# DECENTRALIZED AUTONOMOUS ELECTRIC MOBILITY-ON-DEMAND SERVICES FOR INDIVIDUALS WITH PHYSICAL AND COGNITIVE DISABILITIES

## FINAL PROJECT REPORT

by

Syrine Belakaria, Mustafa Ammous, Sameh Sorour, and Ahmed Abdel-Rahim
University of Idaho

Sponsorship
PacTrans and the University of Idaho

for
Pacific Northwest Transportation Consortium (PacTrans)
USDOT University Transportation Center for Federal Region 10
University of Washington
More Hall 112, Box 352700
Seattle, WA 98195-2700

In cooperation with US Department of Transportation- Office of the Assistant Secretary for Research and Technology (OST-R)

**Disclaimer**

**The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Pacific Northwest Transportation Consortium, the U.S. Government and matching sponsor assume no liability for the contents or use thereof.**

# Technical Report Documentation Page

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| | 01701508 | |

**4. Title and Subtitle**

Decentralized Autonomous Electric Mobility-on-Demand Services for Individuals with Physical and Cognitive Disabilities

**5. Report Date**

**6. Performing Organization Code**

**7. Author(s)**

Syrine Belakaria, Mustafa Ammous, Sameh Sorour, ORCID# 0000-0002-3936-7833 and Ahmed Abdel-Rahim, ORCID# 0000-0001-9756-554X

**8. Performing Organization Report No.**

2017-S-UI-3

**9. Performing Organization Name and Address**

PacTrans

Pacific Northwest Transportation Consortium

University Transportation Center for Region 10

University of Washington More Hall 112 Seattle, WA 98195-2700

**10. Work Unit No. (TRAIS)**

**11. Contract or Grant No.**

69A3551747110

**12. Sponsoring Organization Name and Address**

United States of America

Department of Transportation

Research and Innovative Technology Administration

**13. Type of Report and Period Covered**

Research Final

**14. Sponsoring Agency Code**

**15. Supplementary Notes**

Report uploaded at www.pacTrans.org

**16. Abstract**

This report discusses the foundation of innovative decentralized mobility services for individuals with physical or cognitive disabilities using disability-friendl,y autonomous electric mobility on-demand (AEMoD) services. By leveraging both the Internet-of-things (IoT) and its associated fog control capabilities, this framework will enable real-time, localized, autonomous, disability-aware, and battery-level-based dispatching and charging decisions for a fleet of AEMoD services distributed in multiple city zones. Through IoT-enabled fog control, computational resources are pushed closer to customers in each city zone, which enables the collection of real-time information about the AEMoD services, tracking of their state-of-charge, and the collection of service requests from customers. Driven by these collected data, these fog controllers will provide decentralized, efficient, and responsive dispatching and charging decisions within each zone to serve customers, prioritizing their needs in a timely manner, while maintaining suitable AEMoD state-of-charge for subsequent trips.

The proposed fog-based architectures for localized AEMoD system operations provides a good solution for the communication/ computation delays that restrict massive AEMoD operations in big cities. These emerging architectures will soon become widely used, allowing all localized operational decisions to be made with very low latency by fog controllers located close to the end applications (e.g., each city zone for AEMoD systems). The proposed architecture also employs an optimized, multi-class charging and dispatching queuing model, with a partial charging option for AEMoD vehicles, to provide the best solution to the AEMoD charging delay challenges for each zone. The stability conditions of this model and the optimal number of classes were derived. The decisions on the proportions of each class of vehicles to partially/fully charge or directly serve customers are optimized to minimize the maximum and average system response times by using convex optimization and Lagrangian analysis. Analysis results showed the merits of our proposed model and optimized decision scheme in comparison to both the always-charge and the equal-split schemes. Furthermore, the comparison between the maximum and average problem solutions exhibited negligible variance, which favored the use of the maximum solution because of its lower complexity.

**17. Key Words**

Autonomous mobility on-demand, electric vehicle, fog-based architecture, dispatching, charging, queuing systems.

**18. Distribution Statement**

No restrictions.

| 19. Security Classification (of this report) | 20. Security Classification (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified. | Unclassified. | | NA |

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

## Table of Contents



## List of Figures

## List of Tables

## List of Abbreviations

| | |
|---|---|
| AEMoD | Autonomous electric mobility on-demand |
| AI | Artificial intelligence |
| BCMP | Baskett, Chandy Muntz, Palacios |
| IoT | Internet of things |
| KKT | Karush-Kuhn-Tucker |
| MCC | Mobile cloud computing |
| MEC | Mobile edge computing |
| MoD | Mobility on demand |
| MPC | Model predictive control |
| PEV | Plug-in electric vehicle |
| RAN | Radio Access Network |
| SoC | State of charge |

## Executive Summary

Despite significant advances in vehicle automation and electrification, the deployment of autonomous electric mobility-on-demand (AEMoD) services in big cities encounters two major bottlenecks, namely, communication/computation delays and charging delays. Additionally, with respect to services for individuals with cognitive and physical disabilities, none of the AEMoD routing and scheduling algorithms incorporate or prioritize special users' needs in the optimization process. This project established a foundation for innovative, decentralized mobility services that also consider the needs of individuals with physical or cognitive disabilities. Leveraging both the Internet-of-things (IoT) and its associated fog control capabilities, this framework will enable real-time, localized, autonomous, disability-aware, and battery-level-based dispatching and charging decisions for a fleet of AEMoD services distributed in multiple city zones.

In order to target communication/computation delays, the work presented in this report exploited emerging fog-based architectures for localized AEMoD system operations. These architectures will soon become widely used, allowing all localized operational decisions to be made with very low latency by fog controllers located close to the end applications (e.g., each city zone for AEMoD systems). As for charging delays, an optimized, multi-class charging and dispatching queuing model, with a partial charging option for AEMoD vehicles, was developed for use in each zone. The stability conditions of this model and the optimal number of classes were derived. Decisions about the proportions of each class of vehicles to partially/fully charge or directly serve priority customers were optimized to minimize system response times by using convex optimization and Lagrangian analysis. Analysis results showed the merits of our proposed model and optimized decision scheme in comparison to both the always-charge and the

equal-split schemes. Furthermore, a comparison between the maximum and average problem solutions exhibited negligible variance, which favored the use of the maximum solution because of its lower complexity.

# Chapter 1 Introduction

Urban transportation systems are experiencing tremendous challenges due to the high demands of private vehicle ownership, which result in dramatic increases in road congestion, parking demand [Mitchell et al. 2010], increased travel times [Schrank et al. 2012], and carbon footprint [Emissions Gap Report 2013; U.E.P. Agency 2014]. These problems clearly call for revolutionary solutions to sustain the future of private vehicle mobility. Mobility ondemand (MoD) services have been successful at providing a partial solution to the increased private vehicle ownership problem [Digitalist Magazine 2016] by providing one-way vehicle sharing between dedicated pick-up and drop-off locations for a monthly subscription fee and without the need for customers to purchase vehicle insurance or to incur maintenance costs. The electrification of such MoD vehicles can also gradually reduce the carbon footprint problem. However, the need to make extra trips to pick up, drop off, and occasionally refueling /charge these MoD vehicles has significantly affected the convenience of this solution and reduced its effectiveness at solving urban traffic problems.

Expected to be game-changers for the success of these services are significant advances in vehicle automation and wireless connectivity. With more than 10 million self- driving vehicles expected to be on the road by 2020 [Digitalist Magazine 2016], and the vision of governments and automakers to inject more wireless connectivity and coordinated optimization on city roads, private vehicle ownership has been forecasted to significantly decline by 2025, as individuals' private mobility will further depend on the concept of autonomous electric MoD (AEMoD) [Navigant Research 2016; Forbes 2016]. In short, AEMoD systems will enable customers to simply press some buttons on an app to promptly get an autonomous electric vehicle to transport them door-to-door, with no pick-up/drop- off and driving responsibilities, no dedicated parking needs, no carbon emission, no vehicle insurance and maintenance costs, and extra in-vehicle work/leisure times. With all of these ecological, economical, and customer-oriented qualities, AEMoD systems are expected to significantly prevail in attracting millions of customers across the world and in providing on-demand and hassle-free private urban mobility.

Despite the great expectations for wide deployment of AEMoD service in the next decade, the timeliness of such service (i.e., promptness in providing a ready vehicle to each requesting customer with minimum or bounded delays), and therefore its entire success, is threatened by two major bottlenecks. First, the expected massive demand for AEMoD services will result in excessive, if not prohibitive, computational and communication delays if cloud-based approaches are employed for micro-operation of such systems (e.g., collecting requests, and optimizing dispatching and charging decisions). Moreover, the typical full-battery charging rates of electric vehicles will not be able to cope with the significant numbers of vehicles involved in these systems, thus resulting in instabilities and unbounded customer delays.

## Chapter 2 Background and Literature Review

### 2.1 Background

Mobility-on-demand services have been studied from several perspectives, since the performance of these systems depends on many factors, such as charging resources, customers' special needs, and waiting time. Recent work has addressed important problems related to AMoD systems by building different operation models for them. Zhang et al. (2015) proposed two different models: a distributed queuing model that spatially averaged the customers' queues into one queue, and a lumped model that exploited the theory of Jackson Networks. These models were employed to analyze the re-balancing between the stations.

