. (2021) as, to the best of our knowledge, their study is the only one dealing with meeting points in a dynamic ridesharing system anticipating future developments. Using forecasts of future trips based on historical data the authors construct multiple routes in an offline phase and choose among these routes in an online phase. Thus, their focus lies on shuttle dispatching, i.e., determining the order of served meeting points. The assignment of requests to meeting points is assumed to be given and is therefore not part of their decision. In contrast, the problem we consider centers around the matching of multiple requests to shared trips as well as their assignment to meeting points. In a sequential setting, our approach precedes that of Tafreshian et al. (2021) and complements it by an anticipatory assignment of requests to meeting points.

2.4. Research gap

We put an emphasis on two specific characteristics of the body of literature: Firstly, only a few studies have considered the incorporation of meeting points in stochastic dynamic ridesharing problems. Secondly, from a methodological perspective, there is still a lack of anticipatory policies for request grouping in on-demand ridesharing systems. Drawing on both observations, we contribute to the literature by introducing an on-demand taxi-ridesharing problem with meeting points which we formalize as a sequential decision process. Furthermore, we present an effective anticipatory policy and analyze its behavior and performance on scenarios deriving from real-world data.

3. Model

In this section, we first introduce the reader to the problem. We then formalize the problem as a sequential decision process and present an illustrative example.

3.1. Problem description

We deal with a system where customer requests for a taxi are entering the system dynamically. A request is characterized by the time it enters the system, the desired starting time of the trip, origin/destination of the customer, and the number of passengers. Decisions are made about assigning a new request to an existing trip (if possible) or planning a new trip. Furthermore, decisions are made online, i.e., all incoming requests need to be responded to immediately, and once a decision is implemented, it cannot be reversed. As an incoming request marks a new decision epoch, we deal with an event-driven system. We consider a static set of meeting points that serve as pick-up and drop-off points to/from which customers walk to their origins/destinations. We denote the combination of pick-up and drop-off points as a meeting point pair, representing a complete trip, since no intermediate stops are allowed. The objective is to maximize the distance saved by sharing requests to previously planned trips, i.e., saving the distance which would have been traveled by vehicles in a traditional taxi system where no sharing occurs, which is a common objective in literature (Barann et al., 2017; Stiglic et al., 2015). Therefore, the objective represents a system provider view, and customer-related factors, such as the distance customers walk to the meeting points, are accounted for in hard constraints. Requests are only suitable to be shared with an existing trip if the following matching conditions are fulfilled: Firstly, a certain level of temporal compatibility must be assured, i.e., the departure of the existing trip may only differ up to δ^t minutes from the planned departure time of the request. Secondly, spatial compatibility at the origin and destination must be assured, i.e., there is a maximum accepted walking distance (further referred to as δ^d) from the origin to the pick-up point and from the drop-off point to the destination. Thirdly, the ratio of walking time to/from the meeting points and the time it would take to walk directly from origin to destination must be below limit ϕ, as otherwise, the customer will likely walk directly. Lastly, enough vehicle capacity must be available. Assuming that all customers know that the taxi-ridesharing provider offers shared trips at a lower cost than traditional taxi trips, which may involve walking to and from the pick-up and drop-off points, we presume a sharing-willingness of 100% of all customers. No special preferences of customers are taken into account (e.g., regarding the preferred gender of fellow passengers, smokers, etc.) and a uniform average walking speed of all customers is assumed to calculate the duration of walking distances from and to meeting points. Requests can be made at any time of day via an online platform and the desired departure time is not restricted. We assume an unlimited, homogeneous taxi fleet (with uniform seat capacity). Taxi locations are neglected, and it is assumed that a taxi is always on time at the pick-up point. After all, the focus of our work lies not on routing, but on matching customer requests with similar spatiotemporal characteristics and their anticipatory assignment to meeting points through a central system.

3.2. Sequential decision process

In this subsection, we model the problem as a sequential decision process. Table 1 summarizes introduced tuples, parameters, and decision variables.

Decision Epochs. A decision epoch (also referred to as *epoch*), denoted by $k = 1, \ldots, K$, arises when the k^{th} request enters the system. Thus, the number of epochs K is a random variable.

States. The state of the system at decision epoch k is defined by the tuple $S_k = (C_k, J_k)$ where C_k contains information about customer request k and J_k is a set of planned trips which are not yet departed at epoch k:

- Variable C_k is a tuple containing $(t_k, s_k, a_k, b_k, n_k)$, where t_k is the time the request enters the system, s_k is the time the customer wants to start

Table 1: Sets, Parameters and Decision Variables

Notation	Definition
	State
S_k	state at decision epoch k
C_k	tuple containing information about the k^{th} customer request
t_k	request time of k^{th} request
s_k	time at which the k^{th} request wants to start his/her trip
a_k	origin of the k^{th} request
b_k	destination of the k^{th} request
n_k	number of passengers of the k^{th} request
J_k	tuple of planned trips at decision epoch k
m_j^k	number of available seats of trip j at decision epoch k
	Decisions
t'_j	time trip j departs
p_j	meeting point where trip j starts
d_j	meeting point where trip j ends
y_{kj}	binary decision variable indicating if the k^{th} request is shared to trip j
x_k	decision tuple defined as (y_{kj}, J_k^x)
	Post-decision state
S_k^x	post-decision State k
J_k^x	updated tuple of not yet started planned trips after decision k
	Reward
$R(S_k, y_{kj})$	reward arising from implementing decision y_{kj} in state S_k
	Transition
ω_{k+1}	stochastic exogenous information at decision epoch $k+1$
	Parameters
δ^d	maximal walking distance to/from a meeting point
δ^t	maximal time difference between desired departure time and planned departure time
ϕ	maximal allowed ratio between walking time to/from meeting points and the time walking directly from origin to destination
κ	initial capacity of a taxi

the (in-vehicle) trip, a_k is the origin of the travel, b_k is the destination of the travel, and n_k is the number of passengers of the k^{th} customer request. A travel of a customer is the process of walking to a pick-up point, the taxi trip, and walking from a drop-off point to the final destination.

- The planned trips at epoch k that have not yet started are denoted by set J_k. A trip $j \in J_k$ is the taxi trip from a pick-up point to a drop-off point. A trip $j \in J_k$ is defined by the tuple (t'_j, p_j, d_j, m_j^k), where t'_j is the time the trip departs, p_j is the pick-up point, d_j is the drop-off point, and m_j^k is the number of available seats at trip j at decision epoch k. When the first request C_1 arrives in S_1, J_1 is an empty set.

Actions. The action in decision epoch k is denoted by x_k and can take two forms: The planning of a new trip or sharing a request to a previously planned trip. Therefore, the decision is described by the tuple $x_k = (y_{kj}, J_k^x)$, where y_{kj} indicates potential sharing and J_k^x is the updated set of planned trips. More precisely, the decision variable y_{kj} is described as follows:

$$y_{kj} = \begin{cases} 1, & \text{if } k^{\text{th}} \text{ request is shared with trip } j \\ 0, & \text{otherwise} \end{cases}$$

Constraint (1) assures that a request is grouped to at most one trip.

$$\sum_{j \in J_k} y_{kj} \leq 1 \tag{1}$$

$$y_{kj} \in \{0, 1\} \quad \forall j \in J_k \tag{2}$$

Furthermore, if $y_{kj} = 1$ the following constraints need to be fulfilled:

$$n_k \leq m_j^k \tag{3}$$

$$|s_k - t'_j| \leq \delta^t \tag{4}$$

$$\theta(a_k, p_j) \leq \delta^d \tag{5}$$

$$\theta(d_j, b_k) \leq \delta^d \tag{6}$$

$$t_k + \theta(a_k, p_j) \leq t'_j \tag{7}$$

$$\frac{\theta(a_k, p_j) + \theta(d_j, b_k)}{\theta(a_k, b_k)} \leq \phi \tag{8}$$

Constraint (3) assures enough car capacity. Constraint (4) assures that the time the travel of request in decision epoch k starts corresponds with the time the trip j departs, where δ^t is the maximal allowed time difference between desired departure time and planned departure time. Constraints (5) and (6) limit the walking time from the travel origin to a pick-up point and from a drop-off point to the travel destination to be at most δ^d respectively, where θ is a function which measures the walking time between two locations. Constraint (7) assures that the customer has enough time to walk to the pick-up point (with the request time t_k as the earliest possible start of walking). Constraint (8) limits the ratio between walking time to/from the meeting points and the time it would take to walk from the origin to the destination directly, where ϕ is the maximal allowed ratio. This ratio should be in the range between 0 and 1 as a ratio above 1 would mean that walking directly from the origin to the destination would be shorter than walking required by using the taxi service. We note that the ratio would remain unchanged if we would apply walking distances rather than walking time.

If the request in decision epoch k is not shared with any trip, i.e., $\sum_{j \in J_k} y_{kj} = 0$, a new trip $|J_k|+1$ is planned and a corresponding new trip needs to be added to J_k, leading to the post-decision tuple of planned trips J_k^x. Therefore, we need to decide upon a pick-up point, $p_{|J_k|+1} \in M$ and drop-off point, $d_{|J_k|+1} \in M$, where M is a set of predefined meeting points. Additionally, we need to decide upon a time the trip departs $t'_{|J_k|+1}$. The decisions need to fulfill the following constraints:

$$|s_k - t'_{|J_k|+1}| \leq \delta^t \tag{9}$$

$$\theta(a_k, p_{|J_k|+1}) \leq \delta^d \tag{10}$$

$$\theta(d_{|J_k|+1}, b_k) \leq \delta^d \tag{11}$$

$$t_k + \theta(a_k, p_{|J_k|+1}) \leq t'_{|J_k|+1} \tag{12}$$

$$\frac{\theta(a_k, p_{|J_k|+1}) + \theta(d_{|J_k|+1}, b_k)}{\theta(a_k, b_k)} \leq \phi \tag{13}$$

$$t'_{|J_k|+1} \in \mathbb{N} \tag{14}$$

$$p_{|J_k|+1}, d_{|J_k|+1} \in M \tag{15}$$

Corresponding to Constraints (4)-(8) in the sharing decision, in case a new trip is planned, Constraints (9)-(13) limit the time differences, maximal walking time and ratio between walking to meeting points and walking directly from the origin to the destination. Constraint 14 states that the departure time of the trip is a natural number (e.g., measured in seconds or minutes), and constraint (15) states that our chosen pick-up and drop-off points must be in the set of pre-defined meeting points M.

Rewards. The reward that arises from taking action x_k in state S_k consists of the distance between the origin and destination of the request arising in decision epoch k which is saved by sharing. It is the distance we save compared to a traditional taxi system where no sharing is performed. Therefore, the following reward function is defined:

$$R(S_k, y_{kj}) = \sum_{j \in J_k} \theta^d(a_k, b_k) \cdot y_{kj}, \tag{16}$$

where θ^d is the driving distance between two locations. The walking distance of customers is not included in the reward function but is only included as a hard constraint i.e., we focus on the system from an operator's perspective.

