

19 Autonomous Mobility-on-Demand Systems for Future Urban Mobility

Marco Pavone

19.1 Introduction

19.1.1 Overview

This chapter discusses the operational and economic aspects of autonomous mobility-on-demand (AMoD) systems, a transformative and rapidly developing mode of transportation wherein robotic, self-driving vehicles transport passengers in a given environment. Specifically, AMoD systems are addressed along three dimensions: (1) modeling, that is analytical models capturing salient dynamic and stochastic features of customer demand, (2) control, that is coordination algorithms for the vehicles aimed at throughput maximization, and (3) economic, that is fleet sizing and financial analyses for case studies of New York City and Singapore. Collectively, the models and methods presented in this chapter enable a rigorous assessment of the *value* of AMoD systems. In particular, the case study of New York City shows that the current taxi demand in Manhattan can be met with about 8,000 robotic vehicles (roughly 70 percent of the size of the current taxi fleet), while the case study of Singapore suggests that an AMoD system can meet the personal mobility needs of the *entire* population of Singapore with a number of robotic vehicles roughly equal to 1/3 of the current number of passenger vehicles. Directions for future research on AMoD systems are presented and discussed.

19.1.2 Personal urban mobility in the 21st century

In the past century, *private* automobiles have dramatically changed the paradigm of *personal urban mobility* by enabling *fast* and *anytime* point-to-point travel within cities. However, this paradigm is currently challenged due to a combination of factors such as dependency on oil, tailpipe production of greenhouse gases, reduced throughput caused by congestion, and ever-increasing demands on urban land for parking spaces [1]. In the US, urban vehicles consume more than half of the oil consumed by *all* sectors [2], and produce 20 percent of the total carbon dioxide emissions [3, 4]. Congestion has soared dramatically in the recent past, due to the fact that construction of new roads has not kept up with increasing transportation demand [5]. In 2011, congestion in metropolitan areas increased urban Americans' travel times by 5.5 billion hours (causing a 1 percent loss of US GDP [6]), and this

Marco Pavone

Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, 94305 e-mail: pavone@stanford.edu

figure is projected to increase by 50 percent by 2020 [6]. Parking compounds the congestion problem, by causing additional congestion and by competing for urban land for other uses. The problem is even worse on a global scale, due to the combined impact of rapid increases in urban population (to reach 5 billion, more than 60 percent of the world population, by 2030 [7]), worldwide urban population density, and car ownership in developing countries [1]. As a result, private automobiles are widely recognized as an *unsustainable solution for the future of personal urban mobility* [1].

19.1.3 The rise of mobility-on-demand

The challenge is to ensure the same benefits of privately-owned cars while removing dependency on non-renewable resources, minimizing pollution, and avoiding the need for additional roads and parking spaces. A lead to a solution for this problem comes from realizing that most of the vehicles used in urban environments are *over-engineered* and *underutilized*. For example, a typical automobile can attain speeds well over 100 miles per hour, whereas urban driving speeds are typically slow (in the 15- to 25-miles per hour range [5, 8]). Furthermore, private automobiles are parked more than 90 percent of the time [5]. Within this context, one of the most promising strategies for future personal urban mobility is the concept of *one-way vehicle sharing* using small-sized, electric cars (referred to as mobility-on-demand -MoD-), which provides stacks and racks of light electric vehicles at closely spaced intervals throughout a city [1]: when a person wants to go somewhere, she/he simply walks to the nearest rack, swipes a card to pick up a vehicle, drives it to the rack nearest to the selected destination, and drops it off.

MoD systems with electric vehicles directly target the problems of oil dependency (assuming electricity is produced cleanly), pollution, and parking spaces via higher utilization rates. Furthermore, they ensure more flexibility with respect to two-way rental systems, and provide *personal, anytime* mobility, in contrast to traditional taxi systems or alternative one-way ridesharing concepts such as carpooling, vanpooling, and buses. As such, MoD systems have been advocated as a key step toward sustainable *personal* urban mobility in the 21st century [1], and the very recent success of Car2Go (a one-way rental company operating over 10,000 two-passenger vehicles in 26 cities worldwide [9]) seems to corroborate this statement (see Figure 19.1, left).

MoD systems, however, present a number of limitations. For example, due to the spatio-temporal nature of urban mobility, trip origins and destinations are unevenly distributed and as a consequence MoD systems inevitably tend to become *unbalanced*: Vehicles will build up in some parts of a city, and become depleted at others. Additionally, MoD systems do *not* directly contribute to a reduction of congestion, as the *same* number of vehicle miles would be traveled (indeed more, considering trips to rebalance the vehicles) with the *same* origin-destination distribution.

19.1.4 Beyond MoD: autonomous mobility-on-demand (AMoD)

The progress made in the field of autonomous driving in the past decade might offer a solution to these issues. Autonomous driving holds great promise for MoD sys-



Fig. 19.1 Left figure: A Car2Go vehicle used in a traditional (i.e., non-robotic) MoD system. Right figure: Self-driving vehicle that Google will use in a 100-vehicle AMoD pilot project within the next two years. Image credit: Car2Go and Google.

tems because robotic vehicles can rebalance themselves (eliminating the rebalancing problem at its core), autonomously reach charging stations when needed, and enable system-wide coordination aimed at throughput optimization. Furthermore, they would free passengers from the task of driving, provide a personal mobility option to people unable or unwilling to drive, and potentially increase safety. These benefits have recently prompted a number of companies and traditional car manufacturers to aggressively pursue the “AMoD technology,” with activities ranging from the design of vehicles specifically tailored to AMoD operations [10, 11], to the expected launch by Google of a 100-vehicle AMoD pilot project within the next two years [12] (see Figure 19.1, right).