Zhang and Pavone (2016) proposed a lumped spatial-queuing model. Several non-practical assumptions were made in order to treat the problem as a Jackson Network. The work presented by Zhang et al. (2016a) cast an AMoD system into a closed multi-class BCMP queuing network model, solving the routing problem for rebalancing vehicles on congested roads. Many key factors, however, were not considered in this work in order to simplify the mathematical resolution. In addition, none of this reserch considered the computational architecture for massive demands on such services, vehicle electrification, and the influence of charging limitations on stability.

Zhang et al. (2016b) presented a model predictive control (MPC) approach to optimize the dispatching and scheduling of the vehicles in AMoD systems. It is valuable to apply MPC algorithms to minimize the future waiting time of customers, but the proposed system was optimized without consideration for AMoD vehicle electrification. The MPC technique was also used in more recent work (Tang and Zhang (2017), in which a finite-horizon dynamic programming algorithm was proposed  to provide optimal schedules for plug-in electric vehicle

(PEV) charging, given statistical information on the vehicles' future charging demands. The focus of this research was more on reducing algorithm complexity in comparison to similar algorithms proposed by Rao and Yao (2014) and Bansal et al. (2014).

Other researchers have designed an artificial neural network to predict the quality of service of an MoD system that utilizes a small number of vehicles to meet campus demand [Kumru et al. 2017; Kim et al. 2017; Miller and How 2017]. They showed that combined predictive positioning and ridesharing approaches could achieve effective MOD fleet management performance, but this performance would not be practical at a city scale.

Treleaven et al. (2013) addressed the vehicle dispatching problem on the basis of distances separating vehicles from customers. They employed a combination of the Euclidean bipartite matching problem and random permutation theory to minimize the trip cost, but without considering charging limitations. Charging AEMoD vehicles has also been studied from different perspectives. Some researchres [Korkas et al. 2018; Tushar et al. 2016] proposed optimization models to reduce cost in terms of power and energy. Korkas et al. (2016) provided a valuable analysis of an approximate dynamic programming system with feedback-based optimization for the charging process.

Tushar et al. (2016) presented a time variant cost optimization for charging at Photovoltaics charging stations. Rao and Yao (2014) studied the involvement of smart grids for energy cost optimization not only with closed loop and open loop methods, but also by using artificial intelligence techniques. These techniques allowed the researchers to introduce several agents and complex models, such as considering the vehicles themselves as sources of energy that could contribute to the grid. Artificial intelligence (AI) might produce valuable outcomes, but the cost and complexity of deploying these methods would be high. Our work was different from that of Korkas et al., Tushar et al., and Rigas et al. (2015) because it aimed to optimize system response/waiting times to improve customer satisfaction.

## 2.2 Project Contributions

In this project, we targeted the two limitations related to timeliness that could hinder the success of AEMoD systems, namely communication/computation delays and possible system instability due to the charging process.

To resolve the first limitation, we suggest exploiting the new and trendy fog-based networking and computing architectures [Cisco 2015; Mao et al. 2017; Mach and Becvar 2017]. While long propagation delays remain a key drawback for centralized cloud computing, mobile edge computing (MEC) with proximate access is widely agreed to be a key technology for supporting various applications for the next-generation Internet with millisecond-scale reaction time [Fettweis 2014]. The advantages provided by this technology will allow the handling of vehicular networks [Huang 2017] that require instantaneous decision making, such autonomous mobility [Markakis et al. 2017b]. Consequently, these architectures can also be involved in handling AEMoD system operations in a distributed way. This approach will push the operational decision load closer to end customers in each city zone, thereby reducing computational complexity and communication delays. Fortunately, this architecture suits the nature of many AEMoD fleet operations that are mostly local, such as dispatching and charging. Indeed, AEMoD vehicles will be usually directed to pick up customers close to their locations and to charge at nearby charging stations. This makes fog-based architectures well-suited for localized solutions to guarantee low communication and computation latencies for such local management operations.

With this limitation resolved by these soon-to-be-deployed technologies, this project focused on resolving the second timeliness limitation by proposing a multi-class dispatching and charging approach in each service zone. The proposed approach classifies incoming vehicles according to their state of charge (SoC) and smartly manages their charging options according to the available charging resources in the zone. This management is done by introducing the option of no charging or partial charging for vehicles with non-depleted batteries and by assigning the full charging option only to vehicles will fully depleted batteries. This multi-class system also allocates these vehicles to different classes of customers according to the suitability of the vehicles' SoC for the customers' trip distance.

Given these novel system operation architectures, the question then becomes: What is the optimal proportion of vehicles from each class to either dispatch (i.e., with no charging) or to partially/fully charge, both to maintain charging stability and to minimize the maximum or average response times of the system? To address this question, a queuing model representing the proposed fog-based, multi-class charging and dispatching scheme was developed. The stability conditions of this model and the number of classes that fit the charging capabilities of the service zone were then derived. Decisions about the proportions of each class of vehicles to partially/fully charge or directly serve customers were then optimized to minimize the maximum and average system response times. Maximum response time minimization was formulated as a stochastic linear problem, and and average response time minimization was formulated as a convex optimization problems. Optimal decisions were analytically derived for each problem by using Lagrangian analysis. Finally, the merits of our proposed optimized decision scheme were tested and compared to both the always-charge and the equal split schemes.

# Chapter 3 Proposed System Model

## 3.1 Fog-Based Architecture

Fog-based architectures have recently emerged as novel distributed edge computing architectures both for mitigating the communication and computational burdens on backhaul networks and cloud servers, respectively, and for reducing the delays for system analytics and decision making. These architectures push computational resources closer to the end entities, thus providing them with low complexity and latency analytics and optimization solutions through local communication with those resources. The concept of mobile edge computing (MEC) was first proposed by the European Telecommunications Standard Institute (ETSI) in 2014. It was defined as a new platform that provides IT and mobile cloud computing capabilities (MCC) within the Radio Access Network (RAN) in close proximity to mobile subscribers [Patel et al. 2014]. This led to the emergence of a new research area called fog computing and networking [Chaing and Zhang 2016]. It has been widely agreed and proved that MEC will solve the delay disadvantages of mobile cloud computing[Fettweis 2014]. Mao et al. (2017) provided a clear comparison that showed the benefits offered by MEC in comparison to MCC. The supportable latency of MEC is less than 10 milliseconds in comparison to larger than 100 milliseconds for MCC. It has also been shown that the fog computing has $10^2$ to $10^4$ times higher computation capabilities than the minimum requirements for applications with heavy computational complexity such as gaming [Satyanarayanan et al. 2009], autonomous driving, and instantaneous decision making [Patel et al. 2014; Online]. Moreover, with its shorter distance to end users, fog computing supports less backhaul usage, thus alleviating congestion [Tran et al. 2015; Markakis et al. 2017a]. In addition to these advantages, fog-based architectures are highly energy efficient with respect to supporting computation offloading and are therefore considered to be green technologies [Mach and Becvar 2017; Somov and Giaffreda 2015; Markakis et al. 2017b].

As mentioned earlier, our proposal to employ a fog-based architecture for AEMoD systems was justified by the fact that many AEMoD operations (e.g., dispatching and charging) are localized, with very high demand and the the need for instantaneous decision-making. Indeed, the vehicles located in any city zone are the ones that can reach customers in that same zone within a limited time frame. They will also charge at nearby charging points within the zone. Figure 3-1 illustrates a candidate fog-based architecture that can support real-time micro-operational decisions (e.g., dispatching and charging) for AEMoD systems with extremely low computation and communications delays. The fog controller in each service zone is responsible for collecting information about customer requests, vehicle in-flow to the service zone, vehicle state of charge (SoC), and the available full-battery charging rates in the service zone. Given the collected information, the architecture can promptly make dispatching, and charging decisions for these vehicles in a timely manner.

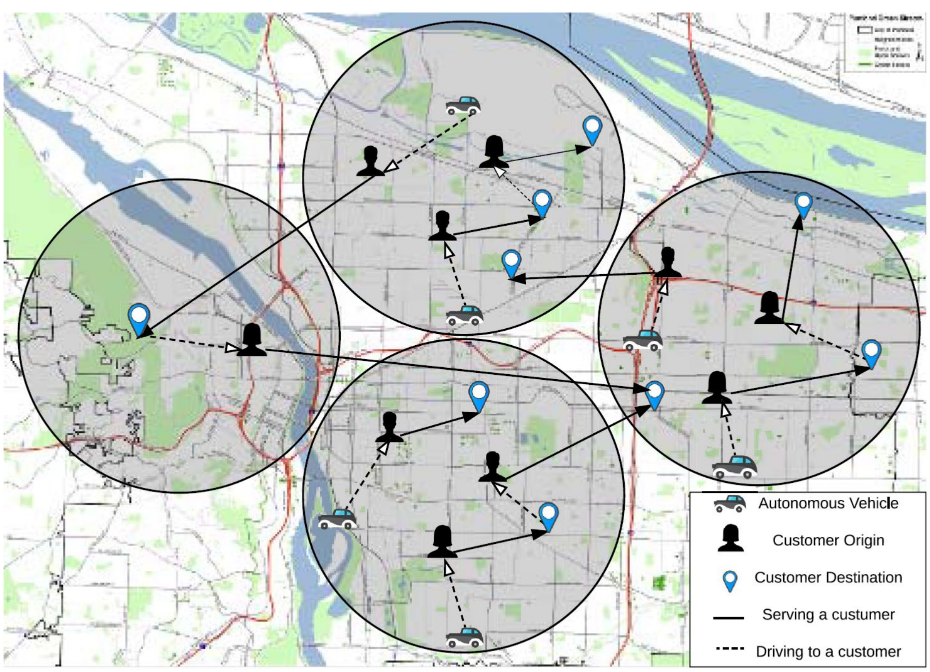

**Figure 3-1** Fog-based architecture for AEMoD system operation.

## 3.2 Multi-Class Dispatching and Charging Model

To guarantee the stability and timeliness of future AEMoD systems, given the relatively limited charging resources in comparison to demand volumes, it is critical to answer two important operational questions:

(1) How to cope with the available charging capabilities of each service zone, given the number of system vehicles?

