



**HAL**  
open science

## Towards Data-Driven On-Demand Transport

Malcolm Egan, Jan Drchal, Jan Mrkos, Michal Jakob

► **To cite this version:**

Malcolm Egan, Jan Drchal, Jan Mrkos, Michal Jakob. Towards Data-Driven On-Demand Transport. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 2018, 5 (14), pp.1-10. 10.4108/eai.27-6-2018.154835 . hal-01839452

**HAL Id: hal-01839452**

**<https://hal.science/hal-01839452>**

Submitted on 16 Jul 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Towards Data-Driven On-Demand Transport

Malcolm Egan<sup>1,\*</sup>, Jan Drchal<sup>2</sup>, Jan Mrkos<sup>2</sup> and Michal Jakob<sup>2</sup>

<sup>1</sup>CITI, INSA Lyon, INRIA, Université de Lyon, France

<sup>2</sup>Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

## Abstract

On-demand transport has been disrupted by Uber and other providers, which are challenging the traditional approach adopted by taxi services. Instead of using fixed passenger pricing and driver payments, there is now the possibility of adaptation to changes in demand and supply. Properly designed, this new approach can lead to desirable tradeoffs between passenger prices, individual driver profits and provider revenue. However, pricing and allocations—known as mechanisms—are challenging problems falling in the intersection of economics and computer science. In this paper, we develop a general framework to classify mechanisms in on-demand transport. Moreover, we show that data is key to optimizing each mechanism and analyze a dataset provided by a real-world on-demand transport provider. This analysis provides valuable new insights into efficient pricing and allocation in on-demand transport.

Received on 09 January 2018; accepted on 30 March 2018; published on 27 June 2018

**Keywords:** on-demand transport, market mechanisms, auctions, taxis

Copyright © 2018 Malcolm Egan *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.27-6-2018.154835

## 1. Introduction

Catching a cab has changed over the last few years. In New York City at the turn of the millennium, hailing a cab from the side of the road was the norm. Or in Sydney, it was common to call a taxi dispatcher for a journey to the airport. Now in 2017, it is increasingly common to instead open an app on a smartphone to book transport via Uber, Lyft, or one of several other on-demand transport services.

What has changed? After the now widespread adoption of internet-enabled smartphones and secure online payments, it is possible for both potential passengers and drivers to easily communicate with a provider or directly with each other. This offers the potential for passengers to be transported, irrespective of their location and the locations of available drivers. The problem that remains is to determine which driver should transport which passenger and how much they should pay—the allocation and pricing.

The high level of connectivity that we now take for granted can also assist in allocation and pricing. In particular, the use of smartphone apps has made it possible for providers to collect huge amounts of data,

such as the time and location of each passenger's pick-up and drop-off, as well as the prices that passengers are prepared to pay and the fares drivers are willing to accept. The combination of this data is leading to new insights into the spatial-temporal profile of on-demand transport systems [1–4] and ultimately opens the way for increasingly efficient allocation and pricing.

In fact, data-driven allocation and pricing is now an active area of research and plays a key role in real-world on-demand transport services [5–10]. By accounting for the real-time behavior of passengers and drivers, it is possible to improve the reliability of on-demand transport services in terms of waiting times, driver profits, and also provider revenue.

In this paper, we develop a general framework to design allocation and pricing algorithms within on-demand transport, with a focus on profit-driven services; that is, we focus on taxi-like services rather than on dial-a-ride services [11]. In particular, we show that existing on-demand transport mechanisms are different variants of two-sided markets. Our framework provides a unifying perspective for existing allocation and pricing algorithms, which can be viewed as implementations of market mechanisms from economics in the context of on-demand transport.

\*Corresponding author. Email: [malcom.egan@insa-lyon.fr](mailto:malcom.egan@insa-lyon.fr)

Key examples are auctions [12] and posted price mechanisms [13].

A key feature of our framework is that it clarifies the requirements of different mechanisms, which forms a basis for the mechanism selection problem that providers face. In fact, we identify allocation and pricing algorithms within our framework that have desirable properties, but have not yet received significant attention.

The algorithms in our framework all rely on information either obtained directly from real-time passenger requests and driver reports, or from historical data collected from previous requests. In this sense, effective mechanisms for pricing and allocation in on-demand transport are data-driven. It is therefore important to develop approaches to obtain and exploit this data.

There are two aspects of the mechanisms that require data. The first aspect is the optimization of the pricing and payments for each mechanism in our framework. In auction-based mechanisms data is obtained directly from passengers and driver reports (or bids). However, in practice an alternative approach based on posted price mechanisms—where the provider makes offers, which are accepted or rejected—has proven most popular. In the posted price mechanism, the prices are selected based on historical transactions. When there is limited data, an online pricing approach is desirable, where the price is adapted as new data becomes available.