Rapid advances in vehicle automation technologies coupled with the increased economic and societal interest in MoD systems have fueled heated debates about the potential of AMoD systems and their economic and societal value. How many robotic vehicles would be needed to achieve a certain quality of service? What would be the cost of their operation? Would AMoD systems decrease congestion? In general, do AMoD systems represent an economically viable, sustainable, and societally-acceptable solution to the future of personal urban mobility?

19.1.5 Chapter contributions

To answer the above questions, one needs to first understand how to *control* AMoD systems, which entails optimally routing in real-time potentially hundreds of thousands of robotic vehicles. Such routing process must take into account the spatio-temporal variability of mobility demand, together with a number of constraints such as congestion and battery recharging. This represents a networked, heterogeneous, stochastic decision problem with uncertain information, hence complexity is at its heart. Within this context, the contribution of this chapter is threefold:

1. We present a spatial queueing-theoretical model for AMoD systems capturing salient dynamic and stochastic features of customer demand. A spatial queueing model entails an exogenous dynamical process that generates “transportation requests” at *spatially-localized* queues.

2. We outline two recent, yet promising approaches for the analysis and control of AMoD systems, which leverage the aforementioned spatial queueing-theoretical model. The first approach, referred to as “lumped” approach, exploits the theory of Jackson networks and allows the computation of key performance metrics and the design of system-wide coordination algorithms. The second approach, referred to as “distributed” approach, transforms the problem of controlling a set of spatially-localized queues into one of controlling a single “spatially-averaged” queue and allows the determination of *analytical* scaling laws that can be used to select system parameters (e.g., fleet sizing).
3. We discuss two case studies for the deployment of AMoD systems in New York City and Singapore. These case studies suggest that it is much more affordable (and convenient) to access mobility in an AMoD system compared to traditional mobility systems based on private vehicle ownership.

The chapter concludes with a discussion about future directions for research, with a preliminary discussion about the potential of AMoD systems to *decrease* congestion. The results presented in this chapter build upon a number of previous works by the author and his collaborators, namely [13] for the lumped approach, [14, 15, 16, 17] for the spatial queueing-theoretical framework and the distributed approach, and [13, 18] for the case studies.

The rest of this chapter is structured as follows. Section 19.2 presents a spatial queueing model for AMoD systems and gives an overview of two complementary approaches to control AMoD systems, namely, the lumped approach and the distributed approach. Section 19.3 leverages analysis and control synthesis tools from Section 19.2 to provide an initial *evaluation* of AMoD systems for two case studies of New York City and Singapore. Section 19.4 outlines directions for future research, with a particular emphasis on (and some preliminary results for) congestion effects. Finally, Section 19.5 concludes the chapter.

19.2 Modeling and controlling AMoD systems

19.2.1 Spatial queueing model of AMoD systems

At a high level, an AMoD system can be mathematically modeled as follows. Consider a given environment, where a fleet of self-driving vehicles fulfills transportation requests. Transportation requests arrive according to an *exogenous dynamic process* with associated origin and destination locations within the environment. The transportation request arrival process and the spatial distribution of the origin-destination pairs are modeled as stochastic processes, leading to a probabilistic analysis. Transportation requests queue up within the environment, which gives rise to a network of *spatially-localized* queues dynamically served by the self-driving vehicles. Such network is referred to as “spatial queueing system.” Performance criteria include the availability of vehicles upon the request’s arrival (i.e., the probability that at least one vehicle is available to provide immediate service) or average wait times to receive service. The model is portrayed in Figure 19.2.

Controlling a spatial queueing system involves a *joint task allocation and scheduling problem*, whereby vehicle routes should be dynamically designed to allocate

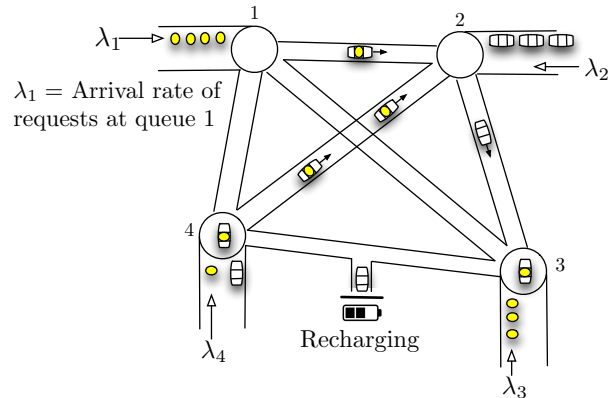


Fig. 19.2 A spatial queueing model of an AMoD system entails an exogenous dynamical process that generates “transportation requests” (yellow dots) at *spatially-localized* queues. Self-driving vehicles (represented by small car icons) travel among such locations according to a given network topology to transport passengers.

vehicles to transportation requests so as to minimize, for example, wait times. In such a dynamic and stochastic setup, one needs to design a *closed-loop control policy*, as opposed to open-loop preplanned routes. The problem combines aspects of networked control, queueing theory, combinatorial optimization, and geometric probability (i.e., probabilistic analysis in a geometrical setting). This *precludes* the direct application of “traditional” queueing theory due to the complexity added by the spatial component (these complexities include, for example, congestion effects on network edges, energy constraints, and statistical couplings induced by the vehicles’ motion [19, 20, 17]). It also precludes the direct application of combinatorial static optimization, as the dynamic aspect of the problem implies that the problem instance is *incrementally revealed over time* and static methods can no longer be applied. As a consequence, researchers have devised a number of alternative approaches, as detailed in the next section.