This is done as our study has the intent to inform operators on the choice of the system's design and present solution methods for this purpose. We note that other objectives (including the customer's perspective) could be investigated in the future.

Transition. We deal with two types of transitions. Firstly, a deterministic endogenous transition occurs when an action has been taken, leading to post-decision state $S_k^x = (C_k, J_k^x)$. In case a new trip is planned, we extend the tuple of planned trips at epoch k that have not yet started, i.e., J^k, by the tuple $(t'_{|J_k|+1}, p_{|J_k|+1}, d_{|J_k|+1}, \kappa - n_k)$ where κ is the vehicle capacity (we assume a homogeneous fleet of vehicles). Therefore, J_k^x is defined as the following: $J_k \cup (t'_{|J_k|+1}, p_{|J_k|+1}, d_{|J_k|+1}, \kappa - n_k)$. In case request k is shared with another trip j, the remaining capacity of this trip is updated accordingly, i.e., $m_j^{k,x} = m_j^k - n_k$. Secondly, a stochastic exogenous transition occurs when a new customer request $k+1$ emerges. This transition is denoted by ω_{k+1} and leads to a new state $S_{k+1} = (C_{k+1}, J_{k+1})$ at time t_{k+1}. Variable C_{k+1} contains the information of the new request. Variable J_{k+1} is the (undeparted) subset of J_k^x with $t'_j \geq t_{k+1}$.

Objective. A solution for our problem is a decision policy π from the overall set of policies Π. A policy π assigns an action $X_k^\pi(S_k)$ to each state S_k. The optimal solution is a policy π^* that maximizes the total expected reward over all epochs:

$$\pi^* = \underset{\pi \in \Pi}{\operatorname{argmax}}\ \mathbb{E}\left[\sum_{k=0}^{K} R(S_k, X_k^\pi(S_k))|S_0\right], \tag{17}$$

when starting in initial state S_0.

3.3. Illustrative example

The problem at hand is illustrated in Figure 1. It shows the system in a state S_k. Assuming a grid where the walking time per segment is 5 minutes, which is equal to the maximal allowed walking time from the origin to the pick-up point as well as from the drop-off point to the destination. Intersections represent

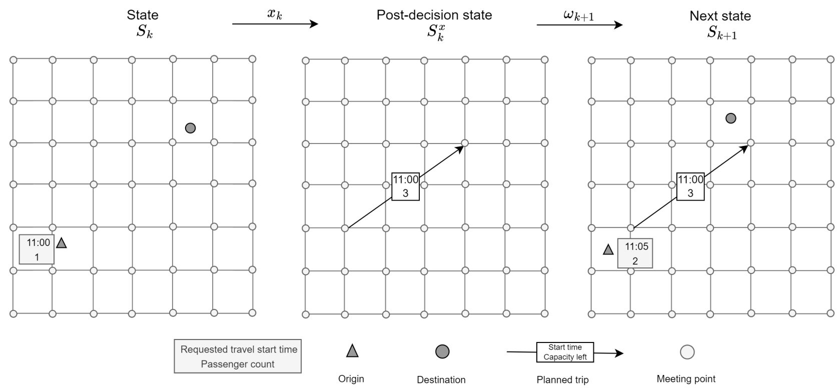


Figure 1: Illustrative example of the problem at hand

meeting points. At decision epoch k, a request for one passenger at 11:00 enters the system and there are no planned trips in the system. We, therefore, need to plan a new trip j for this request. There are multiple possibilities to assign a pick-up and drop-off point to trip j. A simple policy would be to choose the pick-up and drop-off point closest to the origin and destination of the request, respectively. Following this policy leads us to post-decision state S_k^x in which the system solely contains the previously planned trip. In decision epoch $k+1$, a request for two passengers who want to start the trip at 11:05 enters the system. As the origin/destination are within walking distance to pick-up/drop-off point of the trip planned in epoch k and the temporal difference is only 5 minutes, we have the possibility to either share or plan a new trip for the request. In case we share the request emerging in decision epoch $k+1$ to trip j, we save the distance from the origin (triangle) to destination (circle) of the request emerging in $k+1$, i.e., the distance which would have been traveled by a vehicle in a traditional taxi system where no sharing occurs. Therefore, our objective value would increase by the respective saved distance.

4. Solution method

In this section, we present the developed solution method for the problem at hand. We first give a general motivation and overview and then describe the components of our policy in detail.

4.1. Overview

The problem has two different types of decisions. First, if sharing is possible with more than one trip, which trip should the new request be shared with, recalling that the reward of the decision is the same for all trips? Second, if the request cannot be shared and a new trip needs to be generated, how to determine the meeting points the trip starts and ends? Since this second question is significantly more complex, it is the focus of this work and we will give an overview of our approach in the next paragraph. Furthermore, pretests have shown that the sharing decision has a marginal impact on the objective value compared to the new trip decision. Therefore, for the first question, we rely on a straightforward rule. In case sharing is possible with more than one trip, the trip with the earliest departure time is selected. The idea behind this rule is that the sharing potential of a trip decreases the closer the departure time gets.

When determining meeting points for new trips, two aspects should be considered, popularity and coverage. First, decisions should anticipate the expected future demand, thus, the sharing potential for a meeting point pair. Popular pairs that have a high number of expected "shareable" future requests, should be preferred. For a request to be shareable, both meeting points need to be eligible for this request. Even if the pick-up point might fit, a request cannot be shared if the drop-off point is too far away. This leads to another aspect to consider, flexibility via a broader coverage of the service area. Assigning meeting points whose walking reachability partially overlaps with the meeting points of previously planned trips might reduce coverage and sharing potential. Avoiding overlap may lead to a more flexible setup to accommodate future requests.

Both aspects play an important role when determining meeting points. Thus, we propose a combined approach considering both popularity of meeting point pairs and the avoidance of overlap. Furthermore, the importance of an individual aspect might depend on the demand distribution. A focus on popularity might be valuable if overall sharing opportunities are limited, e.g., if the demand is low or customers do not want to walk that far. In such cases, assigning the most-popular pairs might increase the small probability of sharing. In other cases, maintaining coverage by avoiding overlap may play a more important role, e.g., if customers are more flexible or if the demand is high. To account for the different importance of the two aspects, we present a weighted combination of both as our policy and the weight is shifting the focus between them. In the following, we present the specific details of our method with examples. We start with an overview of the general decision-making process in a decision state.

4.2. General decision making process

The general decision-making process is similar in all solution methods developed and is described in Algorithm 1. In epoch $k = 1$, we start with an empty set of planned trips J_1 and initialize the reward to 0. In each epoch, we first check if sharing a request is possible (*GetSharingPartners*). We denote the set of possible sharing partners by J_k^s, which is a subset of all trips planned before epoch k, i.e., J_k. If J_k^s is an empty set, i.e., no sharing is possible, a new trip needs to be planned with the *NewTripPolicy* which we further describe in the next section. As our work focuses on the assignment of requests to meeting points, in all policies, we set the time the new trip starts ($t'_{|J_k|+1}$) to either the desired departure time or the time needed to reach the chosen pick-up point, i.e., $max(s_k, t_k + \theta(a_k, p_{|J_k|+1}))$. We note that further research could investigate policies that optimize the departure time. We then add the new trip to the set of planned trips J_k. If at least one sharing partner is found, we apply the *SharingPolicy* which assigns the k^{th} request to a trip $j \in J_k^s$. Consequently, we update the reward by adding the distance which the k^{th} customer requested to travel to the previous reward. Furthermore, we update the capacity of the chosen trip by subtracting the requested passenger counter of the k^{th} request, n_k, from the previous capacity. We continue with suggesting the *NewTripPolicy*.

Algorithm 1 General Procedure

Input: Empty set of $PlannedTrips = J_1$, Initial decision epoch $k = 1$, Number of decision epochs K, $Reward = 0$

```
while k ≤ K do
    J_k^s ← GetSharingPartners(S_k)
    if J_k^s is empty then
        NewTrip ← NewTripPolicy(S_k)
        J_{k+1} ← J_k ∪ NewTrip
    else
        ChosenSharingTrip ← SharingPolicy(S_k)
        Reward ← Reward + GetDistance(a_k, b_k)
        UpdateCapacity(ChosenSharingTrip, n_k)
    end if
    k ← k + 1
end while
```

4.3. Proposed policy

In this subsection, we will present the proposed new trip policy. The policy combines a lookahead on expected future demand to identify popular meeting points with a policy-function approximation minimizing redundant overlap in the state to increase coverage. We will first elaborate on each individual component of the policy, which we denote as *popularity* and *overlap* components. We then propose a way to combine these two individual components.

Popularity component. In this component of the policy, we first determine the time-dependent popularity of each meeting point pair $M \times M$, by looking at historical data. For each hour of a day, all requests demanding a trip in that hour, which would have been assignable (in terms of walking distance) to a meeting point pair, contribute to the popularity of the pair. Consequently, a request can contribute to the popularity of multiple meeting point pairs. The meeting point pair with the highest popularity is chosen. The component is visualized in Figure 2, where the system is in state S_k. There exists one previously planned trip but assuming that the maximal allowed walking distance δ^d is equal to a segment length, we cannot assign the request to the trip due to spatial incompatibility at the drop-off point. Thus, a new trip needs to be planned. Moreover, suppose that the thickness of the arrows is proportional to the popularity of the meeting point pair. Therefore, we would choose the leftmost meeting point pair (x_1), as pick-up and drop-off point for the new trip. Note that for reasons of clarity, not all possible new trips are visualized.

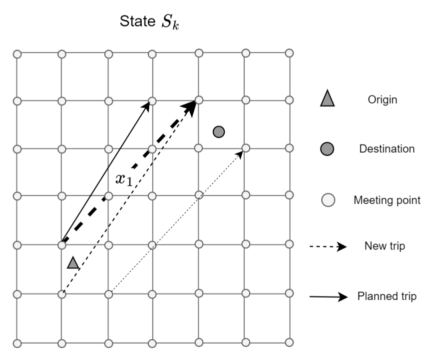


Figure 2: Popularity component. The dashed arrows represent feasible meeting point pairs, with their thickness being proportional to the popularity of the pair.

Overlap component. The previously described component is not considering already planned trips (J_k), when a new trip is being planned. This can lead to inefficient plannings due to redundancies. Let us extend the example of Figure 2 with Figure 3. Taking action x_1 would follow the popularity strategy but leads to a high redundancy at the drop-off area because the walking reachability overlaps, i.e., there is an area which is served by both points. Taking action x_2 would increase the coverage and potentially lead to a higher future sharing potential.