The second aspect is the optimization of allocations of passengers and drivers. While each market mechanism in our framework provides an allocation algorithm for a given set of passengers and drivers, it does not provide a means of selecting this set—known as the market formation problem [14]. Due to the huge scale and heterogeneity of on-demand transport systems, it is not possible to allocate all passengers and drivers simultaneously. As such, market formation is an important practical issue.

At present, there are limited data-driven methods available for the market formation problem, with most commonly used approaches reliant on heuristics guided by experienced practitioners. To gain insight into efficient market formation, we apply machine learning methods to real-world data from the provider Liftago in the Czech Republic. In particular, we identify key features that determine which drivers are willing to be allocated to a passenger. The features include not only factors such as distance but also driver-dependent factors such as historical acceptance rates. A key observation is also that the accuracy of the feature ranking is highly dependent on the availability of data.

In the remainder of the paper, we first detail our classification framework for on-demand transport mechanism in Section 2. Our framework provides a

way to systematically identify the requirements of each pricing and allocation algorithm, which forms a basis for mechanism selection by on-demand transport providers. In Section 3, we discuss the market formation problem and present the results of the analysis of the Liftago dataset. We conclude and outline future directions in Section 4.

## 2. On-Demand Transport Market Mechanisms: A Classification

Two key problems in on-demand transport are: (1) *how to allocate passengers to drivers*; and (2) *how to determine passenger prices and driver payments*. For traditional taxi services, the allocation is based on passengers hailing from the side of the road (the “hackney carriage” model) or through a phone call to a dispatcher who then selects a driver nearby the passenger (the “dispatcher” model). Pricing is typically determined either by regulation or through a negotiation between the driver and each potential passenger.

On the other hand, new on-demand transport services provided by companies such as Uber and Lyft are using a different method. By exploiting a smartphone app, passengers are connected to a number of drivers within their region and are able to select the one they prefer. How much passengers pay and the commission drivers receive is determined by the provider through a pricing mechanism. This allocation and pricing scheme is an example of a posted price mechanism [13], where the provider determines the price and passengers or drivers are able to accept or reject their offer.

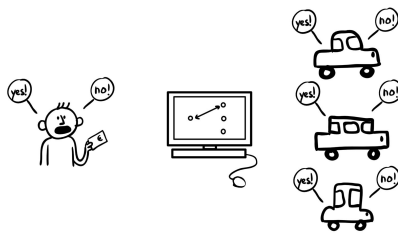
Posted price mechanisms in on-demand transport are themselves a class of two-sided market, where a service is provided by one side (i.e., the drivers) and bought by the other side (i.e., the passengers). Two-sided markets are an active area of research in both economics and computer science, aiming to efficiently allocate resources or services [15]. In this section, we develop a framework to unify the different approaches based on the fact that they are different variants of two-sided markets. Table 1 provides an overview of the framework, presenting the structure and challenges of the four classes of mechanisms: posted-price; double auction; and hybrids.

### 2.1. Posted-Price Mechanisms

Posted price (PP) mechanisms (illustrated in Fig. 1) are currently the most popular approach to two-sided markets within on-demand transport. In this approach, passengers and drivers are each offered a price and a payment, respectively, which they can either accept or reject. More generally, PP mechanisms play an important role in a range of economic problems. The posted price mechanism is summarized in Algorithm 1.

**Table 1.** Summary of Framework for Mechanisms in On-Demand Transport.

Mechanism	Passenger Side	Driver Side	Requirements
Posted price	(One-sided) Posted price	(One-sided) Posted price	Historical data to set prices, driver selection policy.
Double auction	Auction	Auction	Bidding from both passengers and drivers.
Hybrid PP/A	(One-sided) Posted price	Auction	Historical data to set passenger prices, driver bidding.
Hybrid A/PP	Auction	(One-sided) Posted price	Passenger bidding, historical data for driver payments, driver selection policy.



**Figure 1.** Illustration of the posted-price mechanism. In a posted-price mechanism, a single passenger is offered a price by the provider and is able to accept or reject the offer. A number of drivers are then offered the passenger's journey and a commission. If a driver accepts the journey, the passenger is served by the driver.

---

**Algorithm 1** Posted Price Mechanism [13].

---

**Step 1.** For each passenger request, the provider selects a driver.

**Step 2.** The provider offers the passenger a price  $p$  and the driver a payment  $\eta$ .

**Step 3.**

(a) If both the passenger and driver accept ( $p_i \geq p$  and  $\eta \geq \eta_j$ ), the passenger  $i$  is transported by the driver  $j$ .

(b) Otherwise, the passenger request is withdrawn.

---

A key advantage of PP mechanisms is that they require limited information from *each* passenger and driver. Instead, the prices and payments are set based on historical data. This is important as each passenger and driver is not always well informed about the state of the whole on-demand transport system. As such, the price calculations can be done by the provider which

can simplify the use of the service for passengers and drivers.

