SHARED-VEHICLE MOBILITY-ON-DEMAND SYSTEMS:
A FLEET OPERATOR'S GUIDE TO REBALANCING EMPTY VEHICLES

Kevin Spieser (Corresponding author)
Laboratory for Information and Decision Systems (LIDS)
Massachusetts Institute of Technology
32 Vassar Street, Cambridge MA 02142
kspieser@mit.edu

Samitha Samaranayake
Laboratory for Information and Decision Systems (LIDS)
Massachusetts Institute of Technology
32 Vassar Street, Cambridge MA 02142
samitha@mit.edu

Wolfgang Gruel
car2go GmbH
Fasanenweg 15-17, 70771 Leinfelden, Germany
wolfgang.gruel@daimler.com

Emilio Frazzoli
Laboratory for Information and Decision Systems (LIDS)
Massachusetts Institute of Technology
32 Vassar Street, Cambridge MA 02142
frazzoli@mit.edu

5465 words + 4 figures

October 24, 2015

ABSTRACT

We consider the operation of automated mobility-on-demand systems, whereby users share access to a fleet of self-driving vehicles. In these systems, rebalancing, the process by which the supply of empty vehicles is periodically realigned with the demand for transport, is carried out by a fleet operator. Where much of the rebalancing literature skews to the theoretical or simulation-based, we considering shared mobility systems from the perspective of the fleet operator. We test, via simulations, how key performance metrics vary as a function of fleet size and rebalancing policy using rental data from car2go, a free-floating carsharing service operating in markets across Europe and North America. Results reveal that rebalancing can dramatically reduce the number of customer walk aways, even for relatively small fleet sizes. A framework is provided to assess what fleet size is a appropriate for a city factoring in the cost of vehicles, customer walk aways, and the added expense of moving empty vehicles.

INTRODUCTION

In an increasingly urbanized world, ensuring sustainable access to mobility is a serious issue. Road networks and the supporting infrastructure are operating at or near capacity, and the demand for transport continues to rise. Attempts to add parking spaces and expand roadways raise environmental concerns, threaten the livability of cities, and, in many cases, are prohibitively expensive. Fortunately, the emergence of shared-economy markets and ongoing advancements in autonomous vehicle technology may provide a novel option to alleviate the ensuing increase in the demand for personal mobility.

In one-way, mobility-on-demand (MoD) systems, users access mobility by sharing a fleet of vehicles. Using their cellphone, a user can reserve and pickup a nearby vehicle, drive to his or her destination, and then drop off the vehicle. The pickup and dropoff locations need not be the same. Through this arrangement, MoD systems offer the convenience of a private car, without the associated hassles, e.g., maintenance and fueling, and financial investment, e.g., a large fixed cost and ongoing insurance payments (*1, 2*). Existing one-way carsharing services such as car2go and DriveNow (in contrast to traditional car rental services) provide their members with access to a free floating vehicles on-demand (*3*). These services offer an alternative to private automobiles. They not only reduce the total number of vehicles required, but can also contribute to reducing energy consumption and lower emissions in the long run (*4*).

Despite their many perks, MoD system are plagued by an endemic inefficiency (*5*). In most cities, the supply of empty vehicles and the demand for transport are not aligned on the timescales of interest. Left unchecked, carsharing systems experience vehicle surpluses at some locations and shortages at others. A similar imbalance emerges in bike sharing (*6*). However, a bike's small weight and compact design makes it possible to shuffle dozens of bikes with a single utility van. Owing to a car's large size, rebalancing empty vehicles through human labour is an expensive and inefficient proposition. Fortunately, recent advancements in autonomous vehicle technology provide an opportunity to remove this bottleneck (*7*).

Major car manufacturers, research universities, and even software companies have now demonstrated vehicles capable of performing almost all driving-related tasks autonomously (*8*). As of June 2015, autonomous cars from Google have logged more than 1 million miles of unassisted driving (*9*). Moreover, longstanding legal barriers that have limited the impact of self-driving vehicles are beginning to fall. Although most jurisdictions currently require autonomous vehicles to have a safety-driver onboard to intervene in the event of an emergency, impending legislation is likely to relax this requirement (*10*). For Autonomous MoD (AMoD) systems, this freedom would allow vehicles to *rebalance* themselves to more effectively serve the travel demand by facilitating smaller fleet sizes, shortening expected wait times, etc.

Despite the potential of automated MoD systems, fielding fleets of autonomous vehicles raises important logistic questions. For example, how large should autonomous vehicle fleets be and which rebalancing policies should be used? How can future travel demand be incorporated into rebalancing plans and what happens when the future demand is uncertain? To complicate matters, the answer to these questions vary from one party to another. Customers want assurances they can quickly and reliably access a vehicle, attributes that, in the extreme, favor large fleets and frequent rebalancing. Fleet owners are interested in the financial viability of their systems; accordingly, they may be more sensitive to the added costs of such policies. In this paper, we task the fleet operator with striking a favorable balance between consumer satisfaction and corporate objectives[1]. We consider a number of rebalancing policies, paying specific attention to the practical design decisions and tradeoffs facing the fleet operator. To test these algorithms on a meaningful scale, we simulate their ability to serve hypothetical AMoD installations in markets with different characteristics throughout North America. In each case, the travel demand is based on records of actual trips taken by users of car2go fleets already operating in the respective cities.

The remainder of this paper is organized as follows. Section 3 reviews a collection of the relevant literature. Section 4 describes the problem formulation. Section 5 describes a pre-existing general approach to rebalancing and how it can be tailored to respond to (i) demands currently in the system and (ii) demands forecasted to arrive in the near future. To put the aforementioned rebalancing policies to the test, we simulate their performance using actual travel patterns. To this end, Section 6 describes the data sources used and Section 7 discusses the associated simulation results. Finally, Section 8 closes by highlighting key ideas and outlining future work initiatives.

RELATED WORK

Carsharing systems first appeared in Europe in the mid 1900s (*11*). They have gained prominence in North American within the last two decades (*12*), with one-way rental services beginning to gain traction only in the last five to seven years. Considerable work has been done to understand the market for carsharing services in different cities (*13*), with efforts made to relate consumer interest in these services to demographics (*14*), urban geography (*15, 16*), and the quality of service offered (*17*). An excellent survey of these and related issues may be found in (18).