To account for this redundancy, we introduce the *minimal overlap component*

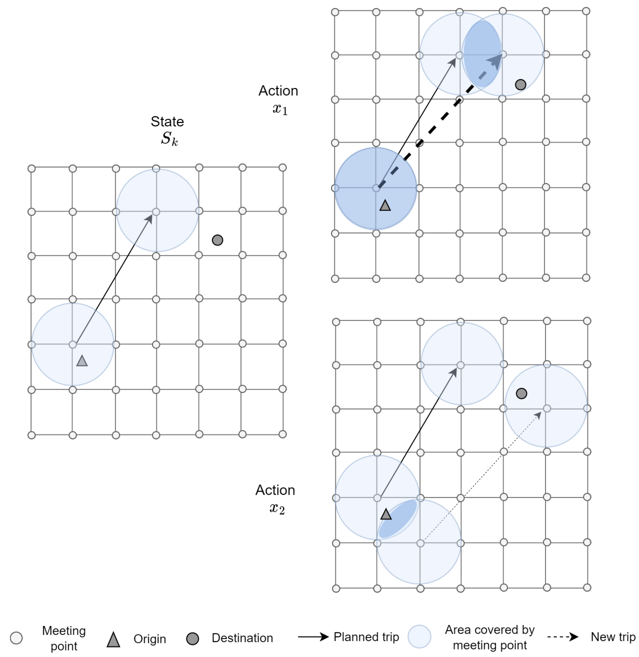


Figure 3: Overlap component. The light blue circles represent the area from where the meeting points of a trip are within walking distance. The dark blue area marks the overlap of covered areas by different trips. The dashed arrows represent feasible meeting point pairs, with their thickness being proportional to the popularity of the pair.

in which we do the following: In the first step, the component identifies all meeting points within walking distance of both the desired departure and arrival locations of a new request to determine all feasible meeting point pairs. For each of these pairs, it then checks if similar trips have already been planned. A planned trip is considered similar if it satisfies all of the following conditions:

a) an overlap exists at the pick-up points,

b) an overlap exists at the drop-off points, and

c) the trip starts in the time window $[s_k - \delta^t, s_k + \delta^t]$.

In the second step, the overlap area between the meeting point pair and the trips already planned is calculated. The meeting point pair leading to the lowest overlap is chosen. Looking at Figure 3, the planned trip in action x_1 leads to high redundancy, especially at the pick-up point. This redundancy is lower for the trip we plan in action x_2. We would therefore prefer x_2 over x_1.

Weighted popularity & minimal overlap policy. As both previously described components should be considered, we combine the popularity component with the minimal overlap component. Let Ω be a vector that contains the overlap area, and P a vector which contains the popularity score of each meeting point pair. We then first normalize both vectors. Max-min normalization is applied to normalize P, which scales every entry in the range of $[0, 1]$:

$$P_i^{norm} = \frac{P_i - \min(P)}{\max(P) - \min(P)} \quad \forall i \in \{1, \ldots, |P|\}. \tag{18}$$

We adjust the max-min normalization for Ω to account for the fact that a high overlap is undesirable. The following adjustment transforms high values to low values and vice versa:

$$\Omega_i^{norm} = \frac{\max(\Omega) - \Omega_i}{\max(\Omega) - \min(\Omega)} \quad \forall i \in \{1, \ldots, |\Omega|\} \tag{19}$$

The property of normalized values being in the range of $[0, 1]$ is maintained. We then calculate the fitness F_i for each meeting point pair i where $\alpha \in [0, 1]$ is a weighting parameter between the two components:

$$F_i = \alpha \cdot \Omega_i^{norm} + (1 - \alpha) \cdot P_i^{norm} \quad \forall i \in \{1, \ldots, |\Omega| = |P|\}. \tag{20}$$

When $\alpha = 0$, only the popularity component is applied. When $\alpha = 1$, only the minimal overlap components is applied. The meeting point pair with the highest fitness is chosen. In case of a tie, the meeting point pair with higher popularity is chosen. In the rare case that this comparison also leads to a tie, the meeting point pair which leads to the lowest walking distance for the customer is chosen. A visual summary of the two components and the weighted policy is given in Figure 4. As previously, the thickness of the arrows is proportional to the popularity of their meeting point pair. The proposed policy therefore chooses the meeting point pair which leads to an intermediate level of overlap and has an intermediate level of popularity (x_3). This keeps the high likelihood of a popular matching, but leads to a broader coverage of the service area.

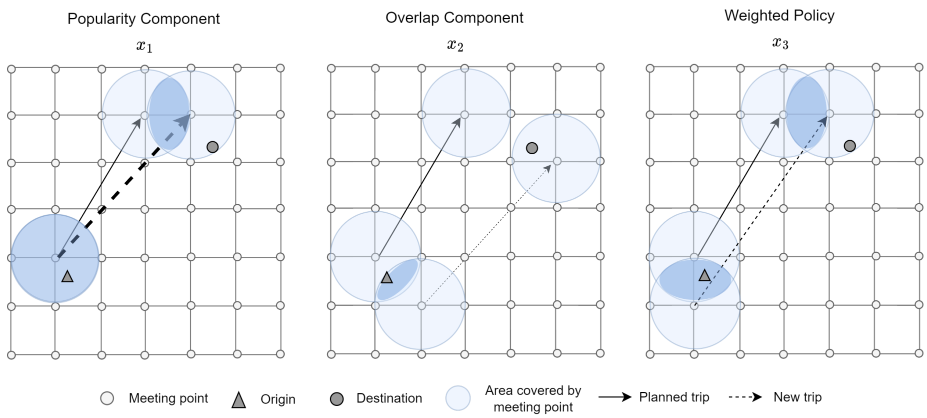


Figure 4: Policy combining popularity and overlap components. The light blue circles represent the area from where the meeting points of a trip are within walking distance. The dark blue area marks the overlap of covered areas by different trips. The dashed arrows represent feasible meeting point pairs, with their thickness being proportional to the popularity of the pair.

5. Experimental design

In this section, we present the data and system settings which are used to create several scenarios to analyze the performance of our policies presented in Section 4. To do so, we apply the nearest-meeting point policy, i.e., a policy without anticipation, which serves as a benchmark policy (see Engelhardt and Bogenberger, 2021).

5.1. System settings & scenarios

We draw on taxi data collected in New-York city (NYC City Taxi & Limousine Commission: TLC, 2017). The dataset is widely used in studies on taxi systems (see e.g. Barann et al., 2017) and represents a traditional taxi system where no sharing occurs. The data includes, among other things, the origin/destination and the number of passengers of taxi trips as well as the time the trips started. We extend the data by adding a request time to each trip. The period between the request time and the time the customer wished to start his/her trip is called the lead time. To generate the lead time, we draw from a uniform distribution. We vary the upper limit of the lead time distribution. The chosen values are presented at the end of this subsection.

For the popularity policy, we determine the popularity of meeting point pairs with data from the first two weeks of September 2015 and test the policies on the last two weeks of September and the first week of October, resulting in 21 test instances. The walking speed is set to 5.1 km/h, which is (roughly) the average normal walking speed of people (Bohannon and Andrews, 2011). As a distance measure, we apply the Haversine distance. Furthermore, we consider a base scenario plus an analysis, where we have parameters that we vary to study the system's sensitivity: Firstly, the meeting point density, i.e., the distance between meeting points. We automatically create an equidistant grid of meeting points with an interval of 600 meters in longitudinal as well as latitudinal directions. We exclude meeting points that are located in major parks, industrial areas, or water areas. Even though this might lead to meeting points located in inaccessible areas, such as buildings, the automatic creation is done as the manual identification of thousands of meeting points is beyond the scope of this study (with an interval of 600 meters, 2306 meeting points have been created). Secondly, we vary the maximum walking time from the customer's origin to a pick-up point and from a drop-off point to the destination of the customer. Thirdly, the geographical area of customer requests is varied. Fourthly, we vary the volume of requests by filtering out customer requests at random. Further, we vary the vehicle capacity κ, the maximal allowed ratio between walking to meeting points and walking directly from origin to destination (ϕ), and the difference between requested trip starting times and the actual trip departure times, i.e., δ^t. In the last scenario, we slightly adjust the problem presented in Section 3, by combining the walking time from/to meeting points instead of regarding them separately. I.e., we replace Constraints (5) and (6) as well as (10) and (11), by single constraints that sum up the walking times to and from meeting points (with $2 \cdot \delta^d$ on the right-hand side instead of δ^d).

In the base scenario, the following values are applied: The meeting point density is set to 600 meters (as visualized in Figure 5), the maximum walking time is set to 450 seconds and we do not restrict the geographical area. Furthermore, we set the volume to 75% to simulate a lower demand volume for such a taxi-ridesharing system compared to traditional taxi-ridesharing systems where no sharing occurs and consider a homogeneous vehicle fleet with a capacity of 4 seats. Moreover, ϕ and δ^t are set to 0.25 and 300 seconds, respectively.

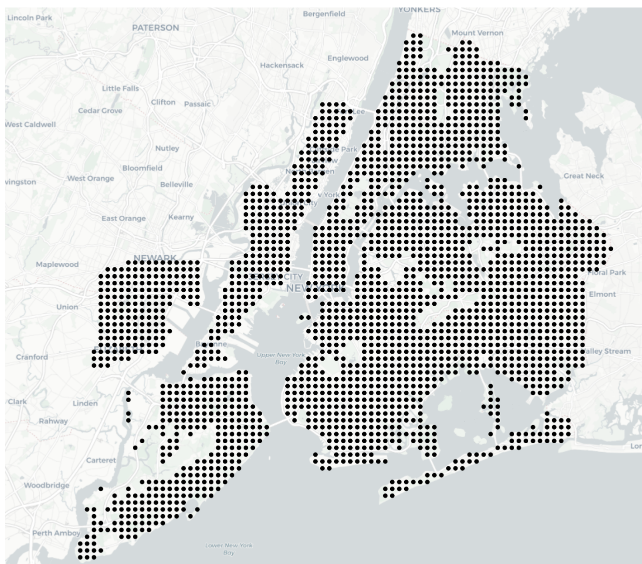


Figure 5: Created meeting points with a distance of 600 meters

For our analysis, we vary the parameters as follows. In the *high* and *low meeting point (MP) density* scenarios, a meeting point distance of 400 and 800 meters is applied, respectively. In the *low walking* and *high walking* scenarios, the maximal allowed walking time is set to 300 and 600 seconds. The *low volume* scenario considers a volume of 50% and the *high volume* scenario a volume of 100%. In the *low car capacity* and *high car capacity* scenario, a vehicle capacity of 2 and 8 is assumed. The lower lead time limit is set to 300 seconds, and we vary the upper lead time (recall that we draw from a uniform distribution).

In the *low lead time* scenario, the upper limit is set to 900 seconds and in the *high lead time* scenario the upper limit is set to 3600 seconds. Moreover, we consider a scenario solely with requests from and to Manhattan and one where only requests are considered which are neither starting nor ending in Manhattan. This is done to perform analyses on cases with different spatiotemporal densities of customer requests, as around 80% of requests do start or end in Manhattan, while Manhattan only makes up for around 7.5% of the total New-York city land area. Further, in the *low ratio* and *high ratio phi* is set to 0.125 and 0.5. In the *low time difference* and *high time difference* scenarios, δ^t is set to 150 and 450 seconds, respectively. Lastly, in the *walking combined* scenario, walking to/from meeting points is combined and not regarded individually, as explained above. A summary of the scenario settings is to be found in Table 2.