PP mechanisms are commonly studied as the multi-period pricing problem [16, 17]. Here, the revenue and proportion of served passengers is optimized accounting for parameters including the number of taxis, average passenger demand, and average waiting times. The basis for this optimization problem is a model relating the provider's revenue to these parameters. An example is the model developed in [18], where the average waiting time of passengers  $W$ , is given by

$$W = \frac{\omega}{pN_T - \frac{DL}{\gamma}}, \quad (1)$$

where  $\omega$  depends on the density of taxi stands,  $p$  is the proportion of available taxis,  $N_T$  is the total number of taxis,  $\gamma$  is the average number of passengers per journey,  $L$  is the average travel time, and  $D$  is the average number of passengers served by the whole taxi system. As such,  $pN_T - \frac{DL}{\gamma}$  represents the average number of available taxis. We remark that [18] focuses on the role of taxi regulation; however, this is under the assumption of a PP mechanism. Recently, this model has also been extended to account for strategic behavior of individual drivers [6, 19].

A drawback of the model introduced in [18] is that the preferences of individual passengers are not explicitly modeled. In fact, the model relies on the average behavior of passengers and drivers. As such, the full statistical information in historical data is not exploited. To remedy this problem, another class of models has been proposed in [7, 20] where each passenger  $i$  is treated as an *agent*, with a maximum price  $p_i$  and a maximum waiting time  $\delta_i$  they are willing to accept, which are random variables drawn from beta distributions. The beta distribution is selected due to the fact that it has bounded support (passengers are not willing to pay arbitrarily high prices), as well as flexibility as it is parameterized by three parameters and includes many standard distributions as special cases (e.g., uniform).

In general socio-economic systems, it is often challenging to understand the underlying processes that cause agents to behave in certain ways. As such, an important methodology is to instead consider *stylized facts*, which are qualitative trends satisfied by the system [21, 22]. In the context of on-demand transport, the model in [7] is motivated by the two following stylized facts, which form common qualitative trends in on-demand transport systems:

**Stylized Fact 1.** The maximum price a given passenger will pay for a journey and her maximum waiting time  $\delta$  do not vary significantly for passengers that regularly use on-demand transport services. In other words, these parameters will in general vary from passenger to passenger, but not for the same passenger.

**Stylized Fact 2.** The probability that a passenger will accept an offer decreases when either the price increases with the waiting time fixed, or the waiting time increases with the price fixed.

The first stylized fact ensures that passengers have a well-defined maximum price and the second stylized fact captures the intuitive notion that if there is a better deal, more passengers will accept. These stylized facts provide an important basis for the passenger models. We note that they are not guaranteed to hold in every situation, and for such scenarios the algorithms may require additional tuning. Nevertheless, these algorithms can still provide improvements over just considering the average behavior as in [6].

As a consequence of the model in [7, 20], the provider can set the price in the PP mechanism to maximize its expected revenue for short waiting times. In particular, the expected revenue is given by

$$R(p) = p\Pr(p \leq p_i), \quad (2)$$

where  $\Pr(p \leq p_i)$  is the probability that passenger  $i$  accepts the offered price  $p$ . A similar approach can be applied to optimize the provider's percentage commission,  $\eta$ , from a journey of driver  $j$  by maximizing

$$C(\eta) = (1 - \eta)\Pr(\eta > \eta_j), \quad (3)$$

where  $C(\eta)$  is the provider's expected commission,  $\eta_j$  is the minimum acceptable percentage commission of driver  $j$  and  $\Pr(\eta > \eta_j)$  is the probability that the driver is willing to accept a commission  $\eta$ .

Observe that the PP mechanism relies on historical data through the distribution of the maximum price

$p_i$ . As the distribution is modeled via the beta distribution, the problem reduces to estimating three parameters. However, it may not always be possible for drivers to have sufficient historical data to reliably estimate the parameters for the distributions of passenger maximum payments and driver minimum percentage commissions. In this situation, the price and commission need to be adapted as new data becomes available.

A promising approach to exploit new data is to use algorithms for the multi-armed bandit problem. In particular, passengers are offered a price within the set  $\mathcal{P} = \{p_1, \dots, p_N\}$ . Using historical observations of the acceptance of each price, the price for the next passenger is updated according to the UCB1 algorithm, originally developed in [23]. This approach was proposed in a pure economic context in [24] and can be adopted in the on-demand transport setting as detailed in Algorithm 2.

---

**Algorithm 2** Multi-armed bandit pricing algorithm [24].

---

**Initialization:** For passenger  $i$ ,  $1 \leq i \leq N$ , allocate price  $p_i$ .

**Loop:** Allocate the price  $p_j$  that maximizes  $\bar{x}_j + \sqrt{\frac{2 \log n}{n_j}}$ , where  $\bar{x}_j$  is the average profit received so far from using price  $p_j$ ,  $n_j$  is the number of times price  $p_j$  has been allocated, and  $n$  is the total number of requests.