Recognizing advances in autonomous vehicle technology, car-sharing researchers have begun to build the functionality associated with self-driving cars into their shared mobility models. The need to relocate empty vehicles in car sharing systems has gained much attention in the last couple of years. Some authors propose user-based solutions for relocating vehicles and make use of economic incentives while other strategies are operator-based (*19, 20*). Approaches range from agent-based models (*21*) to mathematical models (*22*). However, these strategies rely on human labor to rebalance, which is costly and impractical. For this reason, researchers have been eager to engage the use of self-driving cars, for which the vehicles are capable of repositioning themselves. The bulk of this work obeys a natural dichotomy: i) theoretical analysis of abstract models and ii) simulation-based analysis of more realistic systems. The theoretical work contributes to an understanding of the fundamental limitations and tradeoffs present in AMoD systems. It also informs the development of principled control strategies with quantifiable assurances. On the other hand, simulation-based efforts provide a testbed to evaluate the heuristic strategies that often drive practical implementations. In this work, we test rebalancing policies that have been designed to be optimal, given certain assumptions, on real-world test data.

In (*23*), the authors show that rebalancing in most AMoD systems is necessary to prevent the unbounded build up of customers for an otherwise stabilizable arrival process. Working from a fluid model, they provide a stabilizing rebalancing policy that minimizes the number of empty vehicle miles traveled under static conditions. A relaxation is presented that accounts for actual systems whereby customers and vehicles are not divisible. (*24*) provides a rigorous and stochastic approach to deal with a system of self-driving vehicles in which customers are impatient and leave the system if they arrive at a station with no vehicles. They propose a queueing-theoretical approach that models a network of autonomous vehicles within a Jackson network, and give a linear program that equalizes the fleet availability across all stations.

[1]Customer satisfaction is in fact a corporate objective too, because low customer satisfaction will lead to reduced ridership and reduced revenue.

Their algorithms are tested using taxi data from New York City and demonstrate that the current taxi demand in New York City could be served with 40% fewer vehicles. In (*25*), the authors consider the thought experiment of replacing all modes of land-based transit in Singapore with unit capacity self-driving cars. A minimum bound on the fleet size required for stability is provided as well as subsequent estimates to ensure an adequate level of service. A similar project was undertaken in (*26*) that focused on three venues within the United States.

Because AMoD systems involve picking up and and dropping off passengers, they are intrinsically linked to many of the same operational issues explored by the vehicle routing community. The capacitated vehicle routing problem concerns finding a schedule for a vehicle to follow to ensure demands are serviced efficiently (*27, 28, 29, 30*). Vehicle routing problems come in a variety of forms, depending on the number of depots from which vehicles are dispatched, the capacity and capability of vehicles, the communication and information constraints of the scenario, and what transactions constitute serving a demand, see (*31, 32, 33, 34*) for a selection of the many variations considered.

In contrast to the aforementioned approaches, the performance of rebalancing policies can also be evaluated through event-driven software testbeds. In (*26*), the authors investigated different aspects of performance for hypothetical shared-vehicle systems situated in three locations within the United States. Many of the policy decisions are driven by heuristics; for example, customers that appear in the system are matched to the closest vehicle. They conclude that current taxi demand in Manhattan could be satisfied by 9,000 autonomous taxis, i.e. roughly 70% of the current fleet of about 13,000 taxis. At the same time, average wait-times could be pushed to less than one minute. Assuming that the cost of autonomy is only $2,500 per vehicle, they optimistically suggest that the cost of an average taxi trip will drop from $7.80 to $1 due to higher utilization and reduced labor cost.

Fagnant and Kockelman (*7*) performed an agent-based simulation for shared vehicle systems. They generated travel demand artificially for a grid-based urban area using national travel statistics for the US and accounting for traffic congestion. They applied different rebalancing strategies in their simulation and tried to minimize the wait time of the average traveller. Prior knowledge was used to preemptively position vehicles in advance of people appearing. The authors propose that one shared autonomous vehicle can replace roughly eleven conventional vehicles with an increase in total vehicle mileage of 10%, as compared to non-shared-autonomous trips. Building on this work, in (*35*), the same authors apply their methods to the transportation network of Austin, Texas, accounting for traveling speeds throughout the network during different times of the day. Using a synthetically generated travel demand, they investigate the potential implications of shared autonomous fleet operations at a market penetration rate of 1.3%, reporting comparable vehicle reduction ratios and increases in total vehicle mileage traveled. Additional simulation-based approaches have been reported in (*36, 37, 38, 39, 40*).

PROBLEM FORMULATION

This section provides the notation and terminology used to describe AMoD systems. To streamline the presentation as well as reinforce the interconnected nature of system components, we adopt a modular approach, recognizing the the AMoD system is the collection of all components.

The hub network

Consistent with (*23*), we a consider a spatially embedded, hub-based network model $H = (V_H, E_H)$ where $V_H = \{1, \ldots, N\}$ is a finite set of N hubs and $x_i \in \mathbb{R}^2$ is the location of hub i in the plane. $E_H \subseteq V^2$ is a set of edges such that $ij \in E$ if and only if there is a direct link between hubs i and j. The travel time along ij is given by $T_{ij} > 0$. Although this model has limitations, e.g., it can not perfectly capture a free floating carsharing system, its discrete, graph-based nature facilitates a level of analysis not otherwise possible. Moreover, for sufficiently large N, the average distance between any point in the city and the nearest hub becomes small, and the consequences of using a hub-based model less significant.

The demand model

Each customer that enters the system represents a demand for transport that may be described by a triple (i, j, t), where $i, j \in V$, $i \neq j$, are the origin and destination of the demand, and t is its time of arrival. The rate at which customers arrive at hub i at time t is $\lambda_i(t)$. At time t, the rate at which customers arrive at i seeking to travel to j is $\lambda_{ij}(t)$, with

$$\lambda_i(t) = \sum_j \lambda_{ij}(t), \; \forall ij \in E, \; t \geq 0. \tag{1}$$

When a demand arrives at hub i, it immediately enters a queue, with $q_i(t)$ denoting the number of demands waiting at hub i at time t. At each hub, queued demands are served in a first-in, first-out order.