Table 2: Considered Scenarios

Scenario	**MP Density (m)**	$\delta^d(s)$	**Area**	**Vol.**	κ	**Lead time (s)**	ϕ	$\delta^t(s)$	Walking combined
Base scenario	600	450	all NYC	75%	4	[300, 1800]	0.25	300	×
High MP density	**400**	450	all NYC	75%	4	[300,1800]	0.25	300	×
Low MP density	**800**	450	all NYC	75%	4	[300,1800]	0.25	300	×
Low walking	600	**300**	all NYC	75%	4	[300,1800]	0.25	300	×
High walking	600	**600**	all NYC	75%	4	[300,1800]	0.25	300	×
Manhattan	600	450	**only Manh.**	75%	4	[300,1800]	0.25	300	×
Not Manhattan	600	450	**not Manh.**	75%	4	[300,1800]	0.25	300	×
Low volume	600	450	all NYC	**50%**	4	[300,1800]	0.25	300	×
High volume	600	450	all NYC	**100%**	4	[300,1800]	0.25	300	×
Low car capacity	600	450	all NYC	75%	**2**	[300,1800]	0.25	300	×
High car capacity	600	450	all NYC	75%	**8**	[300,1800]	0.25	300	×
Low lead time	600	450	all NYC	75%	4	**[300,900]**	0.25	300	×
High lead time	600	450	all NYC	75%	4	**[300,3600]**	0.25	300	×
Low ratio	600	450	all NYC	75%	4	[300,3600]	**0.125**	300	×
High ratio	600	450	all NYC	75%	4	[300,3600]	**0.5**	300	×
Low time difference	600	450	all NYC	75%	4	[300,3600]	0.25	**150**	×
High time difference	600	450	all NYC	75%	4	[300,3600]	0.25	**450**	×
Walking combined	600	450	all NYC	75%	4	[300,3600]	0.25	300	✔

5.2. Benchmark policies

The main focus of this paper is to describe how anticipatory meeting point assignments can be achieved and how it changes decision making and the performance of the system. To allow this analysis, we compare our method with a non-anticipatory, practically-inspired benchmark policy, the nearest-meeting point policy, which is also used as benchmark in other studies (Engelhardt and Bogenberger, 2021). We assume that the sharing part of the policy remains the same, i.e., we always share when possible and we choose the trip which possesses the earliest departure time. However, when a request enters a system that cannot be shared, the nearest meeting points from the origin and destination serve as pick-up and drop-off point, respectively. In contrast to our proposed method, the nearest-meeting point policy is non-anticipatory, as it neglects customer requests that might arrive in the future and further, it also does not make use of information on trips that have already been planned but have not yet started. Besides this policy, we also test the two components of the policy on their own by setting $\alpha = 0$ and $\alpha = 1$, respectively.

6. Policy evaluation and analysis

In this section, we evaluate and analyze the policy proposed in Section 4. Firstly, we compare the objective values and decisions made when we apply the popularity and overlap components individually, i.e., when $\alpha = 0$ and $\alpha = 1$. Secondly, we determine the optimal value of α of the policy and interpret this value for each scenario. Thirdly, with the optimal value of α in the previous step, we compare customers' walking distances for different scenarios.

6.1. Method analysis

In this section, we investigate the performance of our method and its individual components for the base case and the scenarios.

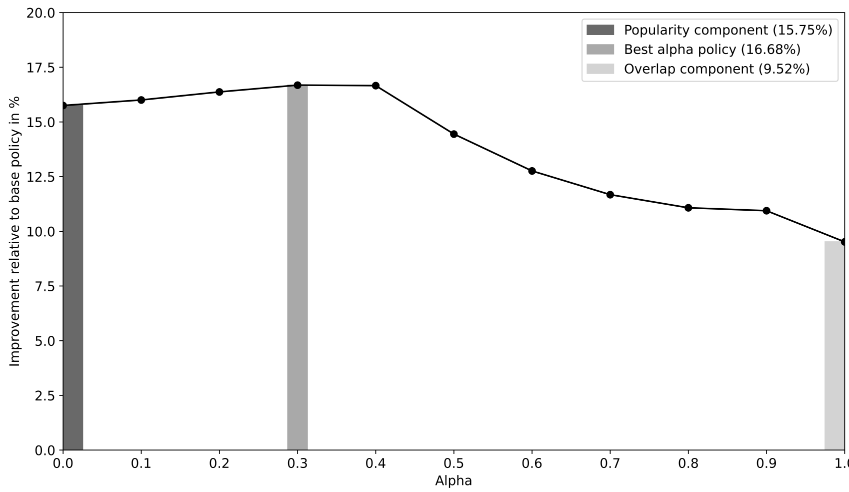


Figure 6: Role of α in base scenario

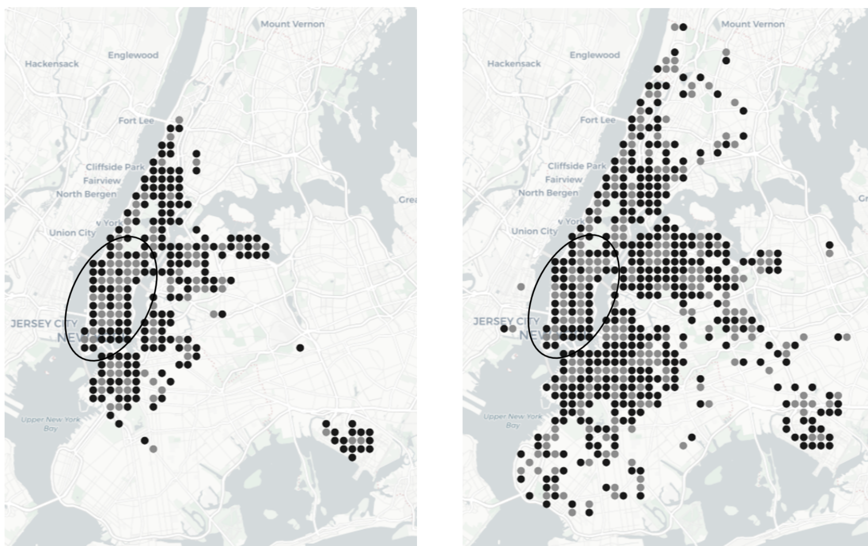
Objective value. Figure 6 shows the objective value for the base scenario, i.e., the relative saved distance compared to the benchmark policy (nearest meeting point policy) in percentage for all tested values of α.

We observe that our policy outperforms the non-anticipatory benchmark regardless of the value of α. Solely considering popularity leads to an improvement of 15.75% compared to the benchmark policy while solely focusing on overlap leads to an improvement of 9.52%. The best value of α lies at 0.3 and results in an improvement of 16.68%. The plots for other scenarios are to be found in Appendix A. These results let us infer the following: Firstly, anticipating future customer requests and ensuring sufficient coverage leads to significantly higher savings compared to the non-anticipatory benchmark policy. Secondly, at least for our setting, in combination, selecting popular meeting points should receive a higher weight than avoidance of overlap.

Decision making. Next, we compare the decisions made when we only apply the popularity and overlap components individually. This is visualized for chosen meeting points in Figure 7 where we consider the base scenario of one exemplary day. Each circle represents a meeting point where the meeting point is grey if it was chosen more often when solely considering popularity ($\alpha = 0$) and black when it was chosen more often when solely considering overlap ($\alpha = 1$). For visual reasons, if a meeting point is chosen equally often, it is not shown. We see that the number of selected customer origins is lower compared to the number of selected customer destinations. Moreover, we can see that when only considering popularity, more meeting points are chosen in busy areas while more meeting points on the border of busy areas are chosen when solely applying the overlap component. This can be observed in lower and middle Manhattan (encircled areas in Figure 7), whose areas at the riverbank are less busy than the areas further away from the riverbank. Further, the focus on the overlap component leads to more used meeting points, i.e., 657 and 1493 pick-up and drop-off points used compared to 648 and 1480 when only focusing on popularity. These findings show that the developed policies work as expected: A high focus on popularity leads to choosing meeting points in areas with high request volumes more often while a high importance on reducing overlap leads to a higher geographical coverage by a) more often selecting meeting points which cover the border of busy areas and b) selecting more meeting points overall. In terms of algorithm efficiency, for a single request, the proposed policy returns an action in milliseconds.

Scenario analysis. We now take a closer look at the individual scenarios. Table 3 shows the improvements of the extreme cases ($\alpha = 0$ and $\alpha = 1$) as well as the best α value for every scenario. A table with the improvements for all values of α is to be found in Appendix A (Table 6).

First of all, we observe that our policy outperforms the benchmark for all scenarios by a large margin. Thus, anticipatory meeting point assignments are superior to non-anticipatory methods regardless of the instances tested. When analyzing the weighting between the two components, we can see that the best value of α is consistently between 0.2 and 0.4. The overlap component plays a more important role ($\alpha = 0.4$) when the density of meeting points is high or



(a) Pick-up points

(b) Drop-off points

Figure 7: Selection frequency of meeting points depending on policy. Each meeting point's color indicates the policy under which it's more frequently chosen: black for overlap-based ($\alpha = 1$) and grey for popularity-based policy ($\alpha = 0$). The encircled area marks lower and middle Manhattan.

Table 3: Influence of α on improvement to benchmark policy

Scenario	$\alpha = 0$	$\alpha = 1$	Best value	Best α
Base Scenario	15.75	9.52	16.68	0.3
High MP density	18.16	10.54	19.45	0.4
Low MP density	7.10	1.89	7.18	0.2
High walking	12.80	8.98	14.01	0.4
Low walking	8.79	3.13	8.83	0.2
Manhattan	9.27	2.59	10.22	0.4
Not Manhattan	51.31	48.32	52.00	0.2
High volume	13.31	7.46	14.21	0.4
Low volume	19.41	12.64	20.45	0.3
High car capacity	14.95	8.81	15.89	0.3
Low car capacity	16.68	10.24	17.61	0.3
High lead time	15.77	9.79	16.69	0.3
Low lead time	16.03	5.93	16.84	0.3
Low ratio	19.54	12.85	20.27	0.3
High ratio	13.27	7.75	14.14	0.3
Low time difference	22.21	5.42	22.85	0.3
High time difference	13.03	8.71	13.94	0.4
Walking combined	8.99	7.73	10.87	0.4
Average	16.47	10.13	17.34	0.32

when customers are allowed to walk more compared to the scenarios in which the meeting point density is low or when the maximal walking distance is more restricted ($\alpha = 0.2$). A possible explanation might be the following: Choosing the meeting point pair with the highest popularity is the “safest” decision as we expect most requests for the chosen meeting point pair. On the contrary, choosing a meeting point pair with low overlap might increase coverage but also result in more “risky” decisions as we do not consider popularity directly. Increased spatial flexibility allows customers to reach a larger number of meeting points, and by doing so, reduces the risk of the decision made when focusing on overlap. Furthermore, we can see from Table 3 that the overall improvement is very high when we do not consider requests from or to highly-populated Manhattan (more than 48% for all values of α). It is therefore even more crucial to anticipate future customer requests when operating such a system in regions with a lower volume of customer requests, likely because a lot of sharing happens automatically in highly popular areas. This hypothesis is supported by the finding that the relative improvement of our policies is higher in scenarios where the volume is low. Combining the two insights of coverage and expected demand, it might be valuable for a provider to consider heterogeneous services for different areas of the city, e.g., by expecting higher walking flexibility in areas with less demand. Besides, the improvement compared to the benchmark policy is higher when the car capacity is low, likely, because, similar to low demand, the sharing potential is relatively small, and smart decision-making becomes more important. Furthermore, we can see that the best α is invariant throughout different upper lead time limits ($= 0.2$). However, solely applying the overlap component leads to a bigger drop in the low lead time scenario. This is likely due to more required customer walking which cannot always be satisfied when the lead time is too low, i.e., customers do not have enough time to reach a meeting point on time. Anticipatory meeting point assignment becomes increasingly important when the maximally allowed ratio between walking to/from meeting points and walking directly and the time difference is low. An explanation for this observation is that many requests cannot be satisfied which leads to a lower trip density and consequently, a smart allocation of requests to these trips becomes more important. Lastly, when walking to/from meeting points is combined, a balance between the two extreme policies outperforms the $\alpha = 0$ policy by a larger margin than for example in the base scenario (8.99% to 10.87% compared with 15.75% to 16.68%). This can be explained by a larger spatial flexibility of customers (similar to the *high walking* scenario).