19.2.2 Approaches for controlling AMoD systems

This section presents two recent, yet promising approaches for the control of spatial queueing systems as models for AMoD systems, namely the lumped approach and the distributed approach. Both approaches employ a number of relaxations and approximations to overcome the difficulties in directly applying results from queueing (network) theory to spatial queueing models. A remarkable feature of these approaches is that they yield *formal performance bounds* for the control policies (i.e., factor of sub-optimality) and scaling laws for the quality of service in terms of model data, which can provide useful guidelines for selecting system parameters (e.g., number of vehicles). These approaches take their origin from seminal works on hypercube models for spatial queues [19], on the Dynamic Traveling Repairman problem [20, 21, 22, 23], and on the Dynamic Traffic Assignment problem [24, 25].

Alternative approaches could be developed by leveraging *worst-case* (as opposed to stochastic) techniques for dynamic vehicle routing, e.g., competitive (online) analysis [26, 27, 28]. This is an interesting direction for future research.

19.2.2.1 Lumped approach

Within the lumped approach [13], transportation requests are modeled by assuming that customers arrive at a set of stations located within a given environment¹, similar to the hypercube model [19]. The arrival process at each station is Poisson with a rate λ_i , where $i \in \{1, \dots, N\}$ and N denotes the number of stations. (Reasonable deviations from the assumption of Poisson arrivals have been found not to substantially alter the predictive accuracy of these models [19].) Upon arrival, a customer at station i selects a destination j according to a probability mass function $\{p_{ij}\}$ (Figure 19.3, left). If vehicles are parked at station i , the customer takes a vehicle and is driven to the intended destination, with a travel time modeled as a random variable T_{ij} . However, if the station is empty of vehicles, the customer immediately leaves the system. Under the assumptions of Poisson arrivals and exponentially-distributed travel times, an AMoD system is then translated into a Jackson network model through an *abstraction procedure* [29, 13], whereby one identifies the stations with single-server queues and the roads with infinite-server queues. (Jackson networks are a class of queueing networks where the equilibrium distribution is particularly simple to compute as the network has a product-form solution [30, 31]). With this identification, an AMoD system becomes a *closed Jackson network with respect to the vehicles*, which is amenable to *analytical* treatment [13] (Figure 19.3, left).

To control the network, for example, to (autonomously) rebalance the vehicles to ensure even vehicle availability, the strategy is to add *virtual customer streams* [13]. Specifically, one assumes that each station i generates “virtual customers” according to a Poisson process with rate ψ_i , and routes these virtual customers to station j with probability α_{ij} . The problem of controlling an AMoD system becomes one of optimizing over the rates $\{\psi_i\}$ and probabilities $\{\alpha_{ij}\}$ which, by exploiting the theory of Jackson networks, can be cast as a *linear program* (hence, this approach extends well to *large* transportation networks). This method *encourages* coordination but does *not* enforce it, which is the key to maintaining tractability of the model [13]. The rates $\{\psi_i\}$ and probabilities $\{\alpha_{ij}\}$ are then used as *feedforward reference signals* in a receding control scheme to control in real-time an entire AMoD system [13], as done for case studies of New York City and Singapore presented in Section 19.3.

19.2.2.2 Distributed approach

The key idea behind the distributed approach [14, 15, 16, 17] is that the number of stations represents a continuum (i.e., $N \rightarrow \infty$), similar to the Dynamic Traveling

¹ Alternatively, to model an AMoD system where the vehicles directly pick up the customers, one would decompose a city into N disjoint *regions* Q_1, Q_2, \dots, Q_N . Such regions would replace the notion of stations. When a customer arrives in region Q_i , destined for Q_j , a free vehicle in Q_i is sent to pick up and drop off the customer before parking at the median of Q_j . The two models are then formally *identical* and follow the same mathematical treatment.

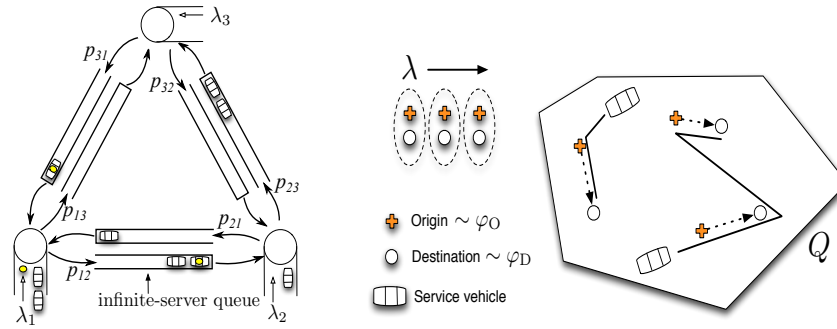


Fig. 19.3 Left figure: In the lumped model, an AMoD system is modeled as a Jackson network, where stations are identified with single-server queues and roads are identified with *infinite-server* queues. (Customers are denoted with yellow dots and servicing vehicles are represented by small car icons.) Some vehicles travel without passengers to rebalance the fleet. Right figure: In a distributed model of an AMoD system, a stochastic process with rate λ generates origin-destination pairs, distributed over a *continuous* domain Q .