In some cases, we are interested in how a rebalancing policy performs when the future demand for transportation is unknown. In these cases, we follow the standard procedure of using separate training and test data sets, with one day of data used for training and a separate (but representative day) used for testing. In other words, a training set is used to make predictions about where demands are likely to appear in the near future, but the simulation is performed using the test dataset. The similarity between the two data sets and the nature of the rebalancing policy dictate how well these policies perform. As a side note, policies that have perfect information about future demand perform no worse than those that do not, because the former can always choose to ignore this information.

Rebalancing tasks

To facilitate rebalancing, each hub i maintains a queue of rebalancing tasks. Each task at i is a pair of the form $(j, z) \in V \times \mathbb{Z}$, specifying that z empty vehicles are to be sent from i to j. The task queue at i may be populated following calls to a rebalancing routine, typically executed periodically every T_P time units. Rebalancing tasks are served in a first-in, first-out order. Empty vehicles at i are first allocated to service demands, with any remaining vehicles used to fulfill rebalancing tasks.

The vehicle fleet

Demands for transport between hubs are served by a fleet of n self-driving vehicles. Vehicles not traveling between hubs are parked at a hub. Let $v_i(t)$ denote the number of vehicle parked at i at time t. Similarly, let $v_{ij}(t)$ denote the number of vehicles en route from i to j at t. Conservation of vehicles requires

$$n = \sum_{i \in V} v_i(t) + \sum_{ij \in E} v_{ij}(t). \tag{2}$$

There are two reasons for a vehicle to travel between hubs i and j: (i) to transport a single demand from i to j, or (ii) to contribute to the rebalancing effort by fulfilling a rebalancing task.

Demand model

It is assumed that the future demand $\lambda_{ij}(t)$ is not known exactly, but that can be approximated. Estimating $\lambda_{ij}(t)$ allows the fleet operator to dispatch vehicles to hubs in preparation for future demands. This scenario is a more realistic representation, because while the fleet operator may have access to historical travel patterns, he does not know all trip specifics in advance. The process of estimating the future demand will be eloborated on in Section 7.

AMoD performance metrics

The customer experience provided by an AMoD system is heavily dependent on the ability to serve demands in a timely manner. Customers that are not quickly matched with vehicles are likely to seek an alternate means of transport. We refer to this event as a *walk-away* and assume each customer has a waiting budget of t_{max}, after which time they walk away and leave the system. The total number of walk-aways over the period of interest is C_{wa}.

Rebalancing trips

A vehicle trip that fulfills a rebalancing task is referred to as a rebalancing trip. For the duration of that trip, the associated vehicle is referred to as a rebalancing vehicle. Following this convention, an empty vehicle en route to pick up a waiting customer and an empty vehicle headed to a hub in preparation for a future customer each constitute a rebalancing trip.

One can conceive of any number of ways to rebalance vehicles. In this paper, we will focus on algorithms that are optimal according to some cost metric. However, there are a number of tuning parameters that do not fit naturally into an optimization framework and must be selected through other means. Throughout the paper, we make a special point of discussing how a fleet operator might perform this task.

REBALANCING A VEHICLE FLEET

We begin by summarizing the rebalancing framework originally reported in (*23*). Later, new functionality will be added by expanding on this formulation. For the time being, we assume an omniscient fleet operator with perfect knowledge of $\lambda_{ij}(t)$, $q_i(t)$, $v_i(t)$, and $v_{ij}(t)$.

For each pair of distinct hubs, $ij \in E$, let n_{ij} represent the number of empty vehicles the fleet operator will send from i to j during rebalancing. The purpose of the rebalancing algorithm is to determine optimal values for n_{ij}. Let n_i^{exc} denote the number of *excess* vehicles at i. Similarly, let n_i^{des} denote the number of *desired* vehicles at i following rebalancing. The following linear program minimizes the amount of work required to realize n_i^{des} at each hub.

$$\min_{n_{ij}} \sum_{ij \in E} T_{ij} n_{ij} \tag{3}$$

$$\text{s.t.} \sum_j n_{ji} - \sum_j n_{ij} \geq n_i^{des} - n_i^{exc}, \; i \in V \tag{4}$$

$$n_{ij} \geq 0, \; ij \in E. \tag{5}$$

In words, (3) represents the total amount of time traveled by empty vehicles to realize the desired fleet distribution. (4) says that after all rebalancing trips are accounted for, each i has n_i^{des} vehicles. (3) In (4), there is some freedom in selecting n_i^{exc}. For example,

$$n_i^{exc}(t) = v_i(t) + \sum_j v_{ji}(t), \tag{6}$$

includes vehicles inbound to i in the calculation. Alternatively, taking $n_i^{exc}(t) = v_i(t)$ permits the fleet operator to shuffle only those vehicles currently residing at hubs. Throughout, we define

$$m = \sum_{i \in V} n_i^{exc}, \tag{7}$$

to be the total number of excess vehicles, recognizing there are multiple ways to define n_i^{exc}. It is important to note that in this formulation, empty vehicles are not rerouted while traveling along links. Rather, it is only after reaching a hub, that a vehicle can change course.

Relative to n_i^{exc}, there is considerably more flexibility in choosing n_i^{des}, and selection of this quantity determines the functionality of the algorithm. However, to ensure (3)–(5) is feasible, we do require that

$$\sum_i n_i^{des} \leq m, \tag{8}$$

i.e., the fleet operator does not attempt to reposition more vehicles than are available. Having selected n_i^{des}, the fleet operator performs rebalancing every T_P time units by solving (3)–(5), determine the optimal n_{ij}, and placing the associated rebalancing tasks in the appropriate queues at each hub. We remark that in the current framework, we do not keep track of how long demands have been waiting at stations when dispatching empty vehicles. Consequently, it is possible that demands that warranted a rebalancing trip will walk away before the associated vehicle arrives. Addressing this limitation is a item of future work.

How the fleet operator chooses T_P can have a significant bearing on performance. Choosing T_P too large limits the ability to redistribute the fleet. Conversely, choosing T_P too small will ignore natural overlaps between pickup and dropoff distributions, spawning more rebalancing trips than necessary.

The rest of this section describes two rebalancing approaches: one that responds to the current state of the system and one that uses knowledge of future demands to preposition vehicles.