---

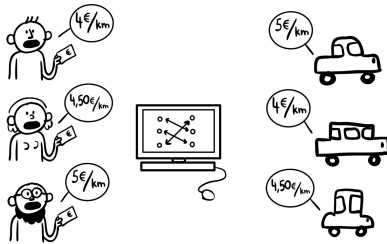
Intuitively, the algorithm trades off exploration of new (potentially higher) prices with exploitation of prices that are known to be successful. The analysis of the performance of multi-armed bandit pricing algorithms in on-demand transport remains to be fully explored.

## 2.2. Double Auction Mechanisms

In economics, a single-sided auction is a means of pricing goods or services where buyers (or sellers) bid for the price they are prepared to pay (or sell) [25]. Classical examples are the English or sealed bid auctions used in property sales or for antiques and artwork. More generally, two-sided or double auctions can be applied in situations where there are both multiple buyers and multiple sellers.

In the context of on-demand transport, a double auction (illustrated in Fig. 2) allows for passengers to disclose the price they are prepared to pay and drivers to set the fare they are prepared to accept in order to provide transport. In contrast, PP mechanisms only allow passengers and drivers to accept or reject prices set by the provider. For this reason, double auctions can lead to more efficient allocation of services.

In fact, the role of a double auction mechanism (to simultaneously price and allocate both passengers and drivers) implemented by a provider is to decide when passengers and drivers can be matched and how much they should pay, given their bids.



**Figure 2.** Illustration of the double auction mechanism. In double auction mechanisms, each passenger makes a bid for the price of their journey, while each driver makes a bid for their commission. The provider then acts as a matchmaker to determine which passenger is served by each driver.

A key problem in double auctions is to account for the strategic behavior of the agents (i.e., passengers and drivers) involved. In fact, unlike traditional taxi models [18], the individual preferences of passengers and drivers influence the design of the mechanism. This is caused by the provider's lack of knowledge concerning how much passengers are prepared to pay and the minimum payment drivers are willing to accept, which requires the provider to elicit these preferences from both passengers and drivers via reporting (or bidding).

In the case of PP mechanisms, agents either accept or reject their offers, which means that agents have no incentive to report (or bid) untruthfully—clearly, accepting an unwanted service is not a rational strategy. However, this is not necessarily the case in double auctions and it is desirable to ensure that agents bid truthfully, the provider does not lose money (weak budget balance), and agents that bid truthfully have an incentive to participate (individual rationality). For a detailed discussion of these concepts, see [25].

Double auction mechanisms have been extensively studied within economics and computer science, and fall into two main classes: static and online. In the static case, all passengers and drivers to be allocated are known to the provider. In the dynamic case, passengers and drivers can arrive at different times which must be accounted for by the mechanism. A key desiderata in both cases is to ensure that the mechanism is efficient, which means that the total reward for buyers and sellers for receiving or providing the service is maximized. In

general, optimal efficiency is not possible to achieve while also ensuring truthfulness, budget balance and individual rationality [26].

The application of classical double auctions in on-demand transport has not proven straightforward. One reason for this is that the purpose of the market can differ compared with the classical economic setting. In particular, the provider seeks to make a long-term profit, while the main related criterion in double auction theory is budget balance which only accounts for the short-term profit. As such, verifying the profitability of double auction mechanisms in on-demand transport relies heavily on simulation. This means that specialized on-demand transport simulation tools such as the Mobility Services Testbed [27] are required.

Another difficulty facing the application of double auctions is that allocations depend heavily on the financial resources of each passenger. For this reason, the passengers that are allocated may not be those that desire the service the most. As such, it is likely that a key application of double auction mechanisms to on-demand transport is for services targeted at businesses (to transport staff within a city) to minimize the effect of financial inequality.

At present, the application of double auctions in on-demand transport has focused on the static case where all passengers and drivers are known before the allocation and pricing begins. In particular, the mechanism is based on the approach by McAfee [26], which is known to have the desirable properties of truthfulness, budget balance and individual rationality. In particular, the work in [8–10] adapted the McAfee mechanism and investigated the average number of transactions and efficiency for each mechanism run.

There are two stages of the static double auction mechanism in [8]:

1. Decompose the market consisting of all passengers and drivers into markets.
2. Allocate passengers to drivers and set prices.

The first stage is necessary to ensure that the locations of passengers and drivers in each market lie within the same region. The reason for this is that the McAfee double auction mechanism requires *homogeneous goods*, which means that the passenger initial locations and requests as well as the initial locations of the drivers are similar.