Repairman problem [20, 21, 22, 23]. In other words, customers arrive at any point in a given bounded environment [15, 16], or at any point along the segments of a road map [15]. In the simplest scenario, a dynamic process generates spatially-localized origin-destination pairs (representing the transportation requests) in a geographical region $Q \subset \mathbb{R}^2$. The process that generates origin-destination pairs is modeled as a spatio-temporal Poisson process, namely, (i) the time between consecutive generation instants has an exponential distribution with intensity λ , and (ii) origins and destinations are random variables with probability density functions, respectively, φ_O and φ_D , supported over Q , see Figure 19.3 (right). The objective is to design a routing policy that minimizes the average steady-state time delay between the generation of an origin-destination pair and the time the trip is completed. By removing the constraint that customers' origin-destination pairs are localized at a finite set of points in an environment, one transforms the problem of controlling N different queues into one of controlling a single "spatially-averaged" queue. This considerably simplifies analysis and control, and allows one to derive *analytical* expressions for important design parameters. For example, one can show that a necessary *and* sufficient condition for stability is that the load factor

$$\rho := \lambda [\mathbb{E}_{\varphi_O \varphi_D} [Y - X] + \text{EMD}(\varphi_O, \varphi_D)] / vm \quad (19.1)$$

is *strictly* less than one, where m is the number of servicing vehicles, v is the average speed of the vehicles, $\mathbb{E}_{\varphi_O \varphi_D} [Y - X]$ is the expected distance between origin and destination locations, and $\text{EMD}(\varphi_O, \varphi_D)$ is the Earth mover's distance between densities φ_O and φ_D [32], representing the minimum distance, on average, a vehicle must travel to realign itself with an *asymmetrical* travel demand [16]. Intuitively, if distributions φ_O and φ_D are imagined as describing two piles each consisting of a unit of "dirt" (i.e., earth), then $\text{EMD}(\varphi_O, \varphi_D)$ can be thought of as the mini-

imum work (dirt \times distance) required to reshape φ_O into φ_D (see [32] for a formal definition). One can use the above formula to estimate the required fleet size to ensure stability – an example application to a case study of Singapore is presented in Section 19.3. With this approach, it is also possible to obtain formal performance bounds (i.e., factors of sub-optimality) for receding-horizon control policies, in the *asymptotic* regimes $\rho \rightarrow 1^-$ (heavy-load, system saturated) and $\rho \rightarrow 0^+$ (light-load, system empty of customers) [33, 17].

19.2.3 Comparison

The lumped approach and the distributed approach are *complementary* in a number of ways. Both models provide *formal guarantees* for stability and performance. The former is more realistic (a road topology can be readily mapped into this model) and provides a natural pathway to synthesize control policies. The latter provides significant mathematical simplifications (as one only needs to study a spatially-averaged queue) and enables the determination of *analytical* scaling laws that can be used to select system parameters (e.g., fleet sizing). In the next section we exploit the interplay between these two approaches to characterize AMoD systems for case studies of New York City and Singapore.

Both approaches appear to be promising tools to systematically tackle the problem of system-wide control of AMoD systems. Several research questions, however, still need to be addressed to fulfill this objective, particularly with respect to inclusion of congestion effects (in Section 19.2.2.1, roads are modeled as infinite server queues, so the travel time for each vehicle is independent of all other vehicles), predictive accuracy, and control synthesis for complex scenarios, as detailed in Section 19.4.

19.3 Evaluating AMoD systems

Leveraging models and methods from Section 19.2, this section studies hypothetical deployments of AMoD systems in two major cities, namely New York City and Singapore. Collectively, the results presented in this section provide a preliminary, yet rigorous evaluation of the benefits of AMoD systems based on real-world data. We mention that both case studies do not consider congestion effects – a preliminary discussion about these effects is presented in Section 19.4.

19.3.1 Case Study I: AMoD in Manhattan

This case study applies the lumped approach to characterize how many self-driving vehicles in an AMoD system would be required to replace the current fleet of taxis in Manhattan while providing quality service at current customer demand levels [13]. In 2012, over 13,300 taxis in New York City made over 15 million trips a month or 500,000 trips a day, with around 85 percent of trips within Manhattan. The study uses taxi trip data collected on March 1, 2012 (the data is courtesy of the New York City Taxi & Limousine Commission) consisting of 439,950 trips within Manhattan. First, trip origins and destinations are clustered into $N = 100$ stations, so that a customer is on average less than 300 m from the nearest station, or approximately a 3-minute walk. The system parameters such as arrival rates $\{\lambda_i\}$, destination prefer-

ences $\{p_{ij}\}$ and travel times $\{T_{ij}\}$ are estimated for each hour of the day using trip data between each pair of stations.

Vehicle availability (i.e., probability of finding a vehicle when walking to a station) is calculated for three cases - peak demand (29,485 demands/hour, 7-8pm), low demand (1,982 demands/hour, 4-5am), and average demand (16,930 demands/hour, 4-5pm). For each case, vehicle availability is calculated by solving the linear program discussed in Section 19.2.2.1 and then applying mean value analysis [29] techniques to recover vehicle availabilities. (The interested reader is referred to [13] for further details). The results are summarized in Figure 19.4.

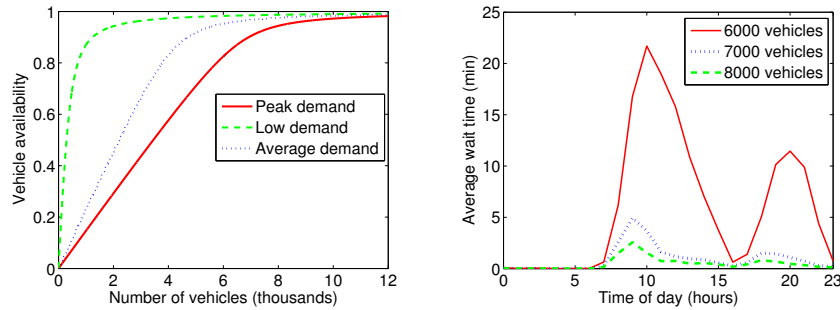


Fig. 19.4 Case study of New York City [13]. Left figure: Vehicle availability as a function of system size for 100 stations in Manhattan. Availability is calculated for peak demand (7-8pm), low demand (4-5am), and average demand (4-5pm). Right figure: Average customer wait times over the course of a day, for systems of different sizes.

