Multi-Class Management with Sub-Class Service for Autonomous Electric Mobility On-Demand Systems

Syrine Belakaria*, Mustafa Ammous*, Sameh Sorour* and Ahmed Abdel-Rahim†‡
*Department of Electrical and Computer Engineering, University of Idaho, Moscow, ID, USA
†Department of Civil and Environmental Engineering, University of Idaho, Moscow, ID, USA
‡National Institute for Advanced Transportation Technologies, University of Idaho, Moscow, ID, USA
Email: {bela7898, ammo1375}@vandals.uidaho.edu, {samehsorour, ahmed}@uidaho.edu

***Abstract*—Despite the significant advances in vehicle automation and electrification, the next-decade aspirations for massive deployments of autonomous electric mobility-on-demand (AEMoD) services are still threatened by two major bottlenecks, namely the computational and charging delays. Recent works have suggested the use of fog-based and a multi-class dispatching and charging architectures (classes being defined based on customer trip distance and vehicle charge) to address these two issues, respectively. This paper aims to resolve a clear and unrealistic limitation in these recent works, namely the matched charge-to-trip only service, by enabling sub-class service; i.e., allowing vehicles to serve customer classes with trips needing less charge. A queuing model representing the new multi-class management scheme is first introduced. The stability conditions of the system in a given city zone are then derived. Decisions on the proportions of each class's vehicles to charge and to partially/fully charge, or directly serve customers with sub-class service are then optimized in order to minimize the maximum response time of the system. Results show the merits of our optimized model compared to a previously proposed scheme and other non-optimized policies.**

***Keywords*—Autonomous Mobility On-Demand; Electric Vehicles; Queuing Systems.**

I. Introduction

Urban transportation systems are facing tremendous challenges nowadays due to the dominant dependency and massive increases on private vehicle ownership, which result in dramatic increases in road congestion, parking demand [1], [2], and carbon footprint [3] [4]. These challenges can be mitigated with the significant advances of vehicle electrification, automation, and connectivity. With more than 10 million self-driving cars expected to be on the road by 2025 [5], it is forecasted that vehicle ownership will significantly decline by 2025, as it will be replaced by the novel concept of Autonomous Electric Mobility on-Demand (AEMoD) services [6], [7]. In such system, customers will simply need to press some buttons on an app to promptly get an autonomous electric vehicle transporting them door-to-door, with no pick-up/drop-off and driving responsibilities, no dedicated parking needs, no vehicle insurance and maintenance costs, and extra in-vehicle leisure times. With these qualities, AEMoD systems will succeed in attracting millions of subscribers and providing hassle-free private urban mobility.

Despite the great aspirations for wide AEMoD service deployments by early-to-mid next decade, the timeliness (and thus success) of such service is threatened by two major bottlenecks. First, the expected massive demand of AEMoD services will result in excessive if not prohibitive computational and communication delays if cloud based approaches are employed for the micro-operation of such systems. Moreover, the typical full-battery charging rates of electric vehicles will not be able to cope with the gigantic numbers of vehicles involved in these systems, thus resulting in instabilities and unbounded customer delays. Several recent works Recent works have addressed important problems in AMoD systems by building different operation models for them like a distributed spatially averaged queuing model and a lumped Jackson network model [9] also the system was [13] cast into a closed multi-class BCMP queuing network to solve the routing problem on congested roads. Many key factors were not considered in these works in order to simplify the mathematical resolution. None of these papers considered the computational architecture for massive demands on such services, the vehicle electrification, and the influence of charging limitations on its stability.