Feedback rebalancing

Feedback rebalancing is succinctly summarized as follows: every T_P, (i) take stock of the outstanding demands in the system and (ii) send empty vehicles to the demands along the most efficient route. The approach is called *feedback rebalancing*, because empty vehicles are routed in response to the current state of the system being *fed back* to the fleet operator. Recall that $q_i(t)$ is the number of outstanding demands at hub i at time t. Let

$$Q(t) = \sum_{i \in V} q_i(t) \tag{9}$$

denote the total number of demands in the system at time t. If $Q_i(t) \leq m$, all demands can be matched with an empty vehicle. However, if $Q_i(t) > m$, there are too few vehicles to match all demands. In this case, there are multiple way to proceed. Here, we opt to allot excess vehicles to hub i in proportionate to $q_i(t)$, i.e.,

$$n_i^{des}(t) = \begin{cases} q_i(t) & , \; Q_i(t) \leq m \\ \dfrac{q_i(t)}{Q_i(t)} \cdot m & , \; \text{otherwise.} \end{cases} \tag{10}$$

The attributes of feedback rebalancing lie in its simplicity: it only requires knowledge of the system state and avoids extraneous rebalancing by routing empty vehicles only when there is an immediate need. However, it fails to leverage knowledge the fleet operator may have regarding the upcoming travel demand.

Feedback + proportional predictive rebalancing

In the event there are more excess vehicles than demands, i.e., $Q(t) < m$, let $m' = m - Q(t)$. That is, m' represents the number of excess vehicles still available to the fleet oeprator after allocating $Q(t)$ vehicles for feedback rebalancing. Assume λ_{ij} is perfectly known over the time interval $[t, t+\tau]$ for $\tau \geq 0$. We refer to parameter τ as the look-ahead window. Let $\lambda_i(t,\tau)$ denote the total number of arrivals at i over $[t, t+\tau]$. Feedback plus proportional rebalancing distributes the m' excess vehicles, not matched during feedback, to hubs in proportion to $\lambda_i[t,\tau]$, i.e.,

$$n_i^{des} = \begin{cases} q_i(t) + \dfrac{\lambda_i(t,\tau)}{\sum\limits_{j \in V} \lambda_j(t,\tau)} \cdot m' & \text{if } Q(t) < m \\ \dfrac{q_i(t)}{Q(t)} \cdot m & \text{otherwise.} \end{cases} \tag{11}$$

This approach is predictive in that in preemptively moves excess vehicles in preparation for demands arriving in $[t, t+\tau]$. Accordingly, care must be taken when selecting τ. Choose τ too small, and rebalancing vehicles pulling into hubs will not be representative of the awaiting demands. Choose τ too large, and the vehicle distribution is overly influenced by demands yet to appear. As a rule of thumb, we advocate selecting τ to be on the same order as the average trip length in $[t, t+\tau]$.

Feedback + predictive rebalancing augments the pragmatic sensibility of (10) by capitalizing on access to λ_{ij}. However, what happens when λ_{ij} is uncertain? Here, predictive rebalancing could send at least some vehicles to the *wrong* hubs.

DESCRIPTION OF TRANSPORTATION DATASETS FROM CITIES

To test the rebalancing algorithms described in Section 5, we used a day of recorded trip data from car2go, one of the largest one-way carsharing service that operates in a number of cities across North America. The carsharing system is free-floating, meaning customers may drop off vehicles anywhere within a prescribed zone. It is reasonable to assume that the first autonomous vehicles for public use are likely to be expensive and tightly regulated. These reasons suggest autonomous vehicles may first appear in the form of AMoD systems. Given their novelty, initial adopters are likely to be people seeking alternatives to traditional mobility models. In many ways, current car2go subscribers fit this demographic, suggesting their travel patterns may be representative of the initial demand experienced by the first AMoD installations, in the respective cities.

The rebalancing algorithms in Sections 5 are predicated on a hub-based network $H = (V_H, E_H)$ and demand model λ_{ij}. The following subsections describe how suitable H and λ_{ij} were determined from the data.

Because AMoD systems are also one-way and free-floating, the recorded rentals may be easily transcribed into a demand model suitable for our purposes. For our simulations, we assume we install an AMoD system in place of the one-way carsharing system, and that the demand model carries over exactly. That is, we do not attempt to model any effects associated with induced demand the new system may bring about.

Creating H

Let the k-th rental during a day be described by a quadruple $(x_k^{pu}, x_k^{do}, t_k^b, t_k^e)$, where x_k^{pu} and x_k^{do} are the points at which the vehicle was picked up and dropped off, respectively. Similarly, t_k^b and t_k^e are the times at which the vehicle rental began and ended, respectively.

A hub-based network was developed by taking the set of all pickup and dropoff points, $P = \{(x_i^{pu}, x_i^{do})\}_i$ for each day of the week (i.e. the hub network for Wednesdays is created using the data from previous Wednesdays), and using a k-means clustering algorithm to form N clusters. V_H was then formed by assigning a hub to the centroid of each cluster.

To determine E_H, we formed the Delaunay graph D of V_H, where

$$V_D = V_H \tag{12}$$

$$E_D = \{ij \in V_D^2 | \exists x \in \mathbb{R}^2 | i, j \in \operatorname*{arg\,min}_{k \in V_D} \{d(k,x)\}\}. \tag{13}$$

H is formed by taking $E_H = E_D$, and thus retaining the notion of hub proximity encoded in D. Note that H is not complete, i.e., $\exists i, j \in V_H | ij \notin E_H$, but is connected, i.e., there is a path between any i and j in H.

Creating λ_{ij}

With access to H, $\lambda_{ij}(t)$ is constructed by mapping each rental record $(x_k^{pu}, x_k^{do}, t_k^b, t_k^e)$ to a hub-based demand (i, j, t_k^b, t_k^e) where i and $j \in V_H$ are the hubs closest to x_k^{pu} and x_k^{do}, respectively. As mentioned, H and $\lambda_{ij}(t)$ is an abstraction of the actual demand, but one that can be made arbitrarily close for sufficiently large N. In our simulations, we use $N = 40$ to provide an adequate spatial resolution.