(2) How to smartly manage the dispatching and charging options of different SoC vehicles, given customers' needs and zone resources, in order to minimize the maximum and/or average system response time?

By system *response time*, we mean the time elapsed between the instant when a customer requests a vehicle and the instant when a vehicle starts moving from its parking or charging spot toward this customer.

Motivated by the fact that different customers can be classified in ascending order of their required trip distances (and thus the SoC needed for their allocated vehicles), we proposed to address the above two questions by introducing a multi-class dispatching and charging scheme for AEMoD vehicles, with an option for no charging or partial charging for vehicles with non-depleted batteries. Arriving vehicles in each service zone are subdivided into different classes in ascending order of their SoC, corresponding to the different customer classes. Different proportions of each class of vehicles will then be prompted by the fog controller either to wait (without charging) for dispatching to their corresponding customer class (i.e., customers whose trips require the SoC range of this class of vehicles) or to partially charge to serve the subsequent customer class. Vehicles arriving with depleted batteries will be allowed to either partially or fully charge to serve the first or last class of customers, respectively. Clearly, the larger the number of classes, the smaller the SoC increase required for a vehicle to move from one class to the next; and the smaller the charging time needed to make this transition, the lower the burden/requirements on the zone charging resources. On the other hand, given a fixed in-flow rate of vehicles to each city zone, more vehicle/customer classes mean less available in-flow vehicles to each customer class, which may result in longer service delays and even instabilities in their waiting queues.

Given this proposed multi-class system solution, the above questions can then be re-phrased as follows:

(1) What is the minimum number of classes that can fit the available charging resources in a given city zone?

(2) What is the optimal proportion of vehicles from each class to dispatch or partially/fully charge both to maintain the overall system stability and to minimize the maximum and/or average *response time* of the system?

To rigorously address these questions, we first modeled our proposed multi-class charging and dispatching solution as a queuing model and set its parameters.

### 3.3 Queuing Model and System Parameters

We considered one service zone controlled by a fog controller connected to the following:

- the service request apps of customers in the zone
- AEMoD vehicles
- $C$ rapid charging points distributed in the service zone and designed for short-term partial charging, and

- one spacious rapid charging station designed for long-term full charging.

AEMoD vehicles enter the service in this zone after dropping off their latest customers in it. Their detection as free vehicles by the zone's controller can therefore be modeled as a Poisson process with rate $\lambda_v$. Customers request service from the system according to a Poisson process. Customers are classified into $n$ classes on the basis of the ascending order of their required trip distance, and vehicles are classified according to the corresponding SoC to cover this distance. From the thinning property of Poisson processes, the arrival process of Class $i$ customers and vehicles, $i \in \{0,\dots,n\}$ are both independent Poisson processes with rates $\lambda_c^{(i)}$ and $\lambda_v p_i$, where $p_i$ is the probability that the SoC of a vehicle arriving to the system belongs to Class $i$. Note that $p_0$ is the probability that a vehicle will arrive with a depleted battery and will therefore not be able to serve immediately. Consequently, $\lambda_c^{(0)} = 0$, as no customer will request a vehicle that cannot travel any distance. On the other hand, $p_n$ is also equal to 0 because no vehicle can arrive to the system fully charged, as it has just finished a previous trip.

Upon arrival, each vehicle of Class $i$, $i \in \{1,\dots,n-1\}$, will park anywhere in the zone until it is called by the fog controller to either (1) serve a customer from Class $i$ with probability $q_i$; or (2) partially charge up to the SoC of Class $i + 1$ at any of the $C$ charging points (whenever any of them become free), with probability $q_i = 1 - q_i$, before parking again to wait to serve a customer from Class $i + 1$. As for Class 0 vehicles that are incapable of serving before charging, they will be directed to either fully charge at the central charging station with probability $q_0$, or partially charge at one of the $C$ charging points with probability $q_0 = 1 - q_0$. In the first case, after charging the vehicle will wait to serve customers of Class $n$, and in the last case, the vehicle will wait to serve customers of Class 1.

**Table 3-1** List of system and decision parameters

| Variables | Definition |
|---|---|
| $\lambda_v$ | Total arrival rate of vehicles |
| $p_i$ | Probability of arrival of a vehicle from Class $i$ |
| $q_0$ | Probability that a battery-depleted vehicle partially charges |
| $\overline{q}_0$ | Probability that a battery-depleted vehicle fully charges |
| $q_i, i \neq 0$ | Probability that a vehicle in Class $i$ is directly dispatched |
| $\overline{q}_i, i \neq 0$ | Probability that a vehicle in Class $i$ partially charges |
| $\mu_c$ | Service rate of fully charging a battery-depleted vehicle |
| $\lambda_v^{(i)}$ | Arrival rate of vehicles of Class $i$ |
| $\lambda_c^{(i)}$ | Arrival rate of customers served by Class $i$'s vehicles |
| $C$ | No. of distributed charging points in the service zone of the fog controller |

The full charging time of a vehicle with a depleted battery is assumed to be exponentially distributed with rate $\mu_c$. Given uniform SoC quantization among the $n$ vehicle classes, the partial charging time can then be modeled as an exponential random variable with rate $n\mu_c$. Note that the larger rate of the partial charging process is not due to a speed-up in the charging process but rather to the reduced time of partially charging. Exponentially distributed charging times for charging electric vehicles have been widely used in the literature [Liang et al. 2014; Zhang et al. 2016a] to model randomness in the charging durations of different battery sizes. The customers belonging to Class $i$, arriving at rate $\lambda_c^{(i)}$, will be served at a rate of $\lambda_v^{(i)}$, which includes the arrival rates of vehicles that (1) arrived to the zone with an SoC belonging to Class $i$ and were directed to wait to serve Class $i$ customers; or (2) arrived to the zone with an SoC belonging to Class $i-1$ and were directed to partially charge to be able to serve Class $i$ customers.

Given the above description and modeling of variables, the entire zone dynamics can thus be modeled by the queuing system depicted in figure 3-2. This system includes $n$ M/M/1 queues for the $n$ classes of customer service, one M/M/1 queue for the charging station, and one M/M/C queue representing the partial charging process at the $C$ charging points.

Having defined the queuing model for the proposed multiclass dispatching and charging system in a city zone, the rest of this report focuses on addressing the two limitation discussed in Chapter 2. Chapter 3 discusses the stability conditions of the system and minimum number of required classes to cope with the charging resources in any arbitrary city zone. The maximum and average response time minimization problems are then formulated and analytically solved in chapters 4 and 5, respectively.

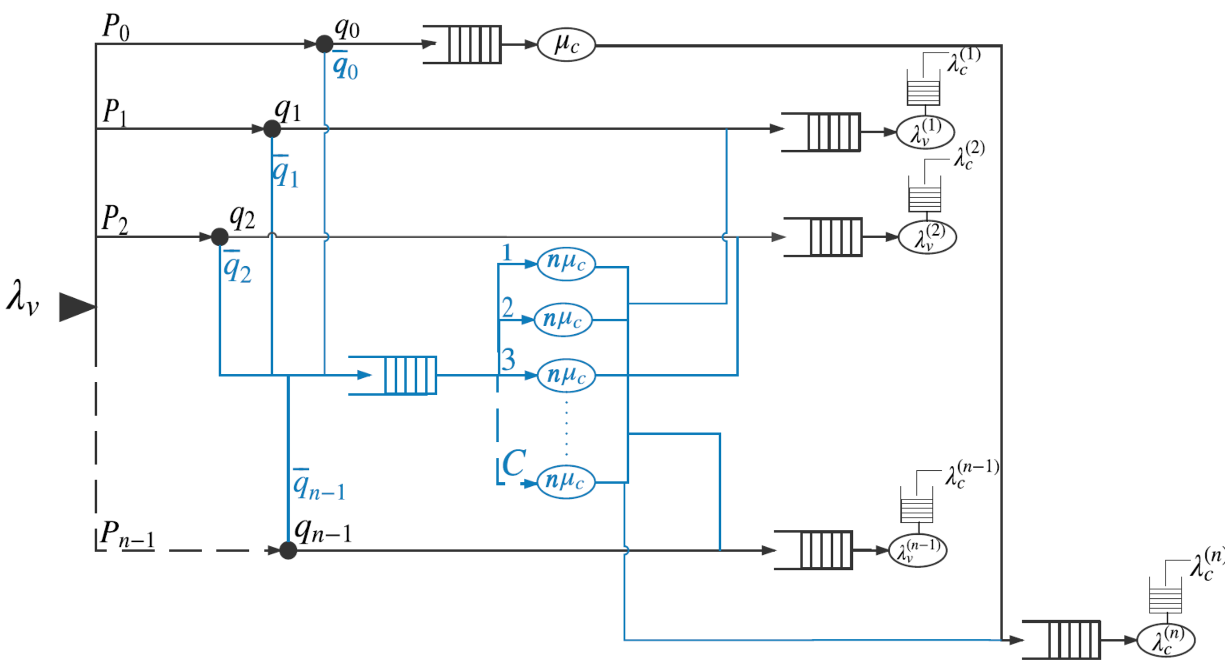


**Figure 3-2** Joint dispatching and partially/fully charging model, abstracting an AEMoD system in one service zone.

## Chapter 4 System Stability Conditions

This chapter discusses how we first deduced the stability conditions of our proposed multi-class dispatching and charging system, using the basic laws of queuing theory. We then derived an expression for the lower bound on the number $n$ of needed classes that fit the charging capabilities of any arbitrary service zone. Each of the $n$ classes of customers is served by a separate queue of vehicles, with $\lambda_v^{(i)}$ being the arrival rate of the vehicles that are available to serve the customers of the $i$-th class. Consequently, it is the service rate of the customers in the $i$-th arrival queues. We can thus deduce the following from figure 3-2 and the system model in the previous chaptee:

$$\lambda_v^{(i)} = \lambda_v \left(p_{i-1}\overline{q}_{i-1} + p_i q_i\right) \quad i = 1, \ldots, n-1$$
$$\lambda_v^{(n)} = \lambda_v \left(p_{n-1}\overline{q}_{n-1} + p_0 q_0\right) \qquad (1)$$

Because we know $q_i$ $q_i$ 1, we substitute $q_i$ by 1 $q_i$ in order to have a system with $n$

$$\lambda_v^{(i)} = \lambda_v \left(p_{i-1} - p_{i-1}q_{i-1} + p_i q_i\right) \quad i = 1, \ldots, n-1$$
$$\lambda_v^{(n)} = \lambda_v \left(p_{n-1} - p_{n-1}q_{n-1} + p_0 q_0\right) \qquad (2)$$

From the well-known stability condition of an M/M/1 queue [Papoulis and Pillai 2002; Leon-Garcia 2008], we have the following:

$$\lambda_v^{(i)} > \lambda_c^{(i)} \quad i = 1, \ldots, n \qquad (3)$$

Before reaching the customer service queues, vehicles will go through a decision step regarding whether to enter these queues immediately or to first partially/fully charge. The stability of the charging queues should be guaranteed to ensure the global stability of the entire system at the steady state. From the model described in the previous chapter and by the well-known stability conditions of M/M/C and M/M/1 queues [Papoulis and Pillai 2002; Leon-Garcia 2008], we have the following stability constraints on the $C$ charging points and central charging station queues, respectively:

$$\sum_{i=0}^{n-1} \lambda_v (p_i - p_i q_i) < C (n\mu_c)$$

$$\lambda_v p_0 q_0 < \mu_c \tag{4}$$

The following lemma illustrates the lower bound on the average in-flow rate of vehicles for a given service zone, given its rate of customer demands on AEMoD services.

*Lemma 1:* For the stability of the entire system, the in-flow rate of vehicles to a given service zone should be more than the total arrival rate of customers belonging to all the classes. In other words,

$$\sum_{i=1}^{n} \lambda_c^{(i)} < \lambda_v \tag{5}$$

Furthermore, the following lemma establishes a lower bound on the number of classes $n$, given the arrival rate of the vehicles $\lambda_v$, the full charging rate $\mu_c$, and the number $C$ of partial charging points.

*Lemma 2:* To guarantee the stability of the charging queues, the number of classes $n$ in the system must obey the following inequality:

$$n > \frac{\lambda_v}{C\mu_c} - \frac{1}{C} \tag{6}$$

## Chapter 5 Maximum Response Time Optimization

The goal of this work was to minimize the maximum expected response time across all system classes.

### 5.1 Problem Formulation

The expected response time of any class was defined as the expected duration between any customers entering a request and a vehicle being dispatched to serve them. From the basic M/M/1 queue analysis of the *i*-th customer class, the expression of this expected response time for the *i*-th class can be expressed as follows:

$$\frac{1}{\lambda_V^{(i)} - \lambda_c^{(i)}} \quad i = 1, \ldots, n \tag{7}$$

Consequently, the maximum of the expected response times across all *n* classes of the system can be expressed as follows:

$$\max_{i \in \{1,\ldots,n\}} \left\{ \frac{1}{\lambda_V^{(i)} - \lambda_c^{(i)}} \right\} \tag{8}$$

It is obvious that the system class with the maximum expected response time is the one that has the minimum expected response rate. In other words, we have

$$\arg \max_{i \in \{1,\ldots,n\}} \left\{ \frac{1}{\lambda_V^{(i)} - \lambda_c^{(i)}} \right\} = \arg \min_{i \in \{1,\ldots,n\}} \left\{ \lambda_V^{(i)} - \lambda_c^{(i)} \right\} \tag{9}$$

Consequently, minimizing the maximum expected response time across all classes is equivalent to maximizing the minimum expected response rate. By using the epigraph form [Boyd and Vandenberghe 2015] of the latter problem, we get the following stochastic optimization problem:

$$\max_{q_0, q_1, \ldots, q_{n-1}} \quad R \tag{10a}$$

$$\text{s.t. } \lambda_V (p_{i-1} - p_{i-1} q_{i-1} + p_i q_i) - \lambda_c^{(i)} \geq R, \quad i = 1, \ldots, n-1 \tag{10b}$$

$$\lambda_V (p_{n-1} - p_{n-1} q_{n-1} + p_0 q_0) - \lambda_c^{(n)} \geq R \tag{10c}$$

$$\sum_{i=0}^{n-1} \lambda_V (p_i - p_i q_i) < C(n\mu_c) \tag{10d}$$

$$\lambda_V p_0 q_0 < \mu_c \tag{10e}$$

$$\sum_{i=0}^{n-1} p_i = 1, \quad 0 \leq p_i \leq 1 \quad i = 0, \ldots, n-1 \tag{10f}$$

$$0 \leq q_i \leq 1 \quad i = 0, \ldots, n-1 \tag{10g}$$

$$R > 0 \tag{10h}$$

The $n$ constraints in (10b) and (10c) represent the epigraph form constraints on the original objective function in the right hand side of (9), after separation [Boyd and Vandenberghe 2015] and substitution of every $\lambda_v^{(i)}$ by its expansion form in (2). The constraints in (10d) and (10e) represent the stability conditions on charging queues. The constraints in (10f) and (10g) are the axiomatic constraints on probabilities (i.e., values being between 0 and 1, and sum equal to 1). Finally, constraint (10h) is a strict positivity constraint on the minimum expected response rate, which also guarantees the stability of customer queues when combined with (10b) and (10c). Indeed, if $R$ is strictly positive, then this guarantees that that the stability conditions in (3) will hold with certainty. Clearly, the above problem is a linear program with linear constraints, which can be solved analytically by using Lagrangian analysis. This is the focus of the next subsection.

5.2 Optimal Dispatching and Charging Decisions

The problem in (10) is a convex optimization problem with a second order differentiable objective and constraint functions that satisfy Slater's condition. Consequently, the KKT conditions provide necessary and sufficient conditions for optimality. Therefore, applying the KKT conditions to the constraints of the problem and the gradient of the Lagrangian function allows us to find the analytical solution of the decisions $q_i$. The Lagrangian function associated with the optimization problem in (10) is given by the following expression:

$$
\begin{aligned}
&L\,(\mathbf{q}, R, \alpha, \beta, \gamma, \omega) \\
&= -R + \sum_{i=1}^{n-1} \alpha_i \left(\lambda_V \left(p_{i-1} q_{i-1} - p_i q_i\right)\right. \\
&\quad \left. + R - \lambda_V p_{i-1} + \lambda_c^{(i)}\right) + \alpha_n \left(\lambda_V \left(p_{n-1} q_{n-1} - p_0 q_0\right)\right. \\
&\quad \left. + R - \lambda_V p_{n-1} + \lambda_c^{(n)}\right) + \beta_0 \left(\sum_{i=0}^{n-1} \lambda_V \left(p_i - p_i q_i\right) - C\,(n\mu_c)\right) \\
&\quad + \beta_1 \left(\lambda_V p_0 q_0 - \mu_c\right) + \sum_{i=0}^{n-1} \gamma_i \left(q_i - 1\right) - \sum_{i=0}^{n-1} \omega_i q_i + \omega_n R
\end{aligned}
\tag{11}
$$

where $\mathbf{q}$ is the vector of dispatching decisions (i.e., $\mathbf{q} = q_0,\ldots,q_{n-1}$ ), and where

- $\alpha = [\alpha_i]$, such that $\alpha_i$ is the associated Lagrange multiplier to the $i$-th customer queues inequality.
- $\beta = [\beta_i]$, such that $\beta i$ is the associated Lagrange multiplier to the $i$-th charging queues inequality.
- $\gamma = \gamma_i$ , such that $\gamma_i$ is the associated Lagrange multiplier to the $i$-th upper bound inequality.
- $\omega = [\omega_i]$, such that $\omega_i$ is the associated Lagrange multiplier to the $i$-th lower bound inequality.

By applying the KKT conditions to the equality and inequality constraints, the following theorem illustrates the optimal solution of the problem in (10).