The second stage of the static double auction approach in [8] is to allocate passengers to drivers using the McAfee mechanism. The allocation algorithm is detailed in Algorithm 3. In the algorithm, Steps 1 and 2 sort the passengers and drivers based on their reported maximum prices and minimum payments, respectively. Note that it is not possible to directly compare the

reports of the passengers and drivers directly due to the fact that drivers report their minimum profit. As such, in Step 1 the price of each passenger is modified to account for the distance of their journey and the initial distance to each driver. Steps 3-5 then determine the prices and payments based on the McAfee mechanism [26].

**Algorithm 3** Double auction mechanism pricing and allocation based on the McAfee mechanism [26].

**Require:** Passengers broadcast  $\{p_i\}_{i=1}^K$  (the maximum price they are prepared to pay for their journey), and drivers broadcast  $\{s_j\}_{j=1}^N$  (the minimum profit they are prepared to receive for their next journey).

**Step 1.** For each commuter, compute  $p'_i = p_i - cR_{\max} - cR_{0,\max}$ , where  $c$  is the cost per kilometer for each driver,  $R_{\max}$  is the maximum journey distance and  $R_{0,\max}$  is the maximum distance of the driver from a passenger.

**Step 2.** Sort passengers:  $p'_{(1)} \geq p'_{(2)} \geq \dots \geq p'_{(K)}$ ; drivers:  $s_{(1)} \leq s_{(2)} \leq \dots \leq s_{(N)}$ .

**Step 3.** Compute the number of efficient trades:  $k^* = \max\{k : p'_{(k)} \geq s_{(k)}, p'_{(k+1)} < s_{(k+1)}\}$  and compute  $p_0 = \frac{1}{2}(p'_{(k+1)} + s_{(k+1)})$ .

**Step 4.** Check the McAfee condition:

- (a) If  $p_0 \in [s_{(k^*)}, p'_{(k^*)}]$ , then the actual prices for drivers and passengers are  $s = p' = p_0$  and all efficient pairs are allocated;
- (b) Otherwise the prices for the drivers and passengers are  $s = s_{(k^*)}$ ,  $p' = p'_{(k^*)}$  and  $k^*$  pairs are allocated.

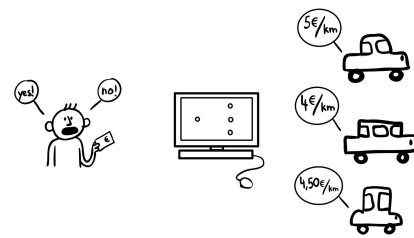
**Step 5.** Passengers are then required to pay  $p = cR_{\max} + cR_{0,\max} + p'$  and each driver  $j$  who transports commuter  $i$  is paid  $s + cR_i + cR_{j,i}$ .

To the best of our knowledge there is no work theory or practice applying online double auctions on-demand transport. However, in certain applications it may be desirable to adopt an online approach, which there is a firm basis in the economics and computer science literature (see, e.g., [15]).

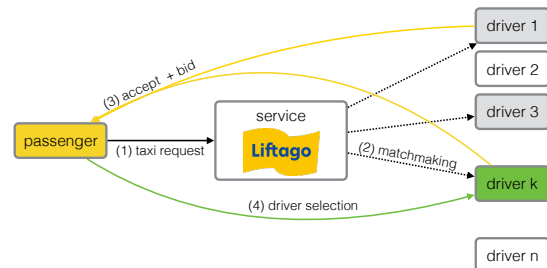
### 2.3. Hybrid Mechanisms

So far, we have examined allocation and pricing for on-demand transport where a PP mechanism or a double auction is applied to both passengers and drivers. It is also possible to consider the class of mechanisms where passengers are priced based on a PP mechanism and drivers based on a single-sided auction mechanism (the hybrid PP/A mechanism), or vice-versa (the hybrid A/PP mechanism). However, investigations into hybrid mechanisms remain ongoing work [28].

A particularly interesting hybrid mechanism is the application of a PP mechanism to price passengers and a single-sided auction to set driver payments (the hybrid PP/A mechanism, illustrated in Fig. 3). Although there has been no formal analysis of this mechanism, it has been put into practice by Liftago in the Czech Republic. Liftago's mechanism proceeds (illustrated in Fig. 4) as detailed in Algorithm 4:



**Figure 3.** Illustration of the hybrid PP/A mechanism. In hybrid PP/A mechanism, a passenger is offered a price by the provider and is free to accept or reject the offer. Each driver then bids for the passenger.



**Figure 4.** Liftago service overview. It takes four steps to issue a ride order: passenger sends a request (1), Liftago designates a set of drivers (2), a subset of which accepts the request and offers a price (3), finally (4) the passenger selects a single offer.

**Algorithm 4** Liftago mechanism description.

- Step 1:** A passenger makes a request to Liftago.
- Step 2:** Liftago identifies a subset of drivers to potentially serve the passenger.
- Step 3:** Each driver bids to serve the passenger, or ignores the request.
- Step 4:** The passenger selects one of the drivers and pays the corresponding bid.