For high vehicle availability (say, 95 percent), one would need around 8,000 vehicles (~ 70 percent of the current fleet size operating in Manhattan, which, based on taxi trip data, we approximate as 85 percent of the total taxi fleet) at peak demand and 6,000 vehicles at average demand. This suggests that an AMoD system with 8,000 vehicles would be able to meet 95 percent of the taxi demand in Manhattan, assuming 5 percent of customers are impatient and leave the system when a vehicle is not immediately available. However, in a real system, customers would wait in line for the next vehicle rather than leave the system, thus it is important to determine how vehicle availability relates to customer waiting times. Customer waiting times are characterized through simulation, using the receding horizon control scheme mentioned in Section 19.2.2.1. The time-varying system parameters λ_i , p_{ij} , and average speed are piecewise constant, and change each hour based on values estimated from the taxi data. Travel times T_{ij} are based on average speed and Manhattan distance between stations i and j , and self-driving vehicle rebalancing is performed every 15 minutes. Three sets of simulations are performed for 6,000, 7,000, and 8,000 vehicles, and the resulting average waiting times are shown in Figure 19.4 (right). Specifically, Figure 19.4 (right) shows that for a 7,000 vehicle fleet the peak averaged wait time is less than 5 minutes (9-10am) and, for 8,000 vehicles, the average wait time is only 2.5 minutes. The simulation results show that high availability (90-95 percent) does indeed correspond to low customer wait time and

that an autonomous MOD system with 7,000 to 8,000 vehicles (roughly 70 percent of the size of the current taxi fleet) can provide adequate service with current taxi demand levels in Manhattan.

19.3.2 Case Study II: AMoD in Singapore

This case study discusses an hypothetical deployment of an AMoD system to meet the personal mobility needs of the *entire* population of Singapore [18]. The study, which should be interpreted as a thought experiment to investigate the potential benefits of an AMoD solution, addresses three main dimensions: (i) minimum fleet size to ensure system stability (i.e., uniform boundedness of the number of outstanding customers), (ii) fleet size to provide acceptable quality of service at current customer demand levels, and (iii) financial estimates to assess economic feasibility. To support the analysis, three complementary data sources are used, namely the 2008 Household Interview Travel Survey -HITS- (a comprehensive survey about transportation patterns conducted by the Land Transport Authority in 2008 [34]), the Singapore Taxi Data -STD- database (a database of taxi records collected over the course of a week in Singapore in 2012) and the Singapore Road Network -SRD- (a graph-based representation of Singapore’s road network).

19.3.2.1 Minimum fleet sizing

The minimum fleet size needed to ensure stability is computed by applying equation (19.1), which was derived within the distributed approach. The first step is to process the HITS, STD, and SRD data sources to estimate the arrival rate λ , the average origin-destination distance $\mathbb{E}_{\varphi_O \varphi_D}[Y - X]$, the demand distributions φ_O and φ_D , and the average velocity v . Given such quantities, equation (19.1) yields that at least 92,693 self-driving vehicles are required to ensure the transportation demand remains uniformly bounded. To gain an appreciation for the level of vehicle sharing possible in an AMoD system of this size, consider that at 1,144,400 households in Singapore, there would be roughly one shared car every 12.3 households. Note however, that this should only be seen as a lower bound on the fleet size, since customer waiting times would be unacceptably high.

19.3.2.2 Fleet sizing for acceptable quality of service

To ensure acceptable quality of service, one needs to increase the fleet size. To characterize such increase, we use the same techniques outlined in Section 19.3.1, which rely on the lumped approach. Vehicle availability is analyzed in two representative cases. The first is chosen as the 2-3pm bin, since it is the one that is the closest to the “average” traffic condition. The second case considers the 7-8am rush-hour peak. Results are summarized in Figure 19.5 (left). With about 200,000 vehicles, availability is about 90 percent on average, but drops to about 50 percent at peak times. With 300,000 vehicles in the fleet, availability is about 95 percent on average and about 72 percent at peak times. As in Section 19.3.1, waiting times are characterized through simulation. For 250,000 vehicles, the maximum wait time during peak hours is around 30 minutes, which is comparable with typical congestion delays during rush hour. With 300,000 vehicles, peak wait times are reduced to less than 15 minutes, see Figure 19.5 (right). To put these numbers into perspective, in

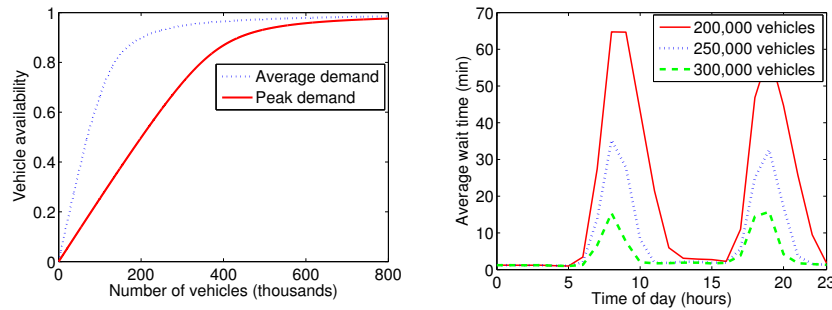


Fig. 19.5 Case study of Singapore [18]. Left figure: Performance curve with 100 regions, showing the availability of vehicles vs. the size of the system for both average demand (2-3pm) and peak demand (7-8am). Right figure: Average wait times over the course of a day, for systems of different sizes.