SIMULATION RESULTS

To gauge the performance of the various rebalancing policies, we tested them using the demand patterns from various car2go markets in North America. For example, Figure 1(a)–1(c) depict the number of walk-aways, the utilization rate, and the number of rebalancing miles traveled for various fleet sizes and for various rebalancing policies in Seattle. The remainder of this section is devoted to commenting on the most pronounced features of these plots and describing the mechanisms that generate them.

To model customer impatience, it is assumed that a demand (customer) walks away once they have waited at least six minutes without receiving service. This number is supported by Uber (*41*). Naturally, as fleet size increases, there are more vehicles scattered throughout the network and the number of walk aways decreases. However, the goal of all rebalancing schemes considered is to match the supply of vacant vehicles with the pending and upcoming demand for transport. We consider three rebalancing strategies: a) Feedback rebalancing, b) Feedback + proportional rebalancing with future demand based on historical data (practical case), c) Feedback + proportional rebalancing with future demand based on perfect predictive capability (best case)

To first summarize the results, we observe that rebalancing allows a fleet operator to significantly reduce the number of vehicles that are needed to serve a fixed demand with the same quality of service.

However, this reduction in fleet size comes with a trade-off. At very small fleet sizes, the vehicles need to rebalance often. As a result, the number of vehicle miles travelled (due to rebalancing) increases compared to the case with no rebalancing.

As Figure 1(a) indicates, any of these rebalancing policies dramatically reduce the number of walk aways for small to medium fleet sizes. It is only for larger fleet sizes that rebalancing efforts provide only a minimal reduction in the number of walk aways relative to a no rebalancing policy. Unsurprisingly, the predicative rebalancing strategies that use knowledge or estimates of future demand do the best job of mitigating walk aways, with the fewest walk aways reported for the preemptive rebalancing strategy that has perfect knowledge of the future travel demand. By and large, these trends extend across all cities, as seen in Figures 2(a)- 4(a).

The next metric of interest, utilization rate, refers to the fraction of time that vehicles are transporting customers over the course a day. Figure 1(b) depicts the utilization rate in Seattle for various fleet sizes and policies and is representative of the analogous plots from other cities. In the case of no rebalancing, when the fleet size is very low, the utilization rate is, as expected, relatively high because the ratio of demands to vehicles is fairly high. However, as the fleet size increases (and quality of service increases), the inclusion of extra vehicles reduces the ratio of vehicle demand to vehicles available and the utilization rate drops. Furthermore, more vehicles become stranded at unpopular stations, which also drives the utilization rate lower. When rebalancing is used on smaller fleet sizes, vehicles are quickly repurposed and manage to stay busy for a greater fraction of the day, leading to higher utilizations rates. As fleet size increases, the utilization rate drops for the same reasons alluded in the no rebalancing case.

The third quantity we consider is the empty vehicle miles traveled (VMT), which are an indication of how intensely the fleet is redistributed. Figure 1(c) depicts the total empty vehicle mileage driven over the course of the day in Seattle. By definition, the empty VMT is zero in the no rebalancing case. For feedback rebalancing, empty VMT tails off as fleet size increases. This is a reflection of the fact that with a greater abundance of vehicles in the system, there is less need to dispatch rebalancing vehicles, and if a dispatch is required, these trips are generally shorter. The reasoning is more involved in the case of the preemptive rebalancing strategies. Focusing first on the perfect information case, for very small fleet sizes, there are more demands in the upcoming look ahead window than there are vehicles available. Consequently, it is not possible to dispatch a vehicle for each upcoming trip at the time of rebalancing. In this regime, as fleet size is increased, more and more vehicles are dispatched, producing the upward trend in the empty VMT curve. However, once the fleet size is on par with the number of demands appearing in the look ahead window, there is a surplus of vehicles throughout the network. Because vehicles are dispatched optimally, i.e., to minimize the empty VMT associated with fleet redistribution, the empty VMT begins to saturate. Continuing, if the fleet is further increased the same effects alluded to in the no rebalancing case begin to take hold. For these reasons, the empty VMT curve for preemptive rebalancing policies typically increases, saturates and then begins to decline steadily as a function of fleet size.

A similar trend occurs for preemptive rebalancing with imperfect information. However, because rebalancing vehicles may be sent to stations where no demands materialize, there is the potential for extra empty vehicle mileage to accrue. Again, for higher fleet sizes, this effect is drowned out by the superior coverage a greater number of vehicles afford and the empty VMT decreases.

Figures 1(a)-1(c), referencing Seattle and provided on the following page, and the associated plots pertaining to other cities on the pages that follow, describe operational metrics as a function of fleet size in isolation. However, a fleet operator will need to consider the combined effects of these metrics. To this end, Figures 1(d)–4(d) describe the tradeoff between fleet size, walkaways and the empty VMT (effort of rebalancing). Understanding this tradeoff can help operators make decisions such as what fleet size and rebalancing policy to utilize in a specific city. In the discussion that follows, we focus on the methodology more so than the actual financial figures and many of the scales have been intentionally presented to display unitless as opposed to dollar values due to the sensitive nature of the underlying data. We only wish to convey the general characteristics of the tradeoffs, as also the appropriate dollar values to use for autonomous vehicles are somewhat subjective and continually evolving[2]. A specific fleet operator can use this framework and tailor the numbers to match the specifics of their operation.

To this end, for each rebalancing policy and fleet size, we define the *operational cost* to be a metric that is comprised of three factors: the cost of purchasing and maintaining the fleet, the cost associated with the empty miles driven by rebalancing vehicles, and the cost associated with customers that walk away without receiving service. The cost to operate the fleet is normalized in the figures to emphasize the shapes of the curves instead of the actual values, which will differ between cities, vehicle fleet sizes, operators, and business models. For example, upon supplying the appropriate cost values, one could generate a plot similar to Figure 1(d). Here, one can contrast the overall expense for various fleet sizes and across different rebalancing policies. Collections of figures analogous to Figures 1(a)–1(d) for different cities are provided in Figures 2(a)–2(d) through 4(a)–4(d). The plots provide an initial indication of the fleet size that is required to achieve a given quality of service with a certain rebalancing policy. The demand profiles in each case are a complex function of the geography, demographics, and social character of the respective markets, and, as a first go, the plots provide an initial insight into how fleets may be sized and operated from a planning level.