Even in competition with traditional taxis and other providers such as Uber, Liftago has managed to remain financially sustainable using the hybrid mechanism. This suggests that a fruitful avenue of research is an analysis of how the hybrid mechanism compares with traditional taxis and other market mechanisms. An initial investigation is currently being carried out in [28].

While Liftago's mechanism has several desirable features, a key challenge is the design of algorithms for **Step 2**. This step is known as market formation and is analogous to the first stage of the double auction mechanism in [8] detailed in Section 2.2. However, unlike in the double auction mechanism the provider does not have bids from both passengers and drivers. As such, there is a reliance on historical transaction data. In the next section, we focus on data-driven aspects of the market formation problem in the context of on-demand transport (see [14] for a more general perspective).

### 3. The Market Formation Problem

Each market mechanism has different features that make it appropriate for different applications of on-demand transport, whether it be a PP mechanism, double auction, or a hybrid. However, all the mechanisms have a common feature: they require a market. In particular, not every passenger and driver can be potentially matched. As such, it is necessary for the provider to make a selection of passengers and drivers that are allocated at the same time.

Ideally, market formation [14] should be performed so that the number of drivers that compete for each passenger is minimized, while still ensuring that each passenger is able to be serviced. The most basic market formation approach simply ensures that *hard constraints* are met. These hard constraints include the maximum travel time between the initial locations of drivers and the pick-up locations of potential passenger. Other hard constraints may be based on other preferences of passengers, which may limit the vehicle class of each driver, or driver reputation. Simply accounting for hard constraints ensures that passengers are not offered a service from any driver that does not satisfy these hard constraints.

In general, hard constraints such as maximum travel time do not rule out a large number of potential drivers [8]. As such, drivers may need to respond to many requests. This can reduce the safety of drivers as they need to concentrate on responding to requests, rather than focusing on driving. Moreover, for a given passenger request, many drivers that are offered a passenger's journey may not want to serve the passenger. This may be due to passenger characteristics such as desired drop-off location or passenger reputation.

To further reduce the number of drivers that can be potentially matched to each passenger, it is possible to exploit heuristics based on the advice of experienced practitioners. For instance, provider experience may suggest that a fixed number of drivers are allowed to be matched with each passenger, where each driver is selected randomly from those that satisfy the hard constraints. More sophisticated heuristics may be based on passenger features such as drop-off location as, for example, passengers that wish to travel to transport hubs (e.g., airports) are significantly more desirable than passengers that wish to travel to locations where there is unlikely to be another passenger [29]. In these cases, it may be desirable to offer more drivers the journey of passengers with undesirable drop-off locations to ensure that the chance they are served is maximized.

Heuristics based on expert support are built using the experience of providers. This approach can be improved by exploiting the availability of historical data, which leads to *data-driven market formation*. In data-driven market formation, systematic machine learning techniques are applied to select drivers and passengers that can be potentially matched.

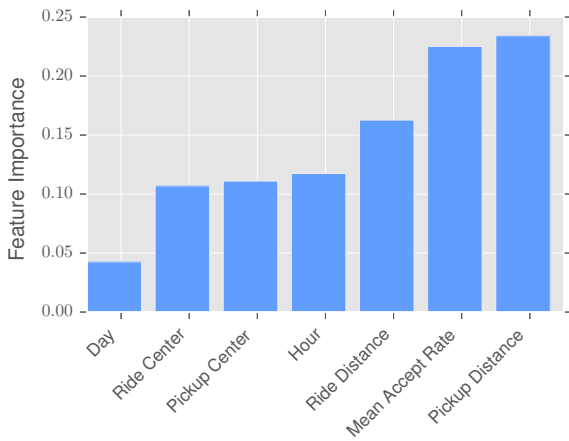
At present, there are limited data-driven algorithms for the market formation problem (initial work in this direction can be found in [3, 4], designed for PP mechanisms). The first step to developing these algorithms is to apply machine learning techniques to real-world data. To this end, we exploit a dataset provided by Liftago, which employs a hybrid mechanism as described in Section 2.3. Although this data is specific to Liftago, the analysis methodology can be applied to any provider's data set and the features provide important insights into the role of both per journey factors such as distance and long-term factors such as the historical acceptance rate of each driver.

Our aim in analyzing the Liftago data set is to understand the contribution of features that determine whether or not a driver is willing to bid for a given passenger's journey. Each feature is detailed in Table 2. We performed a ranking of the features on Liftago's dataset consisting of 253687 passenger requests. The ranking was based on a binary classification model able to predict the probability that a driver will accept and bid for a passenger's request based on the features in Table 2. The classifier is based on the Random Decision Forest Ensemble (RDFE)<sup>1</sup> [31], which provides a means of ranking features by averaging over the expected fraction of samples affected by tree nodes over all 200 trees.