*Theorem 1:* The optimal charging/dispatching decisions of the optimization problem in (10) can be expressed as follows:

$$
q_0^* = \begin{cases} 0 & \text{if } \alpha_1^* > \alpha_n^* \\ 1 & \text{if } \alpha_1^* < \alpha_n^* \end{cases}
$$

$$
q_i^* = \begin{cases} 0 & \text{if } \alpha_{i+1}^* > \alpha_i^* \\ 1 & \text{if } \alpha_{i+1}^* < \alpha_i^* \end{cases} \quad i = 1, \dots, n-1
$$

$$
\text{if } \alpha_1^* = \alpha_n^* \neq 0 \begin{cases} q_1^* = \dfrac{p_0 q_0^*}{p_1} - \dfrac{\lambda_v p_0 - \lambda_c^{(1)} - R^*}{\lambda_v p_1} \\ q_{n-1}^* = \dfrac{p_0 q_0^*}{p_{n-1}} - \dfrac{\lambda_v p_0 - \lambda_c^{(n)} - R^*}{\lambda_v p_{n-1}} \end{cases}
$$

$$
\text{if } \alpha_{i+1}^* = \alpha_i^* \neq 0 \begin{cases} q_i^* = \dfrac{p_{i-1} q_{i-1}^*}{p_i} - \dfrac{\lambda_v p_{i-1} - \lambda_c^{(i)} - R^*}{\lambda_v p_i} \\ q_{i+1}^* = \dfrac{p_i q_i^*}{p_{i+1}} - \dfrac{\lambda_v p_i - \lambda_c^{(i+1)} - R^*}{\lambda_v p_{i+1}} \end{cases}
$$

$$
i = 1, \dots, n-1 \tag{12}
$$

5.3 Maximum Expected Response Time

Again, because the problem in (10) is convex with differentiable objective and constraint functions, then strong duality holds, which implies that the solutions to the primal and dual problems are identical. By solving the dual problem, we can express the optimal value of the maximum expected response time as the reciprocal of the minimum expected response rate of the system. The latter is characterized by the following theorem.

*Theorem 2:* The minimum expected response rate $R*$of the entire system can be expressed as follows:

$$
R^* = \sum_{i=1}^{n} \left( \lambda_v p_{i-1} - \lambda_c^{(i)} \right) \alpha_i^* + \sum_{i=0}^{n-1} \gamma_i^* \tag{13}
$$

5.4 Problem Complexity

This problem, is formulated as a linear program. Theoretically, the complexity of solving algorithms for linear programs (and thus the size of stored information to solve them) has been proved to be $(n^{\frac{3}{4}} \log(\frac{n}{\epsilon}))$ per iteration [Liu and Fan 2010], with *n* as the number of variables and epsilon as a stopping criteria. This complexity expression clearly justifies our suggestion to solve the problem for each zone locally. Indeed, the complexity heavily depends on the number of parameters, which is in this case related to the number of classes *n* defined in the system. Lemma 2 clearly established that *n* scales with the vehicle rate (number of vehicles becoming available for service per unit of time). Clearly, the smaller *n* is, the simpler the problem is.

## Chapter 6 Average Response Time Optimization

The goal of of the work presented in this chapter was to minimize the average expected response time for the system's overall customer classes.

### 6.1 Problem Formulation

As stated earlier, the expected response time for each of the classes in the system is expressed as in (7). Because our system is divided into $n$ classes, the average expected response time across the different classes is expressed as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_v^{(i)}-\lambda_c^{(i)}} \qquad (14)$$

Therefore, minimizing the average expected response time across all the classes of the system, while obeying its stability conditions, can be formulated by the following problem.

$$\underset{q_0,q_1,\ldots,q_{n-1}}{\text{minimize}}\ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_v^{(i)}-\lambda_c^{(i)}} \qquad (15a)$$

$$\text{s.t. } \lambda_v(p_{i-1}-p_{i-1}q_{i-1}+p_iq_i)-\lambda_c^{(i)}>0,$$

$$i=1,\ldots,n-1 \qquad (15b)$$

$$\lambda_v(p_{n-1}-p_{n-1}q_{n-1}+p_0q_0)-\lambda_c^{(n)}>0 \qquad (15c)$$

$$\sum_{i=0}^{n-1}\lambda_v(p_i-p_iq_i)<C(n\mu_c) \qquad (15d)$$

$$\lambda_v p_0 q_0<\mu_c \qquad (15e)$$

$$\sum_{i=0}^{n-1}p_i=1,\quad 0\le p_i\le 1\quad i=0,\ldots,n-1 \qquad (15f)$$

$$0\le q_i\le 1\quad i=0,\ldots,n-1 \qquad (15g)$$

The $n$ constraints in (15b) and (15c) represent the stability constraints in (3) and substituting every $\lambda^{(i)}_v$ by its expansion form in (2). The constraints in (15d) and (15e) represent the stability conditions for charging queues. The constraints in (15f) and (15g) are the axiomatic constraints on probabilities (i.e., values being between 0 and 1, and sum equal to 1).

The above constraints are all linear, but the objective function is obviously not. Nonetheless, the following lemma proves that the optimization problem we have is convex which allows us to find an absolute exact solution analytically and numerically.

*Lemma 3:* Defining the function $f$ as follows:

$$f(q_0, q_1, \ldots, q_{n-1}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_v^{(i)} - \lambda_c^{(i)}} \qquad (16)$$

such that $\lambda_v^{(i)}$ and $\lambda_c^{(i)}$ are defined in (2) and (3), and the function $f$ is convex over the variables $q_0, q_1, \ldots, q_{n-1}$.

Consequently, the problem in (15) is a convex problem with linear constraints, which can be solved analytically by using Lagrangian analysis. This is the focus of the next subsection.

### 6.2 Optimal Dispatching and Charging Decision

As proved above, the problem in (15) is a convex optimization problem with a second order differentiable objective function and constraints that satisfy Slater's condition. Like the the approach presented in Chapter 4, we can apply the KKT conditions to the constraints of the problem and the gradient of the Lagrangian function to find the analytical solution of the decisions $q_i$. The Lagrangian function associated with the optimization problem in (15) is given by the following expression:

$$
\begin{aligned}
&L(\mathbf{q}, \alpha, \beta, \gamma, \omega) \\
&= \frac{1}{n} \sum_{i=1}^{n-1} \frac{1}{\lambda_v (p_{i-1} - p_{i-1} q_{i-1} + p_i q_i) - \lambda_c^{(i)}} \\
&\quad + \frac{1}{n \left( \lambda_v (p_{n-1} - p_{n-1} q_{n-1} + p_0 q_0) - \lambda_c^{(n)} \right)} \\
&\quad + \sum_{i=1}^{n-1} \alpha_i \left( \lambda_c^{(i)} - \lambda_v (p_{i-1} - p_{i-1} q_{i-1} + p_i q_i) \right) \\
&\quad + \alpha_n \left( \lambda_c^{(n)} - \lambda_v (p_{n-1} - p_{n-1} q_{n-1} + p_0 q_0) \right) \\
&\quad + \beta_0 \left( \sum_{i=0}^{n-1} \lambda_v (p_i - p_i q_i) - C(n\mu_c) \right) \\
&\quad + \beta_1 (\lambda_v p_0 q_0 - \mu_c) + \sum_{i=0}^{n-1} \gamma_i (q_i - 1) - \sum_{i=0}^{n-1} \omega_i q_i \qquad (17)
\end{aligned}
$$

where **q** is the vector of dispatching decisions (i.e., $\mathbf{q} = q_0 =, \ldots, q_{n-1}$ ), and where α = [$\alpha_i$], β = [$\beta_i$], γ = $\gamma_i$ , and ω [$\omega_i$] are the vectors of the Lagrange multipliers associated with the inequalities constraints of problem (15) and defined in the same way as explained in Chapter 4.

By applying the KKT conditions to the equality and inequality constraints, the following theorem illustrates the optimal solution of the problem in (15).

*Theorem 3:* The optimal charging/dispatching decisions of the optimization problem in (15) can be expressed as follows:

$$
q_i^* = \begin{cases} 0 & \text{if } \omega_i^* \neq 0 \\ 1 & \text{if } \gamma_i^* \neq 1 \end{cases} \quad i = 0, \ldots, n-1 \qquad (18)
$$

Otherwise, we have the following:

$$
\begin{aligned}
q_0^* &= \frac{\lambda_v \left(p_0 + p_1 q_1^* - p_{n-1} + p_{n-1} q_{n-1}^*\right) - \lambda_c^{(1)} + \lambda_c^{(n)}}{2\lambda_v p_0} \\
q_i^* &= \frac{\lambda_v \left(p_i + p_{i+1} q_{i1}^* - p_{i-1} + p_{i-1} q_{i-1}^*\right) - \lambda_c^{(i+1)} + \lambda_c^{(i)}}{2\lambda_v p_i} \\
& i = 1, \ldots, n-2 \\
q_{n-1}^* &= \frac{\lambda_v \left(p_{n-1} + p_0 q_0^* - p_{n-2} + p_{n-2} q_{n-2}^*\right) - \lambda_c^{(n)} + \lambda_c^{(n-1)}}{2\lambda_v p_{n-1}}
\end{aligned}
\tag{19}
$$

## 6.3 Problem Complexity

This problem is formulated as a convex, non-linear problem. Theoretically, the interior point method complexity for convex, non-linear programs (and thus the size of information stored to solve them) depends on the number of constraints of the problem *m,* This is expressed as $Q(\sqrt{m}\log\left(\frac{m}{\epsilon}\right))$ according to Boyd and Vandenberghe (2015), epsilon being a stopping criterion. This complexity of expression clearly justifies our suggestion to solve the problem for each zone locally. The complexity depends on the number of constraints, which depends strongly on the number of classes *n* defined in the system. Lemma 2 clearly established 2 that *n* scales with the vehicle rate (number of vehicles becoming available for service per unit of time). Clearly, the smaller *n* is, the simpler the problem is.