6.2. Analysis of proposed policy

In this section, we further analyze the value of our policy for different scenarios. We split this analysis into two parts: Firstly, we present key performance indicators (KPIs) that show the sharing potential of the proposed system. Secondly, we analyze the walking distance of customers. In most scenarios, including the base scenario, 287,716 requests arrived per day on average. This number differs for the *Manhattan* (237,609), *not Manhattan* (15,078), *low volume* (192,094), and *high volume* (383,737) scenarios.

KPIs from operator and society perspective. Table 4 shows KPIs from the operator's perspective, i.e., savings that our proposed system attains when being compared to a traditional taxi operation, where no sharing occurs. The respective standard deviations of these values are presented in Appendix A (Table 7). We can see that the proposed policy saves around 204,917 km of traveled distance per day in the base scenario compared to a traditional taxi operation and 26% of all requests can be shared with a previously planned trip. The amount of saved CO_2 can be evaluated by multiplying the average carbon emissions by the distance saved. Applying this formula, the proposed system saves about 35,000 kg of CO_2 per day in the baseline scenario assuming average carbon emissions of 0.17 kg/km (Bruck et al., 2017). A higher meeting point density leads to more distance that can be saved but also to a decrease in the share of saved requests. This seems contradictory but can be explained by the fact that the system is adopted by more customers if the meeting points are close by (due to less walking and consequently satisfying the ratio Constraints 8/13). This is reflected by the percentage of requests that were served, i.e., requests that satisfied all respective constraints, which is shown in the last column. In the high meeting point density scenario, 69% of requests were served compared to 42% in the low meeting point density scenario. Looking at the low/high walking scenarios, we can see that higher spatial flexibility of customers leads to an increase in saved distance requests, i.e., about 4.5 km per request (109,236 km/24,479 requests) are saved in the low walking scenario compared to 5.9 km per request (258,222 km/44,083 requests) saved in the high walking scenario. Another interesting observation is that not only the absolute saved distance and number of requests are lower in the *not Manhattan* scenario, but also the relative numbers. Only 6% (5%) of total requested distance (requests) is saved in areas outside of Manhattan compared to 29% (26%) in Manhattan. This observation might suggest that the economic benefits of using meeting points in a taxi-ridesharing system are higher in areas where the request density is high. This finding is supported when looking at the low/high volume scenarios, as the relative share of saved distance and requests is notably lower in the low-volume scenario. Furthermore, a high car capacity saves more distance compared to the scenario with a low car capacity (236,929 km vs. 182,866 km). Thus, when considering the implementation of meeting points, providers may also consider using (some) vehicles with higher capacities. Furthermore, we can see that the savings do not differ much between the base scenario and the high lead time scenario. Nevertheless, savings are lower when the lead time is low. Therefore, when the mean lead time is above a certain limit, its effect diminishes. The service provider should therefore incentivize people to make requests not at the last minute. Moreover, we can see that the chosen maximal ratio between walking to/from meeting points and walking directly plays a major role in the share of requests that can be served (26% vs. 81%) and the share of requests that can be saved (14% vs. 39%). Consequently, a careful determination of this ratio is advised which could also be based on a customer's individual preference. When walking to and from meeting points is combined, a substantially higher distance is saved (252,522 km) compared to the base scenario. This highlights the importance of exploiting customers' flexibility and the approach should be considered a viable option for our proposed system. From Table 7 we can see that the results are quite robust, as the standard deviations for e.g., the presented ratios of Table 4 are small and do not exceed 0.02.

KPIs from customers' perspective. From Table 4 we can see that customers benefit from walking, leading to higher sharing probabilities and consequently, lower fares could be charged. In the high walking scenario, 28% of all requests are being saved, which is notably more than in the low walking scenario, where only 16% of all requests are being saved. Table 5 shows the average walking distance per request for the benchmark policy and for our policy, respectively. The standard deviations of these numbers are shown in Table 8 (Appendix A). We note that in all our experiments, the distances to the pickup point and from the drop-off point were nearly identical.

In the table, two types of requests are distinguished. Columns *new trip* represents requests that cannot be shared the moment they appear and thus cause the scheduling of a new taxi trip. Columns *saved trip* are those requests that are assigned to a previously scheduled trip. For the benchmark policy, a difference in walking distance between *new trips* and *saved trips* can be noticed. *New trip*-customers need to walk only about 225 (451/2) meters to the pick-up points and from drop-off to their destination in almost all scenarios. The walking distance for *saved trip*-customers is in most cases between 1.25 and 1.75 times higher than the walking distance for new trips for both pickup and drop-off. This difference can be expected since the new trips are assigned to the nearest meeting points. Exceptions are scenarios in which either the density of the meeting points or the walking distance of the customers is varied. For example, a higher density of meeting points leads to a lower average walking distance, while a decrease in meeting point density leads to an increase in walking distance. These effects are higher for *new trip*-customers than for *saved trip*-customers. This is as expected as the benchmark policy assigns customers to the nearest meeting points in case the request cannot be shared.

Table 4: KPIs of our policy

	Saved distance (km)	Share of total requested distance	Saved requests	Share of saved requests	Share of served requests
Base Scenario	204,917	0.26	38,457	0.25	0.54
High MP density	226,648	0.27	46,161	0.23	0.69
Low MP density	171,103	0.24	30,736	0.25	0.42
Low walking	109,236	0.14	24,479	0.16	0.53
High walking	258,222	0.33	44,083	0.28	0.54
Manhattan	116,994	0.29	30,993	0.26	0.49
Not Manhattan	4,231	0.06	361	0.05	0.49
Low volume	110,435	0.21	20,519	0.20	0.54
High volume	313,793	0.30	59,215	0.29	0.54
Low car capacity	182,866	0.25	34,319	0.24	0.50
High car capacity	236,929	0.27	44,430	0.26	0.59
Low lead time	186,630	0.24	35,690	0.23	0.54
High lead time	206,547	0.26	38,680	0.25	0.54
Low ratio	106,813	0.18	10,799	0.14	0.26
High ratio	312,191	0.35	91,170	0.39	0.81
Low time difference	140,940	0.18	26,071	0.17	0.54
High time difference	241,410	0.31	45,780	0.30	0.54
Walking combined	252,522	0.32	43,981	0.28	0.54
Average	187,913	0.25	36,996	0.24	0.53

Table 5: Average walking distances (in meter)

	benchmark policy		proposed policy	
	new trip	saved trip	new trip	saved trip
Base Scenario	451	662	632	670
High MP density	303	614	586	632
Low MP density	589	693	656	675
Low walking	446	509	477	496
High walking	452	776	727	779
Manhattan	444	650	589	628
Not Manhattan	457	696	684	811
Low volume	450	665	634	675
High volume	451	659	630	666
Low car capacity	451	664	632	670
High car capacity	451	661	631	669
Low lead time	451	652	635	654
High lead time	451	667	631	672
Low ratio	426	618	559	691
High ratio	482	689	754	727
Low time difference	451	636	637	678
High time difference	451	626	630	663
Walking combined	451	728	701	750
Average	450	659	635	678

For our policy, overall walking distances are at a higher level compared to those for the benchmark policy, at least for the *new trip*-customers. Looking at *new trip*-customers in the base scenario, the walking distance for our policy is on average around 317 meters to the pickup point and 315 meters from the drop off point. Therefore, customers are walking 181 meters longer to meeting points when our policy is applied compared to the benchmark policy. This can be expected as the nearest meeting point policy achieves minimum walking distances by definition in case a new trip is determined. At the same time, the table shows that for *saved trip*-customers our policy leads to an increase in walking distance of only 8 meters.

While the difference in trip distance between the two request types is less for our policy than for the benchmark policy, the variation across scenarios is larger compared to the benchmark policy. The highest walking distances appear in the *high walking* and *high ratio* as well as in the *not Manhattan* and *walking combined* scenario. This shows, on the one hand, that spatial flexibility of customers is largely exploited and, on the other hand, that a low spatial density of requests also requires long walking distances. In comparison, the meeting point density has less influence on walking distances. The volume of requests as well as car capacity, the maximally allowed time difference and lead time only minimally affect the walking distance compared to the base scenario.

In essence, our policy achieves savings in trip sharing by increasing the walking distance for *new trip*-customers. While the increase in walking may come with slight inconvenience, it also increases the likelihood of sharing a trip, thus, may reduce the cost of travel. This tradeoff should be carefully considered when designing the meeting point system.

7. Conclusions

In this paper, we have shown how the anticipatory assignment of meeting points can reduce travel costs in taxi-ridesharing significantly. Furthermore, we can conclude that such an anticipatory assignment is crucial to design more sustainable on-demand ridesharing systems. Based on real-world taxi data from New York City, we showed that on average 204,917 km and 35,000 kg of CO_2 can be saved per day in a base scenario, compared to a traditional taxi system with no sharing.

We considered a simplified version of the taxi-ridesharing problem in order to analyze the meeting point assignment in adequate detail. This results in a variety of future research directions. In our problem, no intermediate stops between the pick-up and the drop-off point of a shared trip are allowed in oder to maintain high service quality. However, future models could suspend this constraint to analyze the trade-off between less direct routes and reduced walking times. Moreover, we assume that a sufficiently large fleet of taxis was available. While this applies to a metropolitan area such as New York, in other cases, the routing of the fleet may become an important optimization question. Future work may therefore combine our work with dynamic vehicle routing to ensure the availability of vehicles throughout the city, incorporating real-time traffic information and enabling changes in previous decisions in response to congestion or new demands. In our experiments, we have also seen that the capacity of vehicles plays a major role in the potential of taxi-ridesharing. However, since high-capacity vehicles are likely to be more expensive, future research may also focus on the setup and dynamic allocation of the fleet, i.e., anticipatory positioning of empty vehicles, ideally having high-capacity vehicles available in areas of higher sharing potential.