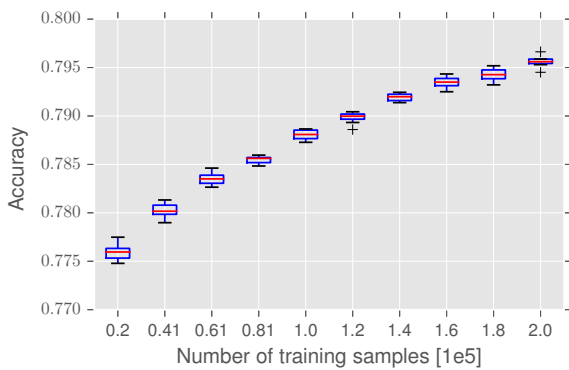
The result of the feature ranking is shown in Fig. 5. Observe that all features have a non-negligible

<sup>1</sup>The implementation is based on the scikit-learn package [30].





**Figure 5.** Feature ranking of Liftago data. The bars show the feature importance of drivers bidding obtained using the random decision forest ensemble algorithm based on 200 trees. The most important features are those approximating ride order distances (Pickup Distance and Ride Distance) and Mean Accept Rate based on per-driver aggregated historical data.



**Figure 6.** Plot of RDFE driver acceptance model accuracy on a test set (up to 20% of the whole dataset).

contribution, which implies that highly efficient market formation algorithms should account for a range of factors, both on a per journey basis and also based on long-term historical data. The analysis also revealed that the most significant factors are the distance from a driver’s initial location to the passengers pick-up location and the mean acceptance rate of the drivers. It is important to note that the pick-up distance is a per journey factor, while the mean acceptance rate is based on long-term driver history.

In practice, the training data depends on how long the provider has been operating. As such, it may not be possible for providers to initially have access to a large dataset. To understand the effect of the size of the training set, the accuracy of the model based on small

**Table 2.** Features derived from the Liftago transactions dataset.

Feature	Description
Pickup Distance	Euclidean distance to pickup
Ride Distance	Euclidian distance from pickup to destination
Pickup Center	Pickup Euclidian distance from Prague center
Ride Center	Pickup Euclidian distance from Prague center
Hour	Time of day
Day	Day of the week
Mean Accept Rate	Driver average accept rate

training sets must be considered. The accuracy is the proportion of correct predictions made by the RDFE model<sup>2</sup>. Fig. 6 is a boxplot of the accuracy for different dataset sizes based on 8 runs of the RDFE with different seeds. Observe that the accuracy is highly dependent on the number of training samples. This suggests that building large training sets is important for market formation in on-demand transport.

#### 4. Conclusions and Future Directions

Data-driven approaches to on-demand transport are now gaining traction as Uber and other providers are adopting market-based mechanisms to allocate and price passengers and drivers. By exploiting historical datasets, it is now possible to improve the efficiency of on-demand transport to ensure it is easier than ever to get from point A to B while maintaining financial sustainability of the provider.

Due to the fundamentally different approach to allocation and pricing, there are new challenges in understanding and improving allocation and pricing in on-demand transport. In this paper, we have provided a framework to classify the different market mechanisms that can be applied and how market formation can be performed. We also have identified key features for market formation based on an analysis of Liftago’s data set, which reveal that both journey-specific and long-term driver behavior should be accounted for. However, there are a number of other aspects that deserve attention.

In particular, the role of subsidies in on-demand transport market mechanisms is not well-understood. As providers need to ensure they have both sufficient supply (drivers) and demand (passengers), it can be necessary for the mechanism to further adjust pricing

<sup>2</sup>The accuracy is a good measure of performance in the case that the classes are well balanced. In the Liftago dataset this is indeed the case with 45% positive and 55% negative.

to ensure that journey offers are accepted by both passengers and drivers.

Selecting and optimizing mechanisms for on-demand transport is challenging due to the huge scale and heterogeneity of the system. For this reason, obtaining large datasets and developing effective simulation tools are a key aspect to improving performance. This is important not just for providers, but also for municipalities and governments to ensure that there is a fair tradeoff between provider revenue, individual driver profits and passenger prices.

Exploiting available data will require effective machine learning techniques at very large scales. Combining these machine learning techniques with economic theory is leading to many new open problems. However, given the importance of transportation, developing efficient data-driven on-demand transport is a valuable challenge.

**Acknowledgements.** This work was partly funded by the Technology Agency of the Czech Republic (grant no. TE01020155). We also thank Liftago for providing the data set and providing permission to publish our analysis.