CONCLUSIONS AND FUTURE DIRECTIONS

The objective of this paper was to conduct a practical assessment of the logistic benefits associated with various rebalancing schemes across a spectrum of mobility markets. Simulation results of various rebalancing schemes revealed how the number of customer walk aways, the utilization rate, and the number of empty vehicle miles varies as a function of fleet size. The notion of operational cost was presented as a way to fuse these metrics into a unified framework that allows the fleet operator to make decisions about which rebalancing policy and what fleet size is most appropriate for a given market.

Given the preliminary nature of the work, there are a number of potentially fruitful research directions. For example, the true cost of walk aways has been captured at only the most rudimentary level. In reality, customers that walk away will be frustrated and less likely to use the AMoD system in the future. On a related note, the demand models used did not account for trips that never began because potential users could not find a vehicle nearby. Conversely, if the AMoD system is able to provide a high quality of service, it stands to reason that demand will increase. It would be useful, especially for the perspective of gauging the feasibility of AMoD systems in a market, to perform an analysis that captures these effects.

The rebalancing policies were predicted on a network of hubs design to approximate an underlying network. However, our model does not account for the actual road network and any peculiarities (e.g., one-way streets) it may have. Nor does it account for congestive effects and the associated variations in travel times. Incorporating these elements would add credence to the analysis.

[2] We used publicly available data [add citations here] to estimate our cost parameters such as fixed vehicle costs, cost per mile of usage etc.

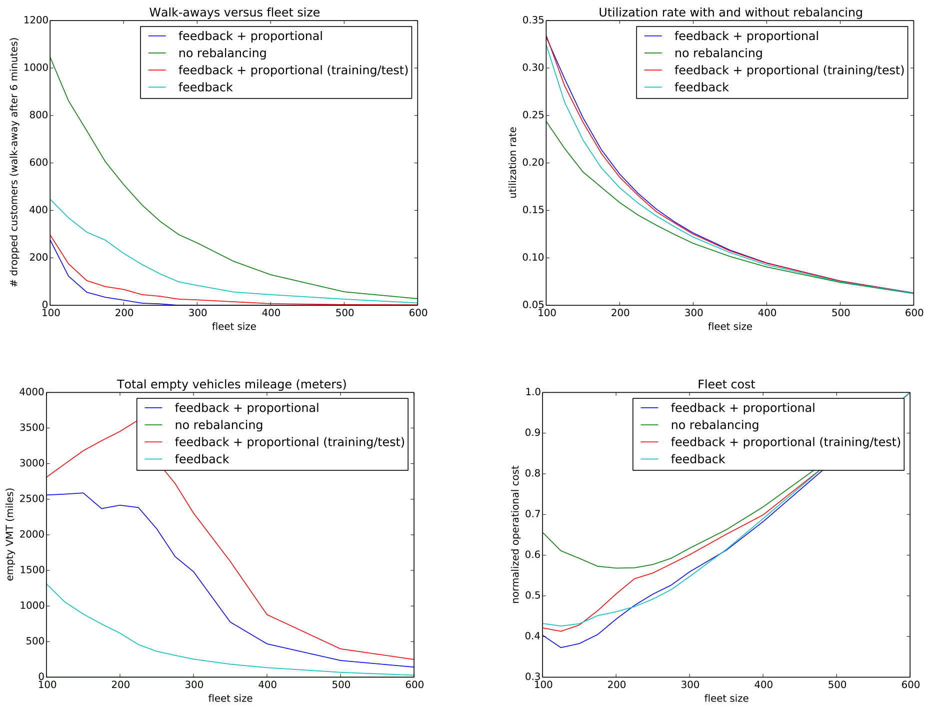

Figure 1: Performance and costs plots versus fleet size for a hypothetical AMoD installation in Seattle.

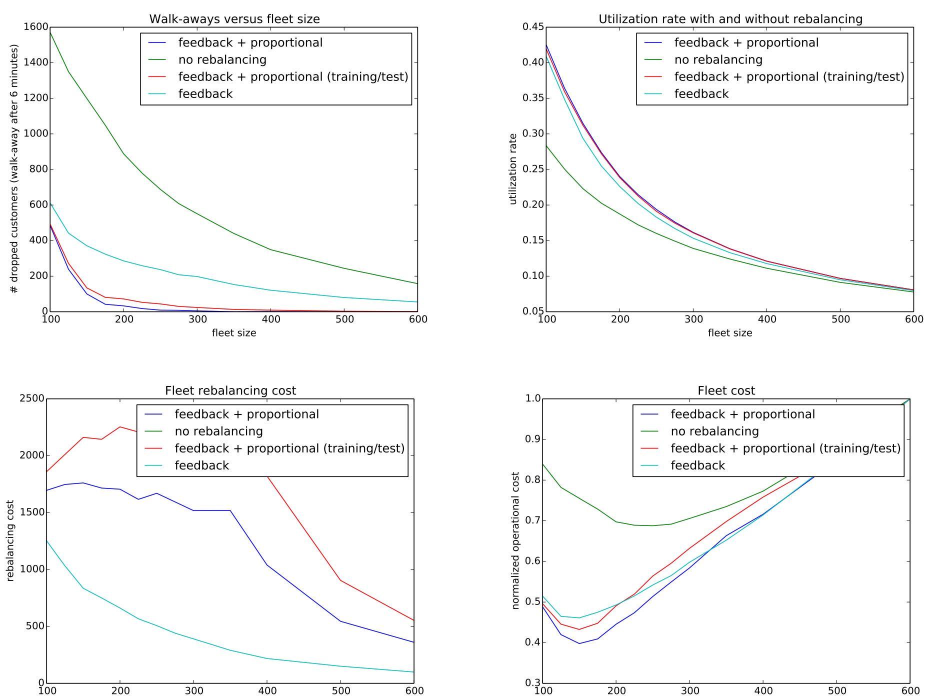

Figure 2: Performance and costs plots versus fleet size for a hypothetical AMoD installation in Calgary.