In [18] [19], we proposed a fog-based multi-class dispatching and charging solution for AEMoD systems to resolve these two issues. The communication/computation limitations were addressed by exploiting the trendy *fog-based networking and computing architectures* [20], soon-to-be-deployed by cellular operators and communication companies. The privileges brought by this technique [18] will allow handling instantaneous decision making applications such as AEMoD system operations in each city zone in a distributed and accelerated way, given full knowledge about customer requests, vehicle inflow to the zone, their state-of-charge (SoC), and the available charging resources in the service zone. The charging limitations were resolved by proposing a multi-class dispatching and charging model, which assigns to each customer a vehicle that has enough charge to serve him/her without in route charging. In this model, arriving vehicles in each service zone are subdivided into different classes in ascending order of their SoC corresponding to the different customer classes. Each

This study was supported by the Pacific Northwest University Transportation center (Pac trans) project KLK864. The authors would like to thank all who contribute to this study from both funding agencies.

class's vehicles are split between either waiting to dispatch to its corresponding customer class, or partially charge to serve the customers from one-higher class. Vehicles arriving with depleted batteries are allowed to either partially or fully charge. Despite the valuable results given by deriving the optimal split proportions for this model (compared to typical non-optimized splitting approaches), its dispatching process is not realistic. Indeed, vehicles in this model only sever trips fitting their SoC (before or after partially charging), ignoring the fact that they can also serve all customer classes requesting smaller trips (and thus less charge). This is not only impractical, but can also result in battery depletion of all vehicles by the end of the service, which can cause instabilities to the entire system.

In this paper, we aim to address this clear limitation of the previous model, by enabling sub-class dispatching; i.e., allowing proportions of each class vehicles to serve to customers from its class as well as all classes requesting shorter trips). The question now is: *What are the optimal charging and sub-classes dispatching proportions to maintain charging stability and minimize the maximum response time of the system?* To address this question, a queuing model representing the proposed multi-class management with sub-class service scheme is first introduced. The stability conditions of this model. Decisions on both the proportions of each class's vehicles to partially/fully charge vs directly dispatch and their proportions of serving own vs sub-classes are then jointly optimized. Finally, the merits of our proposed optimized decision scheme are tested and compared to the proposed work in [18] [19] as well as several non optimized schemes.

II. System Model

We consider one service zone controlled by a fog controller connected to: (1) the service request apps of customers in the zone; (2) the AEMoD vehicles; (3) C rapid charging points distributed in the service zone and designed for short-term partial charging; and (4) 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 thus be modeled as a Poisson process with rate λ_v. Customers request service from the system according to a Poisson process. Both customers and vehicles are classified into n classes based on an ascending order of their required trip distance and the corresponding SoC to cover this distance, respectively. 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 an arriving vehicle to the system belongs to Class i. Note that p_0 is the probability that a vehicle arrive with a depleted battery, and is thus not 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 prior 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) join vehicles that will serve customer

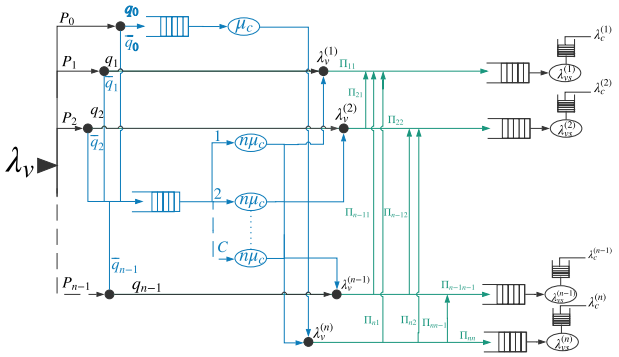

Fig. 1. Joint dispatching and partially/fully charging model, abstracting an AEMoD system in one service zone.

with their current state of charge with probability q_i (The served customer can be from any Sub-class j with $j \leq i$); or (2) partially charge up to the SoC of class $i+1$ at any of the C charging points (whenever any of them becomes free), with probability $\overline{q}_i = 1 - q_i$, before parking again in waiting to serve a customer from any Sub-class j with $j \leq i+1$. 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 C charging points with probability $\overline{q}_0 = 1 - q_0$. In the former and latter cases, the vehicle after charging will wait to serve customers of any Sub-class $j \leq n$ and 1, respectively. Considering the above explanation, Each vehicle, whether decided to serve immediately or decided to charge before serving, will be able to serve: (1) customers from same class with probability Π_{ii}; or (2) customers with a trip distance from any Sub-class with probability Π_{ij}.