2011 there were 779,890 passenger vehicles operating in Singapore [35]. Hence, this case study suggests that an AMoD system can meet the personal mobility need of the entire population of Singapore with a number of robotic vehicles roughly equal to 1/3 of the current number of passenger vehicles.

19.3.2.3 Financial analysis of AMoD systems

This section provides a preliminary, yet rigorous economic evaluation of AMoD systems. Specifically, this section characterizes the total mobility cost (TMC) for users in two competing transportation models. In System 1 (referred to as *traditional system*), users access personal mobility by purchasing (or leasing) a private, human-driven vehicle. Conversely, in System 2 (the AMoD system), users access personal mobility by subscribing to a shared AMoD fleet of vehicles. For both systems, the analysis considers not only the explicit costs of access to mobility (referred to as cost of service -COS-), but also hidden costs attributed to the time invested in various mobility-related activities (referred to as cost of time -COT-). A subscript $i = \{1, 2\}$ will denote the system under consideration (e.g., COS_1 denotes the COS for System 1).

Cost of service: The cost of service is defined as the sum of all explicit costs associated with accessing mobility. For example, in System 1, COS_1 reflects the costs to individually purchase, service, park, insure, and fuel a private, human-driven vehicle, which, for the case of Singapore, are estimated for a mid-size car at \$18,162/year. For System 2, one needs to make an educated guess for the cost incurred in retrofitting production vehicles with the sensors, actuators, and computational power required for automated driving. Based upon the author's and his collaborators' experience on self-driving vehicles, such cost (assuming some economies of scale for large fleets) is estimated as a one-time fee of \$15,000. From the fleet-sizing arguments of Section 19.3.2.2, one shared self-driving vehicle in System 2 can effectively serve the role of about four private, human-driven vehicles in System 1, which implies an estimate of 2.5 years for the average lifespan of a self-driving ve-

Table 19.1 Summary of the financial analysis of mobility-related cost for traditional and AMoD systems for a case study of Singapore [18].

	Cost [USD/km]			Yearly cost [USD/year]		
	COS	COT	TMC	COS	COT	TMC
Traditional	0.96	0.76	1.72	18,162	14,460	32,622
AMoD	0.66	0.26	0.92	12,563	4,959	17,522

hicle. Tallying the aforementioned costs on a fleet-wide scale and distributing the sum evenly among the entire Singapore population gives a COS_2 of \$12,563/year (see [18] for further details about the cost breakdown). According to COS values, it is more affordable to access mobility in System 2 than System 1.

Cost of time: To monetize the hidden costs attributed to the time invested in mobility-related activities, the analysis leverages the Value of Travel Time Savings (VTTS) numbers laid out by the Department of Transportation for performing a cost-benefit analysis of transportation scenarios in the US [36]. Applying the appropriate VTTS values based on actual driving patterns gives $COT_1 = \$14,460/\text{year}$ (which considers an estimated 747 hours/year spent by vehicle owners in Singapore in mobility-related activities, see [18]). To compute COT_2 , this analysis prices sitting comfortably in a shared self-driving vehicle while being able to work, read, or simply relax at 20 percent of the median wage (as opposed to 50 percent of the median wage which is the cost of time for driving in free-flowing traffic). Coupling this figure with the fact that a user would spend no time parking, limited time walking to and from the vehicles, and roughly 5 minutes for a requested vehicle to show up (see Section 19.3.2.2), the end result is a COT_2 equal to \$4,959/year.

Total mobility cost: A summary of the COS, COT, and TMC for the traditional and AMoD systems is provided in Table 19.1 (note that the average Singaporean drives 18,997 km in a year). Remarkably, combining COS and COT figures, the *TMC for AMoD systems is roughly half of that for traditional systems*. To put this into perspective, these savings represent about one third of GDP per capita. Hence, this analysis suggests that it is much more affordable to access mobility in an AMoD system compared to traditional mobility systems based on private vehicle ownership.

19.4 Future research directions

This chapter provided an overview of modeling and control techniques for AMoD systems, and a preliminary evaluation of their financial benefits. Future research on this topic should proceed along two main dimensions: efficient control algorithms for increasingly more realistic models and eventually for real-world test beds, and financial analyses for a larger number of deployment options and accounting for positive externalities (e.g., increased safety) in the economic assessment. Such research directions are discussed in some details next, with a particular emphasis on the inclusion of congestion effects and some related preliminary results.

19.4.1 Future research on modeling and control

A key direction for future research is the inclusion of congestion effects. In AMoD systems, congestion manifests itself as constraints on the road capacity, which in turn affect travel times throughout the system. To include congestion effects, a promising strategy is to study a modified lumped model whereby the infinite-server road queues are changed to queues with a *finite* number of servers, where the number of servers on each road represents the *capacity* of that road [13]. This approach is used in Figure 19.6 on a simple 9-station network, where the aim is to illustrate the impact of autonomously rebalancing vehicles on congestion. Specifically,