We have further shown that the distribution of meeting points and the acceptable walking distance play an important role when sharing trips via meeting points. This leads to two potential extensions. First, while we assume meeting points are equidistantly distributed over the area, the planning of their location may be another interesting avenue of research. Besides achieving good coverage of the service areas, aspects such as accessibility, visibility, or security may be considered. At the same time, decisions could be made about area-specific services (and their prices), e.g., short walking distances in high-density areas and longer walking distances in rather rural areas. Furthermore, walking could be compensated via (dynamic) pricing where customers might be nudged towards longer walking when it can reduce travel costs significantly. Moreover, it is necessary to consider the personal compatibility of ridesharing passengers, which may also be included in the pricing, for example. In our experiments, we see that a substantially higher distance is saved when walking to and from meeting points is combined rather than restricted individually. This highlights the importance of exploiting customers' flexibility and this approach should be considered a viable option for the proposed system.

Besides the potential problem extensions, there might also be potential in extending the proposed methodology and transferring it to related problems. We have shown that the components of our policy, exploiting popularity and ensuring coverage, perform well by themselves and best when combined. While in our policy the weighting of both components is static, future work could investigate a dynamic state-dependent weighting of the two components, e.g., based on the time of the day or the trips in the system. For example, it might be beneficial to focus on popularity in the morning when demand is low and focus on coverage in times when demand is high. Finally, anticipatory decision-making to allow consolidation is not limited to the problem of taxi-ridesharing but also applies to other urban on-demand services such as instant or restaurant-meal delivery where the careful balance between exploiting popularity and ensuring coverage of the service area also applies. Thus, future research may transfer our concepts to related types of on-demand transportation problems.

References

Agatz, N., Erera, A., Savelsbergh, M., Wang, X., 2012. Optimization for dynamic ride-sharing: A review. European Journal of Operational Research 223, 295–303. doi:10.1016/j.ejor.2012.05.028.

Agatz, N., Erera, A.L., Savelsbergh, M.W., Wang, X., 2011. Dynamic ride-sharing: a simulation study in metro Atlanta. Procedia - Social and Behavioral Sciences 17, 532–550. doi:10.1016/j.sbspro.2011.04.530.

Aliari, S., Haghani, A., 2022. Alternative pickup locations in taxi-sharing: A feasibility study. Transportation Research Record , 1–13doi:10.1177/03611981221104690.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D., 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proceedings of the National Academy of Sciences 114, 462–467. doi:10.1073/pnas.1611675114.

Barann, B., Beverungen, D., Müller, O., 2017. An open-data approach for quantifying the potential of taxi ridesharing. Decision Support Systems 99, 86–95. doi:10.1016/j.dss.2017.05.008.

Bohannon, R., Andrews, A., 2011. Normal walking speed: A descriptive meta-analysis. Physiotherapy 97, 182–9. doi:10.1016/j.physio.2010.12.004.

Bruck, B.P., Incerti, V., Iori, M., Vignoli, M., 2017. Minimizing CO2 emissions in a practical daily carpooling problem. Computers & Operations Research 81, 40–50. doi:10.1016/j.cor.2016.12.003.

Engelhardt, R., Bogenberger, K., 2021. Benefits of flexible boarding locations in on-demand ride-pooling systems, in: 2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), IEEE. pp. 1–6. doi:10.1109/MT-ITS49943.2021.9529284.

Fagnant, D.J., Kockelman, K.M., 2018. Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in austin, texas. Transportation 45, 143–158.

Farronato, C., MacCormack, A., Mehta, S., 2018. Innovation at uber: The launch of express pool, Harvard Business School.

Fielbaum, A., Bai, X., Alonso-Mora, J., 2021. On-demand ridesharing with optimized pick-up and drop-off walking locations. Transportation Research Part C: Emerging Technologies 126, 103061. doi:10.1016/j.trc.2021.103061.

Fielbaum, A., Kronmueller, M., Alonso-Mora, J., 2022. Anticipatory routing methods for an on-demand ridepooling mobility system. Transportation 49, 1921–1962.

Haliem, M., Mani, G., Aggarwal, V., Bhargava, B., 2021. A distributed model-free ride-sharing approach for joint matching, pricing, and dispatching using deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems 22, 7931–7942.

Ham, A., 2021. Dial-a-ride problem with meeting point feature known-as express-pool. IEEE Access 9, 86404–86411. doi:10.1109/ACCESS.2021.3089275.

Hosni, H., Naoum-Sawaya, J., Artail, H., 2014. The shared-taxi problem: Formulation and solution methods. Transportation Research Part B: Methodological 70, 303–318. doi:10.1016/j.trb.2014.09.011.

Huang, X., Peng, H., 2018. Efficient mobility-on-demand system with ride-sharing, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 3633–3638.

Kaan, L., Olinick, E.V., 2013. The vanpool assignment problem: Optimization models and solution algorithms. Computers & Industrial Engineering 66, 24–40. doi:10.1016/j.cie.2013.05.020.

Li, X., Hu, S., Fan, W., Deng, K., 2018. Modeling an enhanced ridesharing system with meet points and time windows. PloS one 13, e0195927. doi:10.1371/journal.pone.0195927.

Lin, K., Zhao, R., Xu, Z., Zhou, J., 2018. Efficient large-scale fleet management via multi-agent deep reinforcement learning, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1774–1783.

Lokhandwala, M., Cai, H., 2018. Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of nyc. Transportation Research Part C: Emerging Technologies 97, 45–60.

Lotze, C., Marszal, P., Schröder, M., Timme, M., 2022. Dynamic stop pooling for flexible and sustainable ride sharing. New Journal of Physics 24, 023034. doi:10.1088/1367-2630/ac47c9.

Lyu, Y., Lee, V.C., Ng, J.K.Y., Lim, B.Y., Liu, K., Chen, C., 2019. Flexi-sharing: a flexible and personalized taxi-sharing system. IEEE Transactions on Vehicular Technology 68, 9399–9413. doi:10.1109/TVT.2019.2932869.

Ma, S., Zheng, Y., Wolfson, O., 2013. T-share: A large-scale dynamic taxi ridesharing service, in: 2013 IEEE 29th International Conference on Data Engineering (ICDE), IEEE. pp. 410–421. doi:10.1109/ICDE.2013.6544843.

Martins, L.d.C., de la Torre, R., Corlu, C.G., Juan, A.A., Masmoudi, M.A., 2021. Optimizing ride-sharing operations in smart sustainable cities: Challenges and the need for agile algorithms. Computers & Industrial Engineering 153, 107080. doi:10.1016/j.cie.2020.107080.

Masoud, N., Jayakrishnan, R., 2017. A real-time algorithm to solve the peer-to-peer ride-matching problem in a flexible ridesharing system. Transportation Research Part B: Methodological 106, 218–236. doi:10.1016/j.trb.2017.10.006.

Miskolczi, M., Földes, D., Munkácsy, A., Jászberényi, M., 2021. Urban mobility scenarios until the 2030s. Sustainable Cities and Society 72, 103029. doi:10.1016/j.scs.2021.103029.

NYC City Taxi & Limousine Commission: TLC, 2017. Tlc trip record data. URL: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page. accessed: 14.01.2022.

Powell, W.B., 2011. Approximate Dynamic Programming: Solving the Curses of Dimensionality. Wiley Series in Probability and Statistics. 2nd ed., Wiley, Hoboken, NJ, USA.

Qian, X., Zhang, W., Ukkusuri, S.V., Yang, C., 2017. Optimal assignment and incentive design in the taxi group ride problem. Transportation Research Part B: Methodological 103, 208–226. doi:10.1016/j.trb.2017.03.001.

Qin, Z., Tang, X., Jiao, Y., Zhang, F., Xu, Z., Zhu, H., Ye, J., 2020. Ride-hailing order dispatching at didi via reinforcement learning. INFORMS Journal on Applied Analytics 50, 272–286. doi:10.1287/inte.2020.1047.

Riley, C., Van Hentenryck, P., Yuan, E., 2020. Real-time dispatching of large-scale ride-sharing systems: Integrating optimization, machine learning, and model predictive control. arXiv preprint arXiv:2003.10942 .

Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S.H., Ratti, C., 2014. Quantifying the benefits of vehicle pooling with shareability networks. Proceedings of the National Academy of Sciences 111, 13290–13294. doi:10.1073/pnas.1403657111.

Schreieck, M., Safetli, H., Siddiqui, S.A., Pflügler, C., Wiesche, M., Krcmar, H., 2016. A matching algorithm for dynamic ridesharing. Transportation Research Procedia 19, 272–285. doi:10.1016/j.trpro.2016.12.087.

Simonetto, A., Monteil, J., Gambella, C., 2019. Real-time city-scale ridesharing via linear assignment problems. Transportation Research Part C: Emerging Technologies 101, 208–232. doi:10.1016/j.trc.2019.01.019.

Smet, P., 2021. Ride sharing with flexible participants: a metaheuristic approach for large-scale problems. International Transactions in Operational Research 28, 91–118. doi:10.1111/itor.12737.

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2015. The benefits of meeting points in ride-sharing systems. Transportation Research Part B: Methodological 82, 36–53. doi:10.1016/j.trb.2015.07.025.

Stock, E., 2018. Introducing express pool: Walk a little to save a lot. URL: https://www.uber.com/newsroom/expresspool/.

Tafreshian, A., Abdolmaleki, M., Masoud, N., Wang, H., 2021. Proactive shuttle dispatching in large-scale dynamic dial-a-ride systems. Transportation Research Part B: Methodological 150, 227–259. doi:https://doi.org/10.1016/j.trb.2021.06.002.

Tafreshian, A., Masoud, N., Yin, Y., 2020. Frontiers in service science: Ride matching for peer-to-peer ride sharing: A review and future directions. Service Science 12, 44–60. doi:10.1287/serv.2020.0258.

Tong, L.C., Zhou, L., Liu, J., Zhou, X., 2017. Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing. Transportation Research Part C: Emerging Technologies 85, 451–475. doi:10.1016/j.trc.2017.09.022.

UN, 2019. World urbanization prospects: The 2018 revision. URL: https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.

Van Engelen, M., Cats, O., Post, H., Aardal, K., 2018. Enhancing flexible transport services with demand-anticipatory insertion heuristics. Transportation Research Part E: Logistics and Transportation Review 110, 110–121.

Vosooghi, R., Puchinger, J., Jankovic, M., Vouillon, A., 2019. Shared autonomous vehicle simulation and service design. Transportation Research Part C: Emerging Technologies 107, 15–33.