As widely used in the literature (e.g., [12], [13]), the full charging time of a vehicle with a depleted battery is assumed to be exponentially distributed with rate μ_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 due to the reduced time of partially charging. The customers belonging to Class i, arriving at rate $\lambda_c^{(i)}$, will be served at a rate of $\lambda_{vs}^{(i)}$, which includes summation of proportions of arrival rates of vehicles that: (1) arrived to the zone with a SoC belonging to Class j $\forall j \geq i$ and were directed to wait to serve Class i $\forall i \leq j$ customers; or (2) arrived to the zone with a SoC belonging to Class $j-1$ and were directed to partially charge to be able to serve a Sub-Class i $\forall i \leq j$ customers. Given the above description and modeling of variables, the entire zone dynamics can thus be modeled by the queuing system . This system includes n M/M/1 queues for the n classes of customer service, one M/M/1 queue for the central charging station, and one M/M/C queue representing the partial charging process at the C charging points.

Our goal in this paper is to minimize the maximum expected response time of the entire system. By response time, we mean the time needed that vehicle starts moving from its parking or charging spot towards this customer.

III. System stability conditions

In this section, we first deduce the stability conditions of our proposed joint dispatching and charging system, using the basic laws of queuing theory. Each class of vehicles with an arrival rate $\lambda_v^{(i)}$ will be characterized by its SoC when it is ready to serve customers. Each of the n classes of customers are served by a separate queue of vehicles, with $\lambda_{vs}^{(i)}$ 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 i^{th} arrival queues. We can thus deduce from the system model in the previous section the rate of vehicles with SoC that allows to serve a class i or any sub-class $j \leq i$ that requires lower SoC to serve its customers:

$$\begin{aligned} \lambda_v^{(i)} &= \lambda_v(p_{i-1}\overline{q}_{i-1} + p_i q_i),\ i = 1, \ldots, n-1. \\ \lambda_v^{(n)} &= \lambda_v(p_{n-1}\overline{q}_{n-1} + p_0 q_0) \end{aligned} \tag{1}$$

Since we know that $\overline{q}_i + q_i = 1$ Then we substitute $\overline{q}_i$ by $1 - q_i$ in order to have a system with n variables

$$\begin{aligned} \lambda_v^{(i)} &= \lambda_v(p_{i-1} - p_{i-1}q_{i-1} + p_i q_i),\ i = 1, \ldots, n-1 \\ \lambda_v^{(n)} &= \lambda_v(p_{n-1} - p_{n-1}q_{n-1} + p_0 q_0) \end{aligned} \tag{2}$$

We can also deduce the expression of the rate of vehicles that will actually serve a class of customers i:

$$\lambda_{vs}^{(i)} = \sum_{k=i}^{n} \lambda_v^{(k)} \Pi_{ki},\ i = 1, \ldots, n \tag{3}$$

By injecting the expression of λ_v^k in (2) in (3), we find:

$$\begin{aligned} \lambda_{vs}^{(i)} &= \lambda_v \sum_{k=i}^{n-1} (p_{k-1} - p_{k-1}q_{k-1} + p_k q_k)\Pi_{ki} \\ &+ \lambda_v(p_{n-1} - p_{n-1}q_{n-1} + p_0 q_0)\Pi_{ni},\ i = 1, \ldots, n-1 \\ \lambda_{vs}^{(n)} &= \lambda_v(p_{n-1} - p_{n-1}q_{n-1} + p_0 q_0)\Pi_{nn} \end{aligned} \tag{4}$$