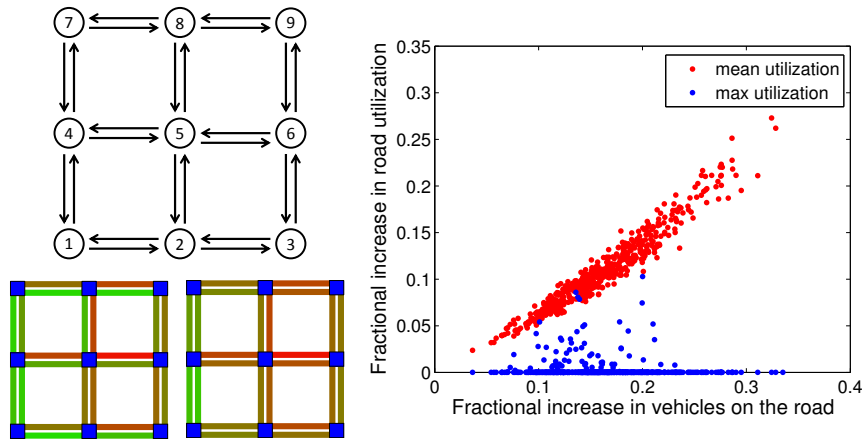


Fig. 19.6 Congestion effects in AMoD systems [13]. Top left: Layout of the 9-station road network. Each road segment has a capacity of 40 vehicles in each direction. Bottom left: The first picture shows the 9-station road network without rebalancing. The color on each road segment indicates the level of congestion, where green is no congestion, and red is heavy congestion. The second picture is the same road network with rebalancing vehicles. Right: The effects of rebalancing on congestion. The x-axis is the ratio of rebalancing vehicles to passenger vehicles on the road. The y-axis is the fractional increase in road utilization due to rebalancing.

the stations are placed on a square grid, and joined by 2-way road segments, each of which is 0.5 km long. Each road consists of a single lane, with a critical density of 80 vehicles/km. Each vehicle travels at 30 km/hour in free flow, which means the travel time along each road segment is 1 minute in free flow. Figure 19.6 plots the vehicle and road utilization increases due to rebalancing for 500 randomly generated systems (where the arrival rates and routing distributions are randomly generated). The routing algorithm for the rebalancing vehicles is a simple open-loop strategy based on the linear program discussed in Section 19.2.2.1. The x-axis shows the ratio of rebalancing vehicles to passenger vehicles on the road, which represents the inherent imbalance in the system. The red data points represent the increase in average road utilization due to rebalancing and the blue data points represent the utilization increase in the most congested road segment due to rebalancing. It is no surprise that the average road utilization rate is a linear function of the number of

rebalancing vehicles. However, remarkably, the maximum congestion increases are much lower than the average, and are in most cases zero. This means that while rebalancing generally increases the number of vehicles on the road, *rebalancing vehicles mostly travel along less congested routes and rarely increase the maximum congestion in the system*. This can be seen in Figure 19.6 bottom left, where rebalancing clearly increases the number of vehicles on many roads but not on the most congested road segment (from station 6 to station 5).

The simple setup in Figure 19.6 suggests that AMoD systems would, in general, not lead to an increase in congestion. A particularly interesting and intriguing research direction is to devise routing algorithms for AMoD systems that indeed lead to a *decrease* in congestion with current demand levels (or even higher). A promising strategy relies on the idea that if AMoD systems are implemented such that passengers are given precise pickup times and trips are staggered to avoid too many trips at the same time, congestion may be reduced. Passengers may still spend the same amount of time between requesting a vehicle and arrival at their destination, but the time spent waiting for the vehicle could be used for productive work as opposed to being stuck in traffic. Specifically, for highly congested systems, vehicle departures can be staggered to avoid excessive congestion, and the routing problem is similar to the simultaneous departure and routing problem [37].

Besides congestion, several additional directions are open for future research. As far as modeling is concerned, those include (i) analysis in a time-varying setup (e.g., with periodically time-varying arrival rates), (ii) inclusion of mesoscopic and microscopic effects into the models (e.g., increased throughput due to platooning or automated intersections), and (iii) more complex models for the transportation requests (e.g., time windows or priorities). On the control side, those include (i) inclusion of recharging constraints in the routing process, (ii) control of AMoD systems as part of a *multi-modal* transportation network, which should address synergies between AMoD and alternative transportation modes and interactions with human-driven vehicles, and (iii) deployment of control algorithms on real-world test beds.

19.4.2 Future research on AMoD evaluation

The AMoD evaluation presented in Section 19.3 already showed that AMoD systems might hold significant financial benefits. Remarkably, such financial benefits might be even larger when one also accounts for the positive externalities of an AMoD system, e.g., improved safety, freeing up urban land for other uses, and even creating a new economy based on infotainment systems onboard the self-driving vehicles. Such additional benefits, however, have not been thoroughly characterized yet and require additional analyses. Another research direction involves the evaluation of AMoD systems for more complex deployment options, e.g., as a last-mile solution within a multi-modal transportation system, or with a more sophisticated service structure, e.g., multiple priority classes.

19.5 Conclusions

This chapter overviewed recent results regarding the modeling, control, and evaluation of autonomous mobility-on-demand systems. Case studies of New York City

and Singapore suggest that it would be much more affordable (and more convenient) to access mobility in an AMoD system compared to traditional mobility systems based on private vehicle ownership. More studies are however needed to devise efficient, system-wide coordination algorithms for complex AMoD systems as part of a multi-modal transportation network, and to fully assess the related economic benefits.

Acknowledgments

The author acknowledges the collaboration with Emilio Frazzoli (MIT), Rick Zhang (Stanford), Kyle Treleaven (MIT), Kevin Spieser (MIT) and Daniel Morton (SMART center) on the results presented in this chapter.