## Chapter 7 Simulation Results

this chapter discusses testing of the merits of our proposed scheme with extensive simulations. The metrics used to evaluate these merits were the maximum and average expected response times of the different classes. For all the simulations, the full-charging rate of a vehicle was set to $\mu_c = 0.033$ mins$^{-1}$, and the number of charging points was $C = 40$.

For the optimized maximum time solution (figure 7-1) and the average response time solution (figure 7-2), the simulations illustrated the interplay of the effect of increasing the number of classes $\lambda_v$ *and* $\sum_{i=1}^{n} \lambda_c^{(i)}$ established in Lemma 1, and effect of increasing the number of classes *n* beyond its strict lower bound introduced in Lemma 2. The figures depict the maximum and average expected response times for different values of $\sum_{i=1}^{n} \lambda_c^{(i)}$ while setting $\lambda_v$ to 15 min$^{-1}$. For this setting, $n = 12$ is the smallest number of classes that satisfies the stability condition in Lemma 2. It is easy to see that the response times for all values of *n* increase dramatically when $\sum_{i=1}^{n} \lambda_c^{(i)}$ approaches $\lambda_v$, thus bringing the system closer to the stability limit established in Lemma 1. As also expected, the figures clearly show that further increasing *n* beyond its stability lower bound increases both the maximum and average response times. As explained earlier, this effect occurs because of the reduced number of available vehicles to each customer class as *n* grows, given the fixed $\lambda_v$. We thus firmly conclude that the optimal number of classes is the smallest value satisfying Lemma 2:

$$n^* = \begin{cases} \dfrac{\lambda_v}{C\mu_c} - \dfrac{1}{C} + 1 & \text{if } \dfrac{\lambda_v}{C\mu_c} - \dfrac{1}{C} \text{ is integer} \\ \left\lceil \dfrac{\lambda_v}{C\mu_c} - \dfrac{1}{C} \right\rceil & \text{Otherwise} \end{cases} \tag{20}$$

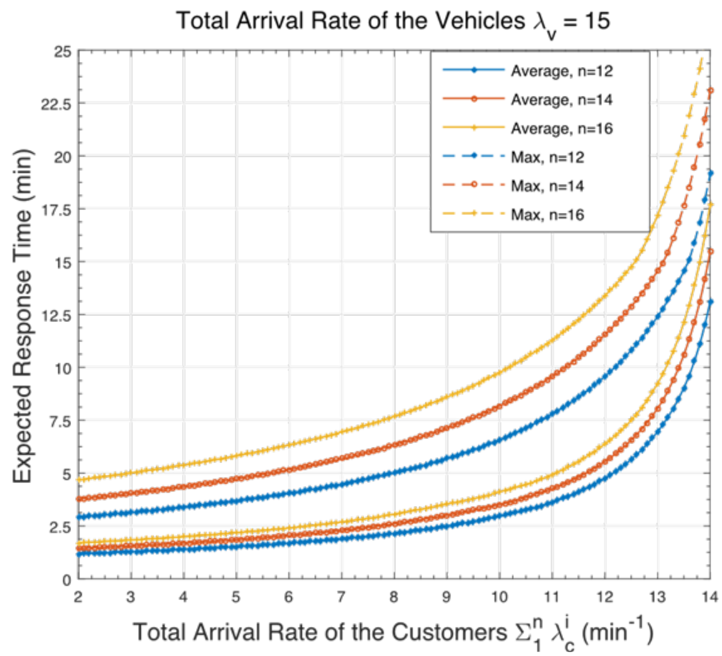


**Figure 7-1** Expected response times using the maximum response time optimization solution

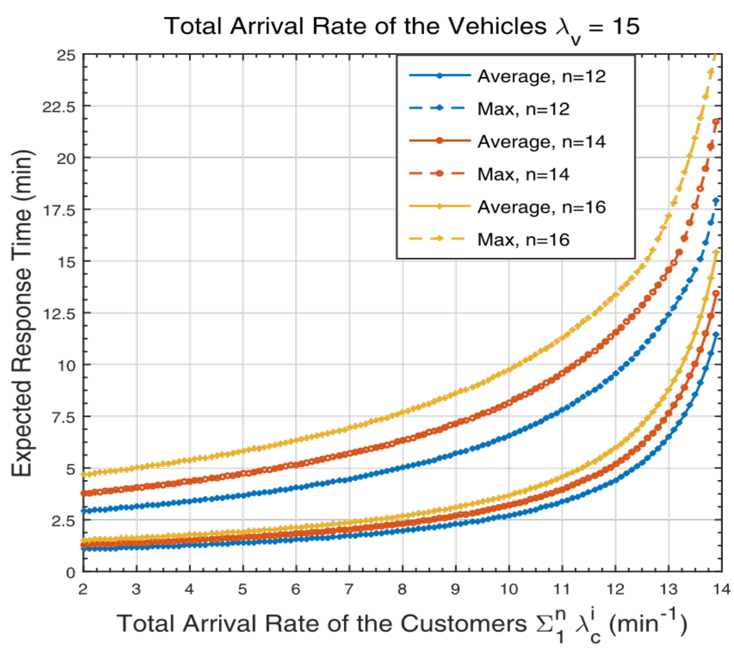


**Figure 7-2** Expected response times using the average response time optimization solution for different

For the maximum and average response time optimization solutions, figures 7-3 and 7-4, respectively, depict the maximum and average expected response time performances for different distributions of vehicle SoC and customer trip distances, given $\lambda_v = 8$ and therefore $n^* = 7$. By decreasing vehicle SoC distribution, we mean that the probability of an arriving vehicle being in class $i$ SoC is lower than that of it being in class $i - 1$ SoC $\forall i \in \{2,\ldots,n\}$. We can infer from both figures that both the maximum and average response times for the Gaussian distributions of trip distances and both Gaussian and decreasing SoCs are the lowest and exhibit the least response time variance. Fortunately, these are the most realistic distributions for both variables. This is justified by the fact that vehicles arrive to the system after trips of different distances, which makes their SoC either Gaussian or slightly decreasing. Likewise, the number of customers requiring mid-length distances is usually larger than the numbers requiring very short and very long distances.