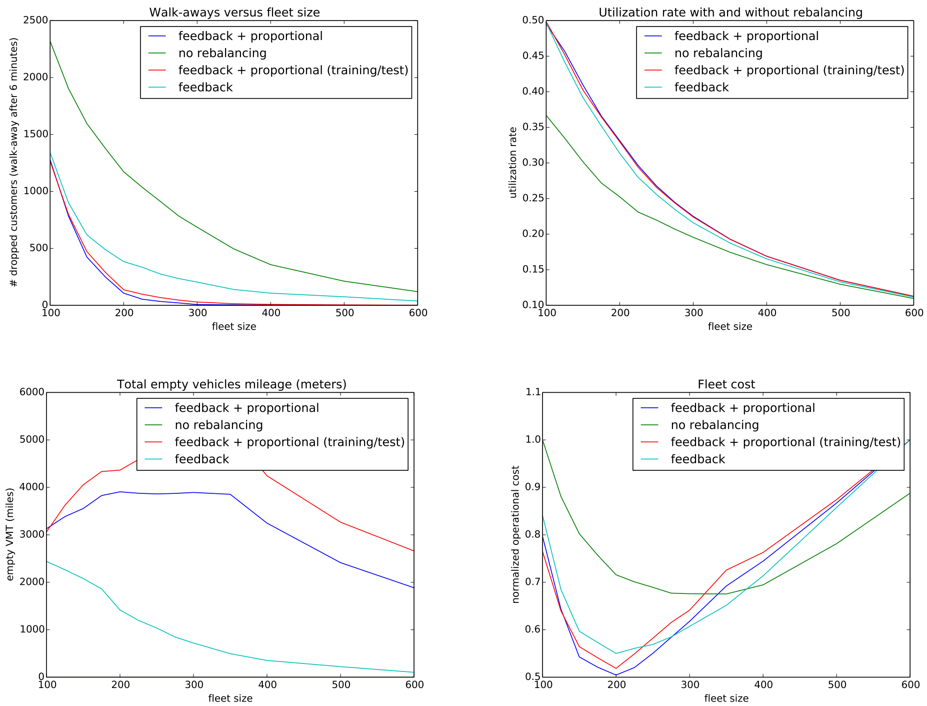

Figure 3: Performance and costs plots versus fleet size for a hypothetical AMoD installation in Vancouver.

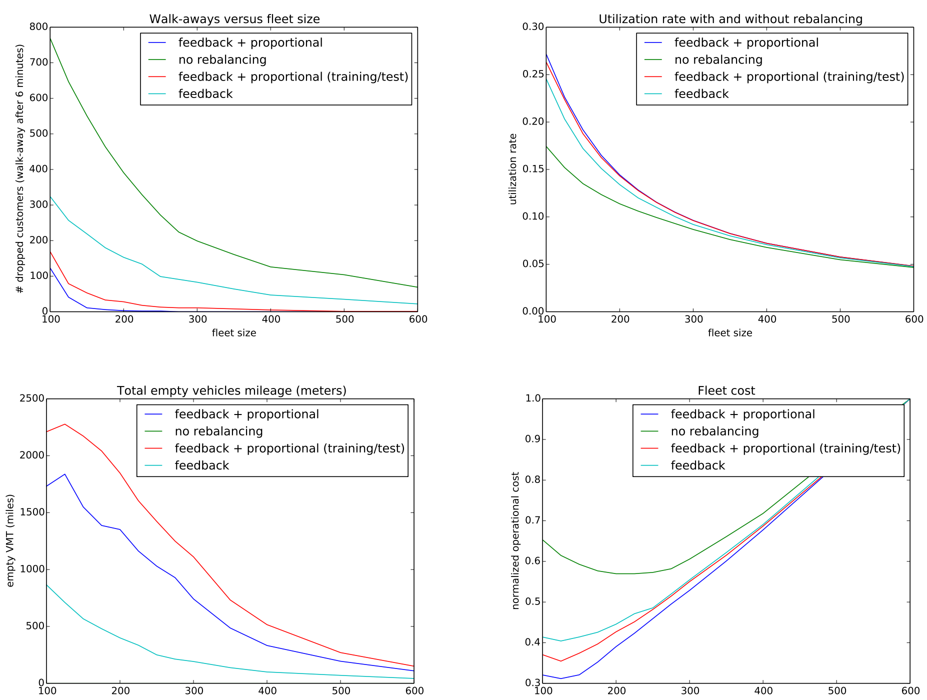

Figure 4: Performance and costs plots versus fleet size for a hypothetical AMoD installation in Washington.

The transportation character of a city is determined in large part by its mass transit infrastructure, e.g., subway, light rail, and bus systems. AMoD systems have the potential to serve the first and last mile problems by dropping off and picking up customers at major hubs, e.g., subway stations and airports. In this framework, mass transit is used to move significant numbers of people over relatively large distances, with autonomous vehicles serving as an enabling technology to provide mobility on a personalized scale. It remains to explore how such a dynamic may impact the fleet size versus performance plots presented in this paper.

REFERENCES

[1] Mitchell, W., C. Borroni-Bird, and L. Burns, *Reinventing the Automobile: personal urban mobility for the 21st century*. MIT Press, 2010.

[2] Shaheen, S. and M. Christensen, The True Future of Mobility had Two Big Barriers to Entry. *citylab.com*, 2014.

[3] Shaheen S., N. C. and H. Micheaux, One-way carsharing's evolution and operator perspectives from the Americas. *Transportation*, , No. 42(3), May 2015, pp. 519–536.

[4] Shaheen, S. A. and A. Cohen, Carsharing and Personal Vehicle Services: Worldwide Market Developments and Emerging Trends. *International Journal of Sustainable Transportation*, 2013.

[5] Weikl, S. and K. Bogenberger, Relocation Strategies and Algorithms for free-floating Car Sharing Systems. *15th International IEEE Conference on Intelligent Transportation Systems*, September 2012.

[6] hemla, F. M. R. W. C., D., Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization*, , No. 10(2), 2013, pp. 718–733.

[7] Fagnant, D. J. and K. M. Kockelman, The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies*, Vol. 40, 2014, pp. 1–13.

[8] KPMG, The Self-Driving Revolution - Are We Ready?, 2015.

[9] Google Self-Driving Car Project Monthly Report. *http://www.google.com/selfdrivingcar/reports/*, June 2015.