From the well-known stability condition of an M/M/1 queue:

$$\lambda_{vs}^{(i)} > \lambda_c^{(i)},\ i = 1, \ldots, n \tag{5}$$

To guarantee customers' satisfaction, the fog controller of each zone must impose an average response time limit T for any class. We can thus express this average response time constraint for the customers of the i-th class as:

$$\frac{1}{\lambda_{vs}^{(i)} - \lambda_c^{(i)}} \leq T \text{ or } \lambda_{vs}^{(i)} - \lambda_c^{(i)} \geq R, \text{ with } R = \frac{1}{T} \tag{6}$$

Before reaching the customer service queues, the vehicles will go through a decision step of either to go to these queues immediately or partially charge. From the system model, we have the following stability constraints on the C charging points and central charging station queues, respectively:

$$\begin{aligned} &\sum_{i=0}^{n-1} \lambda_v(p_i - p_i q_i) < C(n\mu_c) \\ &\lambda_v p_0 q_0 < \mu_c \end{aligned} \tag{7}$$

The following lemma allows the estimation of the average needed vehicles arrival for a given service zone.

Lemma 1: For the entire zone stability, and fulfillment of the average response time limit for all its classes, the average vehicles arrival rate must be lower bounded by:

$$\lambda_v \geq \sum_{i=1}^{n} \lambda_c^{(i)} + nR \tag{8}$$

Proof: The proof of Lemma 1 is in Appendix A of [11]. ■

IV. Joint Charging and Dispatching optimization

A. Problem Formulation

The goal of this paper is to minimize the maximum expected response time of the system's classes. The response time of any class is defined as the average of the duration from any customer request until a vehicle is dispatched to serve him/her. The maximum expected response time is expressed as:

$$\max_{i \in \{1,\ldots,n\}} \left\{ \frac{1}{\lambda_{vs}^{(i)} - \lambda_c^{(i)}} \right\} \tag{9}$$

It is obvious that the system's class having the maximum expected response time is the one that have the minimum expected response rate. In other words, we have:

$$\arg \max_{i \in \{1,\ldots,n\}} \left\{ \frac{1}{\lambda_{vs}^{(i)} - \lambda_c^{(i)}} \right\} = \arg \min_{i \in \{1,\ldots,n\}} \left\{ \lambda_{vs}^{(i)} - \lambda_c^{(i)} \right\} \tag{10}$$

Consequently, minimizing the maximum expected response time is equivalent to maximizing the minimum expected response rate. Using the epigraph form [14] of the latter problem, we get the following stochastic optimization problem:

$$\underset{q_0,\ldots,q_{n-1},\Pi_{11},\ldots,\Pi_{nn}}{\text{maximize}} \quad R \tag{11a}$$

s.t

$$\lambda_v \sum_{k=i}^{n-1} (p_{k-1} - p_{k-1}q_{k-1} + p_k q_k)\Pi_{ki} \tag{11b}$$

$$+ \lambda_v(p_{n-1} - p_{n-1}q_{n-1} + p_0 q_0)\Pi_{ni} - \lambda_c^{(i)} \geq R \tag{11c}$$

$$i = 1, \ldots, n-1 \tag{11d}$$

$$\lambda_v(p_{n-1} - p_{n-1}q_{n-1} + p_0 q_0)\Pi_{nn} - \lambda_c^{(n)} \geq R \tag{11e}$$

$$\sum_{i=0}^{n-1} \lambda_v(p_i - p_i q_i) < C(n\mu_c) \tag{11f}$$

$$\lambda_v p_0 q_0 < \mu_c \tag{11g}$$

$$\sum_{j=1}^{i} \Pi_{ij} = 1,\ i = 1, \ldots, n \tag{11h}$$

$$0 \leq \Pi_{ij} \leq 1,\ i = 1, \ldots, n,\ j = 1, \ldots, i \tag{11i}$$

$$0 \leq q_i \leq 1,\ i = 0, \ldots, n-1 \tag{11j}$$

$$\sum_{i=0}^{n-1} p_i = 1,\ 0 \leq p_i \leq 1,\ i = 0, \ldots, n-1 \tag{11k}$$

$$0 < R \leq \frac{\lambda_v - \sum_{i=1}^{n} \lambda_c^{(i)}}{n} \tag{11l}$$

(11m)