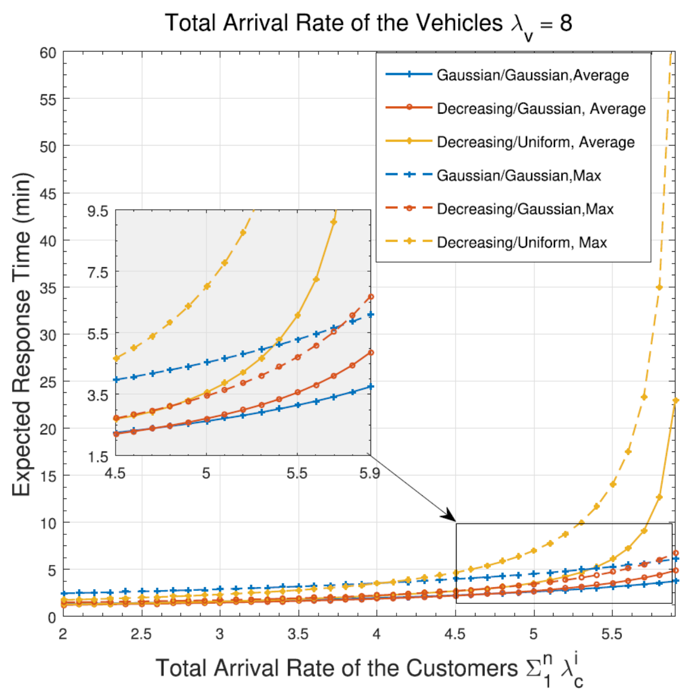


**Figure 7-3** Effect of different customer and SoC distributions on the maximum response time optimization solution.

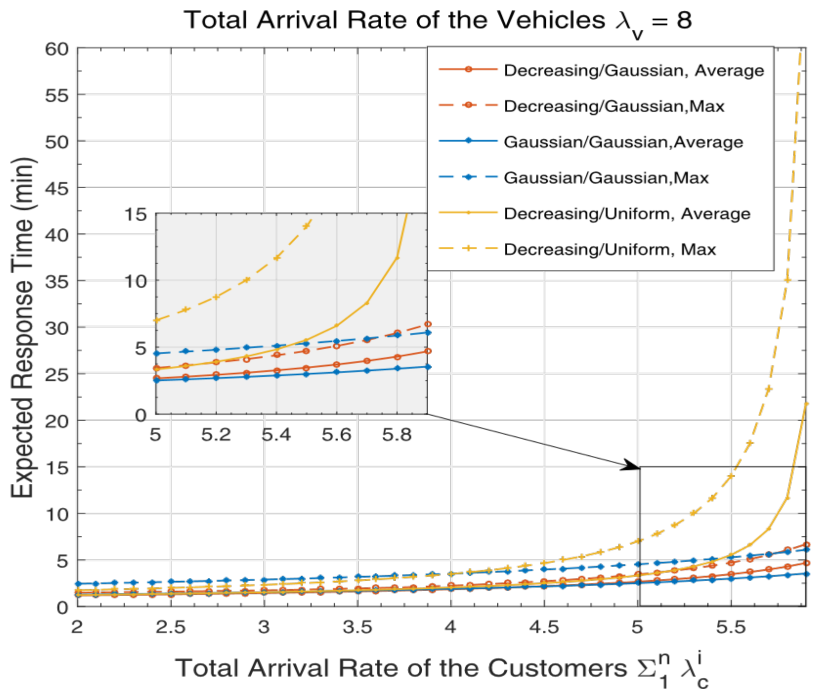


**Figure 7-4** Effect of different customer and SoC distributions on the average response time optimization solution

For the maximum and average response time optimization solutions, figures 7-5 and 7-6, respectively, compare the maximum and average expected response time performances with $\sum_i^n \lambda_c^{(i)}$, for the different decision approaches discussed in Chapter 4, namely the always partially charge decision (i.e., $q_i = 0 \ \forall i$) and the equal split decision (i.e., $q_i = 0.5 \ \forall i$), for $\lambda_v = 8$ and therefore $n = 7$. The latter two schemes represent non-optimized policies, in which each vehicle takes its own fixed action irrespective of system parameters. The figures clearly show superior maximum and average performances for our derived optimal policies in comparison to the other two policies. This is especially true as $\sum_i^n \lambda_c^{(i)}$ gets closer to $\lambda_v$, which represents the most properly engineered scenarios (as large differences between these two quantities result in very low utilization). Gains of 13.3 percent in the average performance and and 21.3 percent in the maximum performance can be noticed in comparison to the always charge policy. This demonstrates the importance of our proposed scheme in achieving lower response times and thus better customer satisfaction.

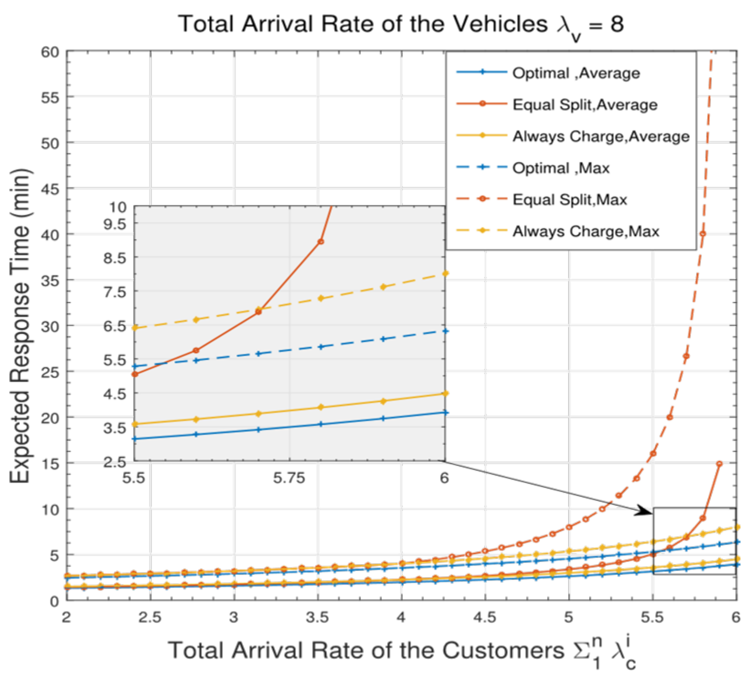


**Figure 7-5** Comparison of the maximum response time optimization solution to non-optimized policies.

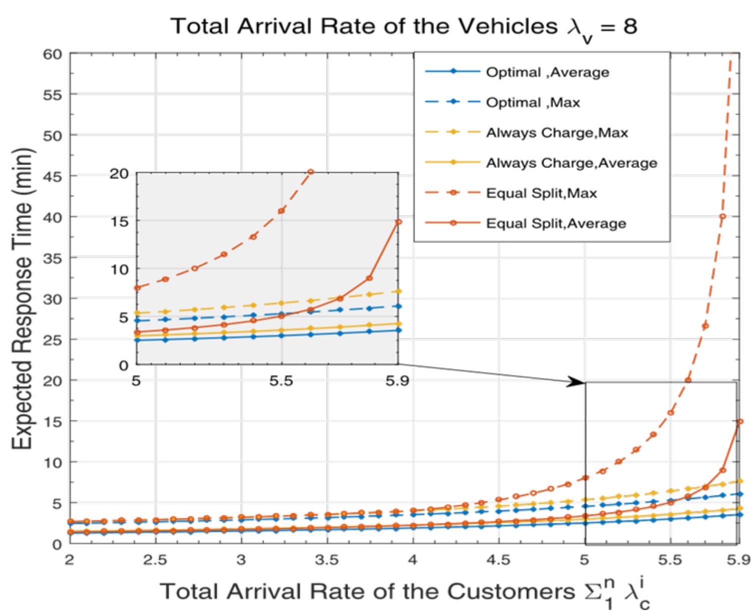


**Figure 7-6** Comparison of the average response time optimization solution to non-optimized policies.

Figure 7-7 compares the maximum and average expected response time performance produced by the maximum and average response time optimization solutions introduced in chapters 4 and 5, respectively, for different values of $\sum_i^n \lambda_c^{(i)}$ while setting $\lambda_v$ to 15 min$^{-1}$ (i.e., $n$ = 12). We can easily see that the maximum expected response times achieved by both solutions are the same. On the other hand, the average expected response time given by the average solution is slightly lower than that of the maximum solution. These results suggest that the variance in performance achieved by both solutions is negligible. Consequently, the one that is obtained with less complexity should be used to almost satisfy the minimum value for both metrics. We know from chapters 4 and 5 that the maximum solution is obtained by solving a linear optimization and the average solution is obtained by solving a convex yet non-linear optimization. It is well known that solving the latter requires more computation that the former. For example, when interior point methods are used, the maximum number of iterations for the maximum solution is 10 and that for the average solution is 25. Therefore, the maximum solution is recommended for use in future AEMoD systems because of its lower complexity and its negligible degradation in its average response time performance in comparison to the average solution.

Because of the significant differences between the related works and our proposed model in the factors considered to affect customer waiting times, we decided that conducting a numerical comparison would be inapplicable and non-meaningful.

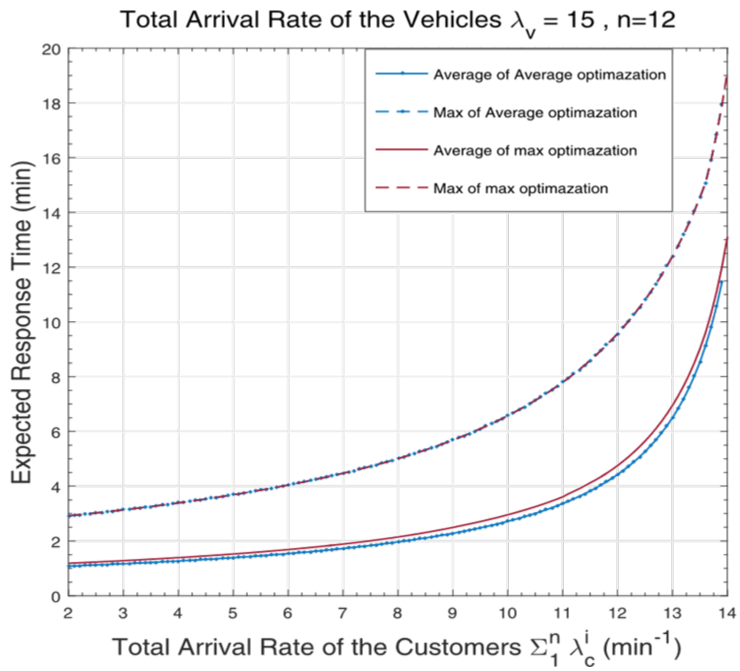


**Figure 7-7** Comparison between the maximum minimization and the average minimization of the expected response time.

# Chapter 8 Summary and Conclusions

In this project, we proposed solutions to the computational and charging bottlenecks that threaten the success of AEMoD systems in attracting a large number of customers and solving private urban transportation problems. We also incorporated user-defined special needs in the optimization process to further accommodate customers with cognitive and physical disabilities in AEMoD operations. The computational bottleneck can be resolved by employing a fog-based architecture to distribute the optimization loads over different service zones and to reduce communication delays, while matching the nature of the dispatching/charging processes of AEMoD vehicles. We also proposed a multi-class dispatching and charging scheme to guarantee suitability between the vehicle charge requirements of customers' trips with special needs and the available resources in each city zone. To efficiently engineer this multiclass solution, we developed a queuing model, derived its stability conditions, and characterized the optimal number of classes to both minimize response time and match zone charging resources. We then formulated the problem of optimizing the proportions of vehicles from each class that would partially/fully charge or directly serve customers as a stochastic linear optimization problem to minimize the maximum expected system response time and as a convex but non-inear optimization problems to minimize the average expected system response time. The optimal decisions for both problems were analytically derived by using Lagrangian analysis. Simulation results demonstrated both the merits of our proposed optimal decision scheme in comparison to typical non-optimized schemes and its performance for different distributions of vehicle SoC and customer trip distances. The comparison between the maximum and average problem solutions exhibited negligible variance, which favored the use of the maximum solution because of its lower complexity.