[10] Fagnant, D. J., K. M. Kockelman, and P. Bansal, Operations of a Shared Autonomous Vehicle Fleet for the Austin, Texas Market. *Transportation Research*, Vol. 28, 2015.

[11] Shaheen, S., D. Sperling, and C. Wagner, A Short History of Carsharing in the 90s. *Journal of World Transport Policy and Practice*, 1999.

[12] Lane, C., PhillyCarShare: First-Year Social and Mobility Impacts of Car Sharing in Philadelphia, Pennsylvania. *Transportation Research Record: Journal of TRB*, 2005.

[13] Costain, C., C. Ardon, and K. Habib, Synopsis of users behavior of a carsharing program: a case study in Toronto. *Proc. of Transportation Research Board Annual Meeting*, 2012.

[14] Brook, D., Carsharing: startup issues and new operational models. *Proc. Transportation Research Board Annual Meeting*, 2004.

[15] Cervero, R., City CarShare: First-Year Travel Demand Impacts. *Transportation Research Record*, 2003.

[16] Stillwater, T., P. Mokhtarian, and S. Shaheen, Carsharing and the built environment: A GIS-based study of one U.S. Operator. *Inst. of Transportation Studies, Univ. of California*, 2008.

[17] Lorimier, A., Understanding the factors affecting vehicle usage and availability in carsharing networks: a case study of Communauto carsharing system from Montreal, Canada. *International Journal of Sustainable Transportation*, 2011.

[18] Jorge, D. and G. Correia, Carsharing systems demand estimation and defined operations: a literature review. *European Journal of Transportation Infrastructure Research*, 2013.

[19] Monica Clemente, A. M. M. W. U., Maria Pia Fanti, The Vehicle Relocation Problem in Car Sharing Systems: Modeling and Simulation in a Petri Net Framework. *34th International Conference, PETRI NETS 2013, Milan, Italy, June 24-28, 2013. Proceedings*, 2013, pp. 250–269.

[20] Weikl, S. and K. Bogenberger, Relocation Strategies and Algorithms for Free-Floating Car Sharing Systems. *Intelligent Transportation Systems Magazine, IEEE*, Vol. 5, No. 4, 2013, pp. 100–111.

[21] F. Ciari, K. A., C. Dobler, Modeling one-way shared vehicle systems: an agent-based approach. *13th International Conference on Travel Behaviour Research, Toronto July 2012*, 2012.

[22] Jorge, D., G. Correia, and C. Barnhart, Comparing Optimal Relocation Operations With Simulated Relocation Policies in One-Way Carsharing Systems. *Intelligent Transportation Systems, IEEE Transactions on*, Vol. 15, No. 4, 2014, pp. 1667–1675.

[23] Pavone, M., S. Smith, E. Frazzoli, and D. Rus, Robotic Load Balancing for Mobility-on-Demand Systems. *International Journal of Robotics Research*, 2012.

[24] Zhang, R. and M. Pavone, Control of Robotic Mobility-On-Demand Systems: a Queueing-Theoretical Perspective. *Proceedings of Robotics: Science and Systems Conference*, 2014.

[25] Spieser, K., K. Treleaven, R. Zhang, E. Frazzoli, D. Morton, and M. Pavone, Toward a systematic approach to the design and evaluation of automated mobility-on- demand systems: a case study in Singapore. *Road Vehicle Automation*, 2014.

[26] Burns, L. D., W. Jordan, and B. A. Scarborough, Transforming personal mobility. *The Earth Institute*, 2013.

[27] Spliet, R., A. F. Gabor, and R. Dekker, The vehicle rescheduling problem. *Computers and Operations Research*, Vol. 43, 2014, pp. 129 – 136.

[28] Bertossi, A. A., P. Carraresi, and G. Gallo, On some matching problems arising in vehicle scheduling models. *Networks*, Vol. 17, No. 3, 1987, pp. 271–281.

[29] Carpaneto, G., M. Dell'amico, M. Fischetti, and P. Toth, A branch and bound algorithm for the multiple depot vehicle scheduling problem. *Networks*, Vol. 19, No. 5, 1989, pp. 531–548.

[30] Lenstra, J. K. and A. H. G. R. Kan, Complexity of vehicle routing and scheduling problems. *Networks*, Vol. 11, No. 2, 1981, pp. 221–227.

[31] Ann-Sophie Pepin, A. H. D. H., Guy Desaulniers, A comparison of five heuristics for the multiple depot vehicle scheduling problem. *Journal of Scheduling*, Vol. 12, No. 1, 2009, pp. 17–30.

[32] Renaud, J., G. Laporte, and F. F. Boctor, A tabu search heuristic for the multi-depot vehicle routing problem. *Computers and Operations Research*, Vol. 23, No. 3, 1996, pp. 229 – 235.

[33] Desaulniers, G., J. Lavigne, and F. Soumis, Multi-depot vehicle scheduling problems with time windows and waiting costs. *European Journal of Operational Research*, Vol. 111, No. 3, 1998, pp. 479 – 494.

[34] Treleaven, K., M. Pavone, and E. Frazzoli, Asymptotically Optimal Algorithms for One-to-one Pickup and Delivery Problems with Applications to Transportation System. *IEEE Transactions on Control*, 2013.

[35] Fagnant, D. J. and K. M. Kockelman, Dynamic ride-sharing and optimal fleet sizing for a system of shared autonomous vehicles. In *Transportation Research Board 94th Annual Meeting*, 2015, 15-1962.

[36] Papanikolaou, D., A New System Dynamics Framework for Modelling behavior of vehicle sharing systems. *Proc. of Symposium on Simulation for Architecture and urban design*, 2011.

[37] Barrios, J., On the performance of flexible carsharing: a simulation-based approach. *j*, 2012.

[38] Barth, M. and M. Todd, Simulation model performance analysis of a multiple shared vehicle system. *Transportation Research Record*, 1999.

[39] Kek, A., R. Cheu, and M. Chor, Relocation Simulation Model for multiple-station shared-use vehicle systems. *Transportation Research Record*, 2006.

[40] Efthymiou, D., C. Antoniou, and Y. Tyrinopoulos, Spatially Aware Model for Optimal Site Selection: method and Application in a Greek Mobility Center. *Transportation Research Record*, 2012.

[41] Myhrvold, C., *Uber Expectations As We Grow*, 2015.