The n constraints in (11d) and (11e) represent the epigraph form's constraints on the original objective function in the right hand side of (10), after separation [14] and substituting every $\lambda_{vs}^{(i)}$ by its expansion form in (4). The constraints in (11f) and (11g) represent the stability conditions on charging queues. The constraints in (11h), (11i), (11j) and (11k) are the axiomatic constraints on the probabilities (i.e., values being between 0 and 1, and sum equal to 1). The Finally, Constraint (11l) is a positivity constraint on the minimum expected response rate. Finally, Constraint (11l) is is a positivity constraint and the upper bound on R introduced by Lemma (1).

B. Lower Bound Analytical Solutions

The optimization problem in (11) is a quadratic non-convex problem with second order differentiable objective and constraint functions. Usually, the solution obtained by using the Lagrangian and KKT analysis for such non-convex problems provides a lower bound on the actual optimal solution. Consequently, we propose to solve the above problem by first finding the solution derived through Lagrangian and KKT analysis, then, if needed, iteratively tightening this solution to the feasibility set of the original problem. The Lagrangian function associated with the optimization problem in (11) is given by the following expression:

$$
\begin{aligned}
L(R,\mathbf{q},\boldsymbol{\Pi},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\omega},\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\delta}) &= -\sum_{i=0}^{n-1}\omega_i q_i - \omega_n(R-\epsilon_2) \\
&+ \sum_{i=1}^{n-1}\alpha_i[\lambda_c^{(i)} - \lambda_v\sum_{k=i}^{n-1}(p_{k-1}-p_{k-1}q_{k-1}+p_k q_k)\Pi_{ki} \\
&\qquad - \lambda_v(p_{n-1}-p_{n-1}q_{n-1}+p_0 q_0)\Pi_{ni} + R] \\
&+ \alpha_n(\lambda_c^{(n)} - \lambda_v(p_{n-1}-p_{n-1}q_{n-1}+p_0 q_0)\Pi_{ni} + R) \\
&+ \beta_0(\sum_{i=0}^{n-1}\lambda_v(p_i - p_i q_i) - C(n\mu_c)) + \beta_1(\lambda_v p_0 q_0 - \mu_c) \\
&+ \sum_{i=0}^{n-1}\gamma_i(q_i-1) + \gamma_n(R - \frac{\lambda_v - \sum_{i=1}^{n}\lambda_c^{(i)}}{n}) - R \\
&+ \sum_{i=1}^{n}\sum_{j=1}^{i}\nu_{ij}(\Pi_{ij}-1) - \mu_{ij}\Pi_{ij} + \sum_{i=1}^{n}\delta_i(\sum_{k=1}^{i}\Pi_{ik}-1)
\end{aligned}
\tag{12}
$$

where:

- $\mathbf{q} = [q_0, \ldots, q_{n-1}]$ is the vector of charing decisions.
- $\boldsymbol{\Pi} = [\Pi_{ij}]$ is the vector of dispatching decisions to serve customers.
- $\boldsymbol{\alpha} = [\alpha_i]$, such that α_i is the associated Lagrange multiplier to the i-th customer queues inequality.
- $\boldsymbol{\beta} = [\beta_i]$, such that βi is the associated Lagrange multiplier to the i-th charging queues inequality.
- $\boldsymbol{\delta} = [\delta_i]$, such that δ_i is the associated Lagrange multiplier to the i-th equality constraint on the dispatching decision.
- $\boldsymbol{\gamma} = [\gamma_i]$, such that γ_i is the associated Lagrange multiplier to the i-th upper bound inequality on the charging decisions and the expected response time.
- $\boldsymbol{\omega} = [\omega_i]$, such that ω_i is the associated Lagrange multiplier to the i-th lower bound inequality on the charging decisions and the expected response time.
- $\boldsymbol{\mu} = [\mu_{ij}]$, such that μ_{ij} is the associated Lagrange multiplier to the j-th lower bound inequality on the dispatching decision Π_{ij}.
- $\boldsymbol{\nu} = [\nu_{ij}]$, such that ν_{ij} is the associated Lagrange multiplier to the j-th upper bound inequality on the dispatching decision Π_{ij}.