For future work, we will study the problem of maximizing system utilization (i.e., minimizing the required in-flow vehicle rate to each city zone) while satisfying a maximum response time constraint. We will also study scenarios in which non-dispatch vehicles in each class serve either lower customer classes or other city zones.

## References


S. Bansal, M. N. Zeilinger, and C. J. Tomlin, “Plug-and-play model predictive control for electric vehicle charging and voltage control in smart grids,” in Proc. IEEE 53rd Conf. Decis. Control, Dec. 2014, pp. 5894–5900.

S. Boyd and L. Vandenberghe, Convex Optimization. 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2015.

M. Chiang and T. Zhang, “Fog and IoT: An overview of research opportunities,” IEEE Internet Things J., vol. 3, no. 6, pp. 854–864, Dec. 2016.

Cisco, “Fog computing and the Internet of Things: Extend the cloud to where the things are,” Cisco, San Jose, CA, USA, White Paper, 2015. [Online]. Available: http://www.cisco.com/c/dam/en_us/solutions/ trends/iot/docs/computing-overview.pdf

Digitalist Magazine. IoT And Smart Cars: Changing The World For The Better, Aug. 30, 2016. [Online]. Available: http://www.digitalistmag.com/iot/2016/08/30/iot-smart-connected-carswillchange-world-04422640

The Emissions Gap Report 2013—UNEP, U. N. E. Programme, Armidale, NSW, Australia, 2013.

G. P. Fettweis, “The tactile Internet: Applications and challenges,” IEEE Veh. Technol. Mag., vol. 9, no. 1, pp. 64–70, Mar. 2014.

Forbes. (Jun. 13, 2016). The Future Is Now: Smart Cars And IoT In Cities. [Online]. Available: http://www.forbes.com/sites/pikeresearch/2016/06/13/the-future-is-now-smartcars/63c0a25248c9

X. Huang, R. Yu, J. Kang, Y. He, and Y. Zhang, “Exploring mobile edge computing for 5G-enabled software defined vehicular networks,” IEEE Wireless Commun., vol. 24, no. 6, pp. 55–63, Dec. 2017.

S.-W. Kim et al., “Autonomous campus mobility services using driverless taxi,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 12, pp. 3513–3526, Dec. 2017.

C. D. Korkas, S. Baldi, S. Yuan, and E.B. Kosmatopoulos,“An adaptive learning-based approach for nearly optimal dynamic charging of electric vehicle fleets,” IEEE Trans. Intell. Transp. Syst., vol. 19, no. 7, pp. 2066–2075, Jul. 2018.

M. Kumru, E. Debada, L. Makarem, and D. Gillet, “Mobility-on-Demand scenarios relying on lightweight autonomous and connected vehicles for large pedestrian areas and intermodal hubs,” in Proc. 2nd IEEE Int. Conf. Intell. Transp. Eng. (ICITE), Singapore, Sep. 2017, pp. 178–183.

A. Leon-Garcia, Probability, Statistics, and Random Processes for Electrical Engineering, 3rd ed. Upper Saddle River, NJ, USA: PrenticeHall, 2008.

H. Liang, I. Sharma, W. Zhuang, and K. Bhattacharya, “Plug-in electric vehicle charging demand estimation based on queueing network analysis,” in Proc. IEEE PES General Meeting|Conf. Expo., Jul. 2014, pp. 1–5.

L. Liu and L. Fan, “The complexity analysis of an efficient interior-point algorithm for linear optimization,” in Proc. 3rd Int. Joint Conf. Comput. Sci. Optim., 2010, pp. 21–24.

P. Mach and Z. Becvar, “Mobile edge computing: A survey on architecture and computation offloading,” IEEE Commun. Surveys Tuts., vol. 19, no. 3, pp. 1628–1656, 3rd Quart., 2017.

Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, “A survey on mobile edge computing: The communication perspective,” IEEE Commun. Surveys Tuts., vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.

E. K. Markakis, K. Karras, A. Sideris, G. Alexiou, and E. Pallis, “Computing, caching, and communication at the edge: The cornerstone for building a versatile 5G ecosystem,” IEEE Commun. Mag., vol. 55, no. 11, pp. 152–157, Nov. 2017a.

E. K. Markakis et al., “Efficient next generation emergency communications over multi-access edge computing,” IEEE Commun. Mag., vol. 55, no. 11, pp. 92–97, Nov. 2017b.

J. Miller and J. P. How, “Predictive positioning and quality of service ridesharing for campus Mobility on Demand systems,” in Proc. IEEE Int. Conf. Robot. Automat. (ICRA), Singapore, May/Jun. 2017, pp. 1402–1408.

W. J. Mitchell, C. E. Borroni-Bird, and L. D. Burns, Reinventing the Automobile: Personal Urban Mobility for the 21st Century. Cambridge, MA, USA: MIT Press, 2010.

Navigant Research. (2016). Transportation Outlook: 2025 to 2050. [Online]. Available: http://www.navigantresearch.com/ research/transportation-outlook-2025-to-2050

[Online]. Available: https://5g-ppp.eu/wp-content/uploads/2014/02/ 5G-PPP-White-Paper-on-Automotive-Vertical-Sectors.pdf

A. Papoulis and S. Pillai, Probability, Random Variables, and Stochastic Processes, 4th ed. New York, NY, USA: McGraw-Hill, 2002.

M. Patel, Y. Hu, and P. Hédé, “Mobile-edge computing—Introductory technical white paper,” White Paper, ETSI, Sophia Antipolis, France, Sep. 2014.

L. Rao and J. Yao, “SmartCar: Smart charging and driving control for electric vehicles in the smart grid,” in Proc. IEEE Global Commun. Conf., Dec. 2014, pp. 2709–2714.

E. S. Rigas, S. D. Ramchurn, and N. Bassiliades, “Managing electric vehicles in the smart grid using artificial intelligence: A survey,” IEEE Trans. Intell. Transp. Syst., vol. 16, no. 4, pp. 1619–1635, Aug. 2015.

M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, “The case for VM-based cloudlets in mobile computing,” IEEE Pervasive Comput., vol. 8, no. 4, pp. 14–23, Oct./Dec. 2009.

D. Schrank, B. Eisele, and T. Lomax, “TTIs 2012 urban mobility report,” Texas A&M Transp. Inst., College Station, TX, USA, Tech. Rep. 498 U.S. Urban Areas, 2012.

A. Somov and R. Giaffreda, “Powering IoT devices: Technologies and opportunities,” IEEE IoT Newslett., Nov. 2015. [Online]. Available: http://iot.ieee.org/newsletter/november-2015/powering-iot-devicestechnologies-and-opportunities.html

W. Tang and Y. J. Zhang, “A model predictive control approach for low-complexity electric vehicle charging scheduling: Optimality and scalability,” IEEE Trans. Power Syst., vol. 32, no. 2, pp. 1050–1063, Mar. 2017.

T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, “Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges,” IEEE Commun. Mag., vol. 55, no. 4, pp. 54–61, Apr. 2017. [33] “5G automotive vision,” 5GPPP, White Paper, Oct. 2015.

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. Autom. Control, vol. 58, no. 9, pp. 2261–2276, Sep. 2013.

W. Tushar, C. Yuen, S. Huang, D. B. Smith, and H. V. Poor, “Cost minimization of charging stations with photovoltaics: An approach with EV classification,” IEEE Trans. Intell. Transp. Syst., vol. 17, no. 1, pp. 156–169, Jan. 2016.

U. E. P. Agency. (2014). Greenhouse Gas Equivalencies Calculator. [Online]. Available: http://www.epa.gov/cleanenergy/energyresources/refs.html

R. Zhang and M. Pavone, “Control of robotic Mobility-on-Demand systems: A queueing-theoretical perspective,” Int. J. Robot. Res., vol. 35, nos. 1–3, pp. 186–203, 2016.

R. Zhang, K. Spieser, E. Frazzoli, and M. Pavone, “Models, algorithms, and evaluation for autonomous Mobility-on-Demand systems,” in Proc. Amer. Control Conf., Chicago, IL, USA, 2015, pp. 2573–2587.

K. Zhang, Y. Mao, S. Leng, Y. Zhang, S. Gjessing, and D. H. K. Tsang, “Platoon-based electric vehicles charging with renewable energy supply: A queuing analytical model,” in Proc. IEEE Int. Conf. Commun. (ICC), May 2016a, pp. 1–6.

R. Zhang, F. Rossi, and M. Pavone, “Model predictive control of autonomous Mobility-on-Demand systems,” in Proc. IEEE Int. Conf. Robot. Automat., Stockholm, Sweden, May 2016b, pp. 1382–1389.

S. Belakaria, M. Ammous, S. Sorour, and A. Abdel-Rahim, “A multiclass dispatching and charging scheme for autonomous electric mobility on-demand,” in Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall), Toronto, ON, Canada, Sep. 2017, pp. 1–5.

International Energy Outlook 2013, U. E. I. Admin., Santa Ana, CA, USA, 2013.

Santos, N. McGuckin, H. Y. Nakamoto, D. Gray, and S. Liss, “Summary of travel trends: 2009 national household travel survey,” National Household Travel Survey, Washington, DC, USA, Tech. Rep. FHWA-PL-ll-022, 2011.