Wang, J., Cheng, P., Zheng, L., Feng, C., Chen, L., Lin, X., Wang, Z., 2020. Demand-aware route planning for shared mobility services. Proceedings of the VLDB Endowment 13, 979–991.

Wang, X., Agatz, N., Erera, A., 2018. Stable matching for dynamic ride-sharing systems. Transportation Science 52, 850–867. doi:10.1287/trsc.2017.0768.

Wang, Z., Hyland, M.F., Bahk, Y., Sarma, N.J., 2022. On optimizing shared-ride mobility services with walking legs. URL: https://arxiv.org/abs/2201.12639, doi:10.48550/ARXIV.2201.12639.

Wei, C., Wang, Y., Yan, X., Shao, C., 2017. Look-ahead insertion policy for a shared-taxi system based on reinforcement learning. IEEE Access 6, 5716–5726. doi:10.1109/ACCESS.2017.2769666.

Yan, L., Luo, X., Zhu, R., Santi, P., Wang, H., Wang, D., Zhang, S., Ratti, C., 2020. Quantifying and analyzing traffic emission reductions from ridesharing: A case study of shanghai. Transportation Research Part D: Transport and Environment 89, 102629. URL: https://www.sciencedirect.com/science/article/pii/S1361920920308142, doi:https://doi.org/10.1016/j.trd.2020.102629.

Yu, X., Shen, S., 2019. An integrated decomposition and approximate dynamic programming approach for on-demand ride pooling. IEEE Transactions on Intelligent Transportation Systems 21, 3811–3820.

Zhao, M., Yin, J., An, S., Wang, J., Feng, D., 2018. Ridesharing problem with flexible pickup and delivery locations for app-based transportation service: Mathematical modeling and decomposition methods. Journal of Advanced Transportation 2018. doi:10.1155/2018/6430950.

Zheng, Y., Li, W., Qiu, F., Wei, H., 2019. The benefits of introducing meeting points into flex-route transit services. Transportation Research Part C: Emerging Technologies 106, 98–112. doi:10.1016/j.trc.2019.07.012.

Zuo, H., Cao, B., Zhao, Y., Shen, B., Zheng, W., Huang, Y., 2021. High-capacity ride-sharing via shortest path clustering on large road networks. The Journal of Supercomputing 77, 4081–4106. doi:10.1007/s11227-020-03424-6.

Appendices

A. Alpha in scenarios 1-8

In this section, we present graphs visualizing the influence of α on the objective value, compared to the benchmark policy for each scenario, i.e. scenarios where we vary the meeting point density (Figure 8 and 9), walking distance (Figure 10 and 11), geographical area (Figure 12 and 13), volume (Figure 14 and 15), car capacity (Figure 16 and 17), upper lead time (Figure 18, and 19), the ratio between walking to/from meeting points and walking directly (Figure 20 and 21), the time difference (Figure 22 and 23) and the scenario in which the walking distance from and to meeting points is combined (Figure 24). The exact numbers corresponding to the above-mentioned figures are given in Table 6. Tables 7 and 8 give the standard deviations corresponding to Tables 4 and 5 in the main manuscript.