## References

- [1] R. Balan, N. Khoa, and L. Jiang, "Real-time trip information service for a large taxi fleet," in *Proc. International Conference on Mobile Systems, Applications, and Services*, 2011.
- [2] L. Moreira, J. Gama, M. Ferreira, J. Mendes-Moreira, and J. Damas, "Predicting taxi-passenger demand using streaming data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1393–1402, 2013.
- [3] S. Guo, Y. Liu, K. Xu, and D. Chiu, "Understanding ride-on-demand service: demand and dynamic pricing," in *Proc. IEEE International Conference on Pervasive Computing and Communications Workshops*, 2017.
- [4] S. Guo, Y. Liu, K. Xu, and D. Chiu, "Understanding passenger reaction to dynamic prices in ride-on-demand service," in *Proc. IEEE International Conference on Pervasive Computing and Communications Workshops*, 2017.
- [5] C. Zeng and N. Oren, "Dynamic taxi pricing," in *Proc. European Conference on Artificial Intelligence*, 2014.
- [6] J. Gan and B. An, "Game theoretic considerations for optimizing taxi system efficiency," *IEEE Intelligent Systems*, vol. 32, no. 3, pp. 46–52, 2017.
- [7] M. Egan and M. Jakob, "Market mechanism design for profitable on-demand transport services," *Transportation Research Part B*, vol. 89, pp. 178–195, 2016.
- [8] M. Egan, M. Schaefer, M. Jakob, and N. Oren, "A double auction mechanism for on-demand transport networks," in *Proc. PRIMA2015: Principles and Practice of Multi-Agent Systems*, 2015.
- [9] L. Zhou and H. Xu, "An efficient double auction mechanism for on-demand transport services in cloud-based mobile commerce," in *Proc. IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, 2017.
- [10] J. Zhang, D. Wen, and S. Zeng, "A discounted trade reduction mechanism for dynamic ridesharing pricing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1586–1595, 2015.
- [11] J.-F. Cordeau, "The dial-a-ride problem: models and algorithms," *Annals of Operations Research*, vol. 153, no. 1, pp. 29–46, 2007.
- [12] R. Myerson and M. Satterthwaite, "Efficient mechanisms for bilateral trading," *Journal of Economic Theory*, vol. 29, pp. 265–281, 1983.
- [13] R. Wang, "Auctions versus posted-price selling," *The American Economic Review*, vol. 83, no. 4, pp. 838–851, 1993.
- [14] C. Brooks and E. Durfee, "Congregating and market formation," in *Proc. International Joint Conference on Autonomous Agents and Multiagent Systems*, 2002.
- [15] D. Parkes, "Online mechanisms," in *Algorithmic Mechanism Design*, 2007.
- [16] H. Yang, M. Ye, W. Tang, and S. Wong, "A multiperiod dynamic model of taxi services with endogenous service intensity," *Operations Research*, vol. 53, no. 3, pp. 501–515, 2005.
- [17] C. Borgs, O. Candogan, J. Chayes, I. Lobel, and H. Nazerzadeh, "Optimal multiperiod pricing with service guarantees," *Management Science*, vol. 60, no. 7, pp. 1792–1811, 2014.
- [18] H. Yang, M. Ye, W. Tang, and S. Wong, "Regulating taxi services in the presence of congestion externality," *Transportation Research Part A*, vol. 39, pp. 17–40, 2005.
- [19] J. Gan, B. An, and C. Miao, "Optimizing efficiency of taxi systems: scaling-up and handling arbitrary constraints," in *Proc. International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2015.
- [20] M. Egan and M. Jakob, "A profit-aware negotiation mechanism for on-demand transport services," in *Proc. European Conference on Artificial Intelligence (ECAI)*, 2014.
- [21] R. Cont, "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, vol. 1, pp. 223–236, 2001.
- [22] L. Tesfatsion, "Introduction to the special issue on agent-based computational economics," *Journal of Economic Dynamics and Control*, vol. 25, no. 3-4, pp. 281–292, 2001.
- [23] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2, pp. 235–256, 2002.
- [24] R. Kleinberg and T. Leighton, "The value of knowing a demand curve: bounds on regret for online posted-price auctions," in *IEEE Symposium on Foundations of Computer Science*, 2003.
- [25] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.
- [26] R. McAfee, "Dominant strategy double auction," *Journal of Economic Theory*, vol. 56, no. 2, pp. 434–450, 1992.

- [27] M. Čertický, M. Jakob, R. Píbil, and Z. Moler, “Agent-based simulation testbed for on-demand transport services (demonstration),” in *Proc. International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2014.
- [28] M. Egan, N. Oren, and M. Jakob, “Hybrid mechanisms for on-demand transport,” *In preparation*.
- [29] H. Yang, C. Fung, K. Wong, and S. Wong, “Nonlinear pricing of taxi services,” *Transportation Research Part A*, vol. 44, no. 5, pp. 337–348, 2010.
- [30] F. Pedregosa *et al.*, “Scikit-learn: machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [31] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.