19.6 References

1. W. J. Mitchell, C. E. Borroni-Bird, and L. D. Burns. *Reinventing the Automobile: Personal Urban Mobility for the 21st Century*. The MIT Press, Cambridge, MA, 2010.
2. International Energy Outlook 2013. Technical report, U.S. Energy Information Administration, 2013.
3. The Emissions Gap Report 2013 - UNEP. Technical report, United Nations Environment Programme, 2013.
4. U.S. Environmental Protection Agency. Greenhouse Gas Equivalencies Calculator. Online: <http://www.epa.gov/cleanenergy/energy-resources/refs.html>, 2014.
5. Our Nation's Highways: 2011. Technical report, Federal Highway Administration, 2011.
6. D. Schrank, B. Eisele, and T. Lomax. TTI's 2012 urban mobility report. Technical report, Texas A&M Transportation Institute, Texas, USA, 2012.
7. UN. World urbanization prospects: The 2011 revision population database. Technical report, United Nations, 2011.
8. A. Santos, N. McGuckin, H. Y. Nakamoto, D. Gray, and S. Liss. Summary of travel trends: 2009 national household travel survey, 2011.
9. CAR2GO. CAR2GO Austin. Car Sharing 2.0: Great idea for a great city. online: <http://www.car2go.com/austin/en/>, 2011.
10. J. Motavalli. G.M. EN-V: Sharpening the focus of future urban mobility. Online: <http://wheels.blogs.nytimes.com/2010/03/24/g-m-en-v-sharpening-the-focus-of-future-urban-mobility>. The New York Times, 24 March 2010.
11. Induct. Navia - the 100% electric automated transport. Online: <http://induct-technology.com/en/products/navia-the-100-electric-automated-transport>, 2013.
12. Google. Just press go: designing a self-driving vehicle. Online: <http://googleblog.blogspot.com/2014/05/just-press-go-designing-self-driving.html>, 2014.
13. R. Zhang and M. Pavone. Control of robotic mobility-on-demand systems: a queueing-theoretical perspective. In *Robotics: Science and Systems Conference*, 2014.
14. M. Pavone. *Dynamic vehicle routing for robotic networks*. PhD thesis, Massachusetts Institute of Technology, 2010.
15. K. Treleaven, M. Pavone, and E. Frazzoli. Models and asymptotically optimal algorithms for pickup and delivery problems on roadmaps. In *Proc. IEEE Conf. on Decision and Control*, pages 5691–5698, 2012.
16. K. Treleaven, M. Pavone, and E. Frazzoli. Asymptotically optimal algorithms for one-to-one pickup and delivery problems with applications to transportation systems. *IEEE Trans. on Automatic Control*, 58(9):2261–2276, 2013.
17. E. Frazzoli and M. Pavone. Multi-vehicle routing. In *Springer Encyclopedia of Systems and Control*. Springer, 2014.
18. K. Spieser, 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. In *Road Vehicle Automation*. Springer, 2014.

19. R. C. Larson and A. R. Odoni. *Urban operations research*. Prentice-Hall, 1981.
20. D. J. Bertsimas and G. J. van Ryzin. A stochastic and dynamic vehicle routing problem in the Euclidean plane. *Operations Research*, 39:601–615, 1991.
21. D. J. Bertsimas and D. Simchi-Levi. A new generation of vehicle routing research: robust algorithms, addressing uncertainty. *Operations Research*, 44(2):286–304, 1996.
22. D. J. Bertsimas and G. J. van Ryzin. Stochastic and dynamic vehicle routing in the Euclidean plane with multiple capacitated vehicles. *Operations Research*, 41(1):60–76, 1993.
23. D. J. Bertsimas and G. J. van Ryzin. Stochastic and dynamic vehicle routing with general interarrival and service time distributions. *Advances in Applied Probability*, 25:947–978, 1993.
24. T. L. Friesz, J. Luque, R. L. Tobin, and B. W. Wie. Dynamic network traffic assignment considered as a continuous time optimal control problem. *Operations Research*, 37(6):893–901, 1989.
25. S. Peeta and A. Ziliaskopoulos. Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, 1:233–265, 2001.
26. E. Feuerstein and L. Stougie. On-line single-server dial-a-ride problems. *Theoretical Computer Science*, 268(1):91–105, 2001.
27. P. Jaillet and M. R. Wagner. Online routing problems: Value of advanced information and improved competitive ratios. *Transportation Science*, 40(2):200–210, 2006.
28. G. Berbeglia, J. F. Cordeau, and G. Laporte. Dynamic pickup and delivery problems. *European Journal of Operational Research*, 202(1):8–15, 2010.
29. D. K. George and C. H. Xia. Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. *European Journal of Operational Research*, 211(1):198–207, 2011.
30. J. R. Jackson. Networks of waiting lines. *Operations Research*, 5(4):518–521, 1957.
31. J. R. Jackson. Jobshop-like queueing systems. *Management science*, 10(1):131–142, 1963.
32. L. Ruschendorf. The Wasserstein distance and approximation theorems. *Probability Theory and Related Fields*, 70:117–129, 1985.
33. M. Pavone, K. Treleaven, and E. Frazzoli. Fundamental performance limits and efficient policies for transportation-on-demand systems. In *Proc. IEEE Conf. on Decision and Control*, pages 5622–5629, 2010.
34. Singapore Land Transport Authority. 2008 Household interview travel survey background information, 2008.
35. Land Transport Authority. Singapore land transit statistics in brief, 2012.
36. HEATCO. Harmonized European approaches for transport costing and project assessment. Online: <http://heatco.ier.uni-stuttgart.de>, 2006.
37. H. Huang and W. H. K. Lam. Modeling and solving the dynamic user equilibrium route and departure time choice problem in network with queues. *Transportation Research Part B: Methodological*, 36(3):253–273, 2002.