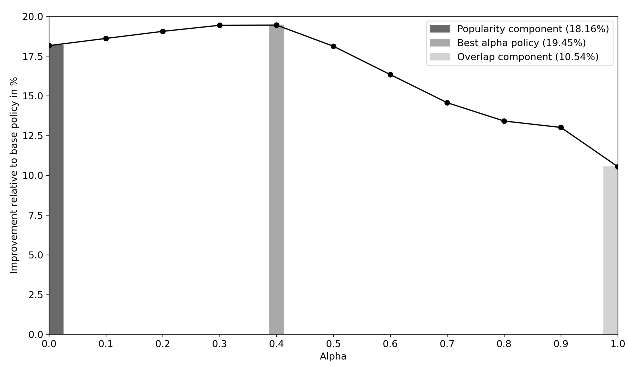


Figure 8: Role of alpha with grid density of 400 meters

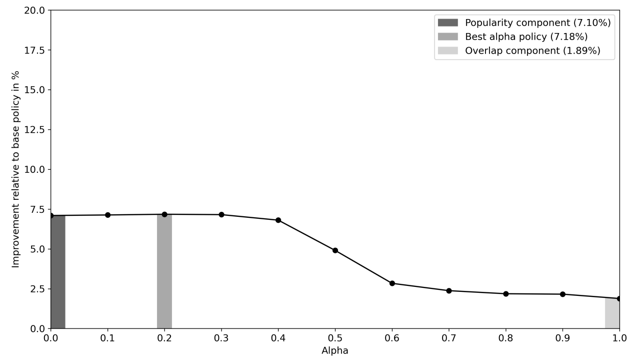


Figure 9: Role of alpha with grid density of 800 meters

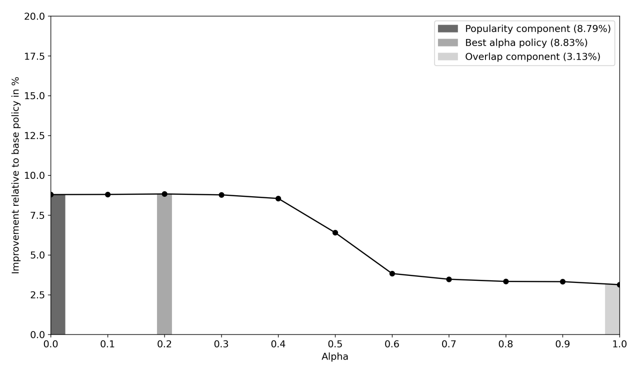


Figure 10: Role of alpha with a maximal walking time of 5 minutes

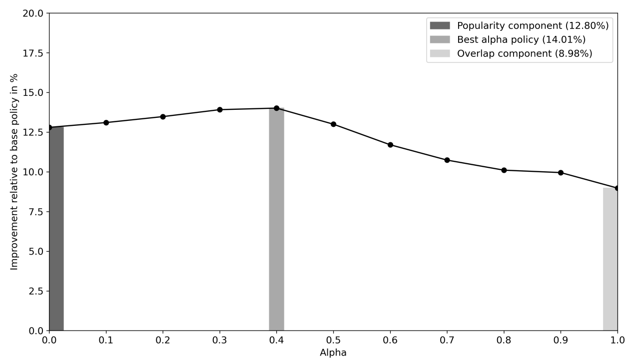


Figure 11: Role of alpha with a maximal walking time of 10 minutes

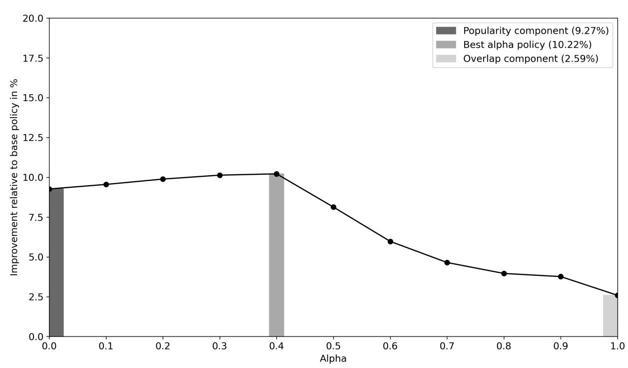


Figure 12: Role of alpha for requests in Manhattan

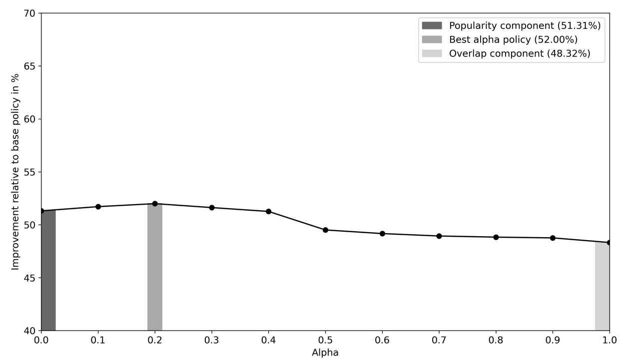


Figure 13: Role of alpha for requests not in Manhattan

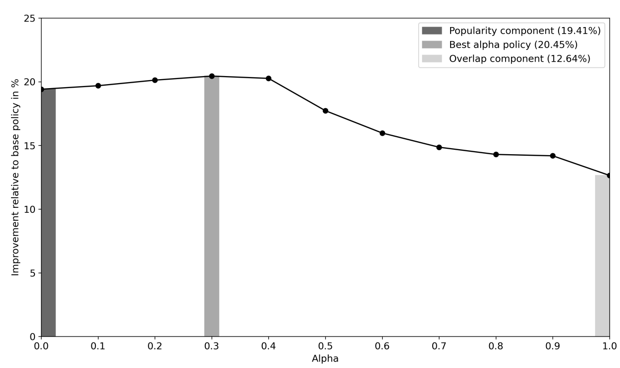


Figure 14: Role of alpha for a volume of 0.5

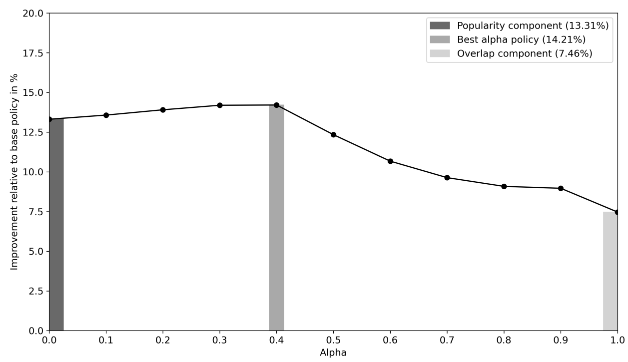


Figure 15: Role of alpha for a volume of 1

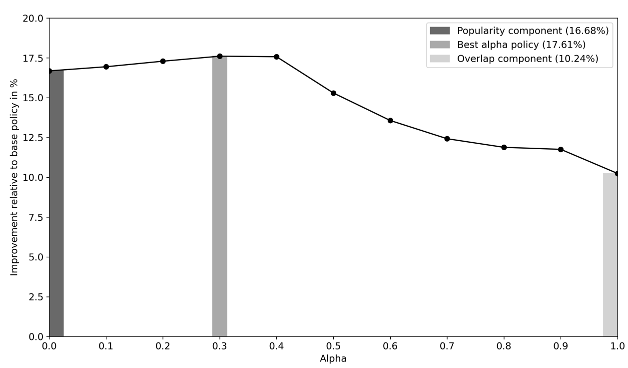


Figure 16: Role of alpha for car capacity of 2

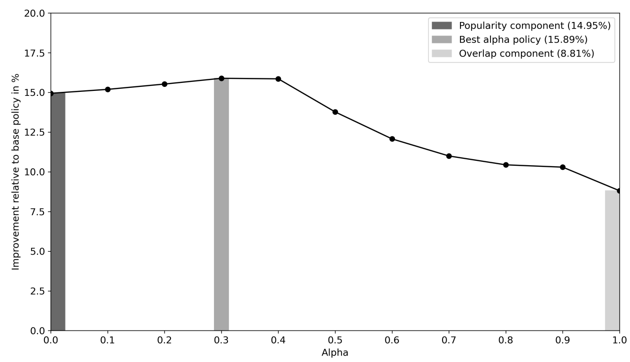


Figure 17: Role of alpha for car capacity of 8

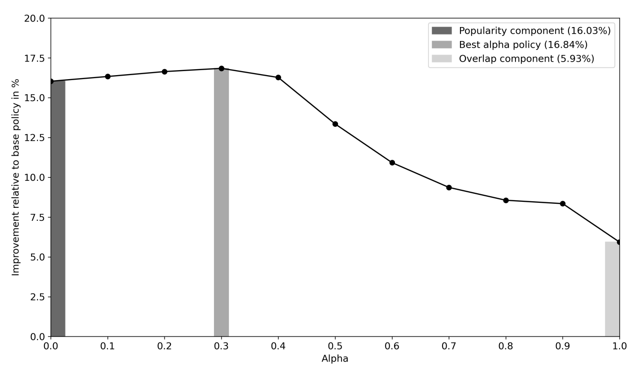


Figure 18: Role of alpha for an upper lead time of 900 seconds

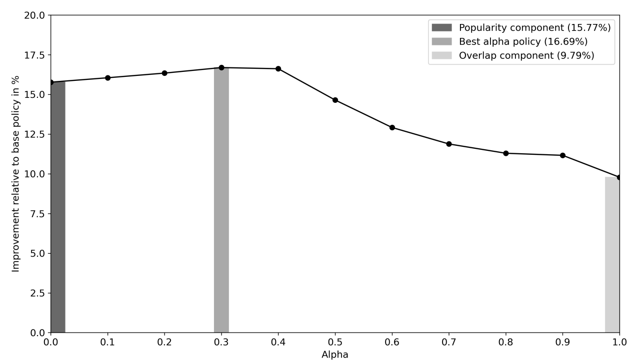


Figure 19: Role of alpha for an upper lead time of 3600 seconds

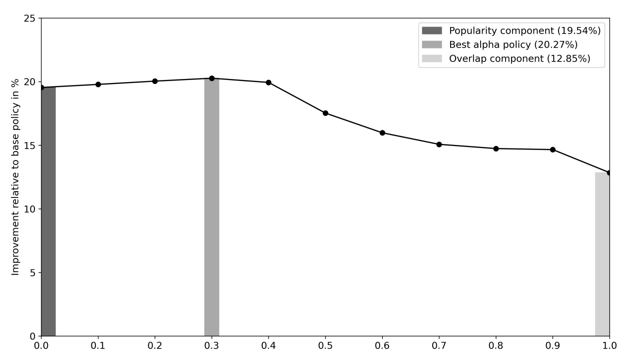


Figure 20: Role of alpha for a maximally allowed ratio of 0.125

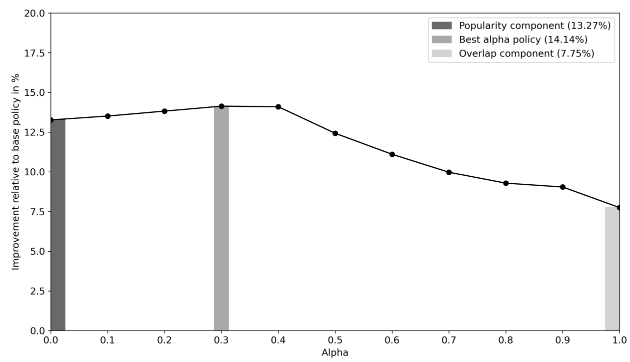


Figure 21: Role of alpha for a maximally allowed ratio of 0.5

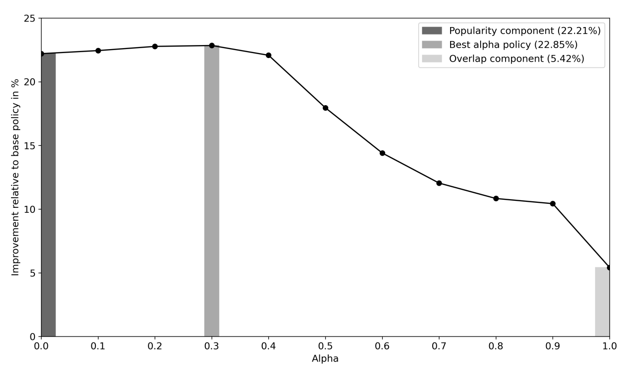


Figure 22: Role of alpha for a maximally allowed time difference of 150 seconds

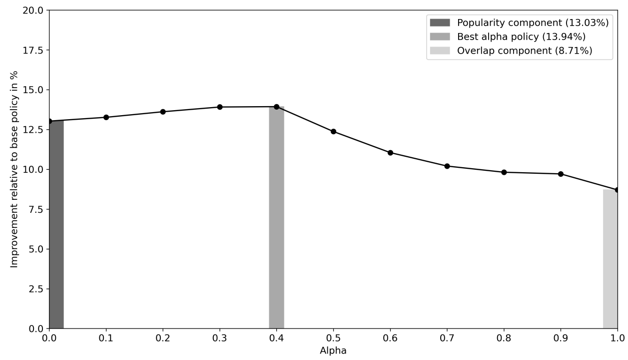


Figure 23: Role of alpha for a maximally allowed time difference of 450 seconds

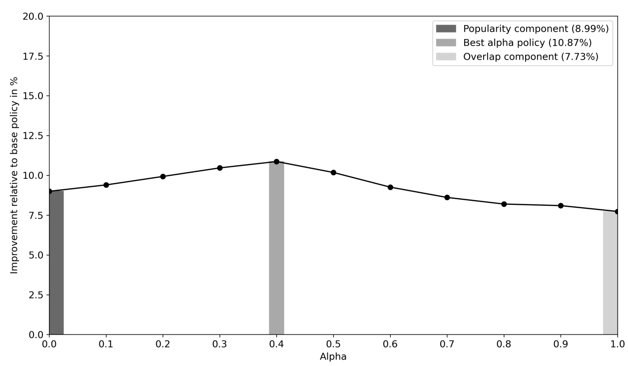


Figure 24: Role of alpha for if pick-up and drop-off walking are combined

Table 6: Influence of alpha on improvement to base policy

Scenario/Alpha	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Base Scenario	15.75	16.00	16.37	**16.68**	16.66	14.45	12.76	11.67	11.07	10.94	9.52
High MP density	18.16	18.61	19.06	19.44	**19.45**	18.12	16.33	14.57	13.41	13.01	10.54
Low MP density	7.10	7.13	**7.18**	7.15	6.81	4.91	2.84	2.38	2.19	2.16	1.89
Low walking	8.79	8.80	**8.83**	8.77	8.54	6.40	3.83	3.47	3.34	3.32	3.13
High walking	12.80	13.10	13.48	13.92	**14.01**	13.00	11.70	10.74	10.10	9.95	8.98
Manhattan	9.27	9.55	9.89	10.14	**10.22**	8.13	5.97	4.65	3.96	3.76	2.59
Not Manhattan	51.31	51.71	**52.00**	51.62	51.26	49.51	49.17	48.94	48.83	48.76	48.32
Low volume	19.41	19.69	20.13	**20.45**	20.27	17.73	15.98	14.86	14.30	14.19	12.64
High volume	13.31	13.57	13.91	14.19	**14.21**	12.34	10.67	9.63	9.09	8.96	7.46
Low car capacity	16.68	16.94	17.29	**17.61**	17.57	15.29	13.57	12.42	11.88	11.76	10.24
High car capacity	14.95	15.19	15.53	**15.89**	15.86	13.77	12.07	11.00	10.44	10.29	8.81
Low lead time	16.03	16.33	16.64	**16.84**	16.27	13.35	10.92	9.36	8.56	8.35	5.93
High lead time	15.77	16.05	16.34	**16.69**	16.62	14.65	12.92	11.88	11.30	11.16	9.79
Low ratio	19.54	19.78	20.04	**20.27**	19.94	17.53	15.98	15.07	14.74	14.66	12.85
High ratio	13.27	13.51	13.82	**14.14**	14.10	12.43	11.11	9.97	9.29	9.04	7.75
Low time difference	22.21	22.45	22.78	**22.85**	22.08	17.96	14.41	12.05	10.84	10.43	5.42
High time difference	13.03	13.27	13.62	13.91	**13.94**	12.37	11.05	10.20	9.82	9.71	8.71
Combined	8.99	9.40	9.93	10.47	**10.87**	10.18	9.26	8.61	8.20	8.10	7.73

Table 7: Standard deviations of KPIs of our policy

	Saved distance (km)	Share of total requested distance	Saved requests	Share of saved requests	Share of served requests
Base Scenario	20854.27	0.01	4863.99	0.01	0.02
High MP density	22410.31	0.01	5815.09	0.01	0.01
Low MP density	18654.80	0.02	4065.24	0.02	0.02
Low walking	12406.80	0.01	3417.69	0.01	0.02
High walking	25099.14	0.02	5309.16	0.01	0.02
Manhattan	20175.60	0.02	5101.38	0.02	0.02
Not Manhattan	1066.09	0.01	91.82	0.01	0.02
Low volume	11745.36	0.01	2704.18	0.01	0.02
High volume	30758.68	0.02	7313.11	0.01	0.02
Low car capacity	19064.74	0.02	4238.00	0.01	0.01
High car capacity	24329.34	0.02	5660.22	0.01	0.02
Low lead time	19348.06	0.01	4623.12	0.01	0.02
High lead time	20763.26	0.01	4844.81	0.01	0.02
Low ratio	15413.65	0.02	1349.33	0.02	0.02
High ratio	29367.90	0.02	10431.70	0.02	0.01
Low time difference	15729.16	0.01	3511.48	0.01	0.02
High time difference	23920.01	0.02	5625.60	0.02	0.02
Walking combined	24691.97	0.02	5358.48	0.01	0.02

Table 8: Standard deviations of average walking distances (in meter)

	benchmark policy		proposed policy	
	new trip	saved trip	new trip	saved trip
Base Scenario	2.5	7.5	5.9	10.6
High MP density	1.0	8.5	8.4	11.1
Low MP density	2.3	6.4	3.4	6.1
Low walking	2.1	4.4	2.7	4.3
High walking	2.6	12.9	9.6	14.8
Manhattan	3.1	8.7	6.7	8.5
Not Manhattan	3.4	23.7	17.7	29.5
Low volume	2.5	7.8	6.0	11.4
High volume	2.5	7.6	6.0	10.1
Low car capacity	2.5	7.9	6.1	10.8
High car capacity	2.4	7.6	5.8	10.6
Low lead time	2.5	7.4	6.2	9.8
High lead time	2.5	7.9	6.1	11.0
Low ratio	2.8	21.3	9.0	33.7
High ratio	1.0	9.0	4.0	7.9
Low time difference	2.6	7.4	6.3	11.2
High time difference	2.5	7.8	6.0	10.5
Combined	2.7	13.5	9.1	13.1