For more accurate resolutions, Three small positive constants ϵ_0, ϵ_1 and ϵ_2 are added to the stability conditions on the charging queues and the positivity condition on the maximum expected waiting time to make them non strict inequalities.

Solving the equations given by the KKT conditions on the problem equality and inequality constraints, the following theorem illustrates the optimal lower bound solutions of the problem in (11).

Theorem 1: The lower bound solution of the optimization problem in (11), obtained from Lagrangian and KKT analysis can be expressed as follows: where ζ_i and ζ_{ij} are the solution that that maximize $\inf_{\mathbf{q},\boldsymbol{\Pi}} L(\mathbf{q},\Pi^*,R^*,\alpha^*,\beta^*,\gamma^*,\omega^*,\mu^*,\nu^*,\delta^*)$

Proof: The proof of Theorem 1 is in Appendix C in [11]. ■

C. Solution Tightening

As stated earlier, the closed-form solution derived in the previous section from analyzing the constraints' KKT conditions does not always match with the optimal solution of the original optimization problem, and is sometimes a non-feasible lower bound on our problem. Unfortunately, there is no method to find the exact closed-from solution of non-convex optimization. However, starting from the derived lower bound, we can numerically tighten this solution by iterating toward the feasible set of the original problem. There are several algorithms to iteratively tighten lower bound solutions, one of which is the *Suggest-and-Improve algorithm* algorithm proposed in [17] to tighten non-convex quadratic problems. We will thus propose to employ this method whenever the KKT conditions based solution is not feasible and tightening is required.

V. Simulation Results

In this section, we test the merits of our proposed scheme using extensive simulations. The metric used to evaluate these merits is the maximum expected response times of the different classes. For all the performed simulation figures, the full-charging rate of a vehicle is set to $\mu_c = 0.033$ mins^{-1}, and the number of charging points $C = 40$.

Fig. 2 depicts the maximum expected response time for different values of $\sum_{i=1}^{n}\lambda_c^{(i)}$, while fixing λ_v to 8 min^{-1}. For this setting, $n = 7$ is the smallest number of classes that satisfy the stability condition in Lemma 2 in [18]. From queuing theory rules [15] [16] the more serving queues a system have,

$$
R^* = \begin{cases} \frac{\lambda_v - \sum_{i=1}^{n} \lambda_c^{(i)}}{n} & \gamma_n^* \neq 0 \\ \epsilon_2 & \omega_n^* \neq 0 \\ \sum_{i=1}^{n-1} \alpha_i^* (\lambda_v \sum_{k=i}^{n-1} (p_{k-1} - p_{k-1} q_{k-1}^* + p_k q_k^*) \Pi_{ki}^* + \lambda_v (p_{n-1} - p_{n-1} q_{n-1}^* + p_0 q_0^*) \Pi_{ni}^* - \lambda_c^{(i)}) & \\ + \alpha_n^* (\lambda_v (p_{n-1} - p_{n-1} q_{n-1}^* + p_0 q_0^*) \Pi_{nn}^* - \lambda_c^{(n)}) & Otherwise \end{cases}
$$

$$
q_0^* = \begin{cases} 0 & \alpha_1^* \Pi_{11}^* - \sum_{i=1}^{n} \alpha_i^* \Pi_{ni}^* - \beta_0^* + \beta_1^* > 0 \\ 1 & \alpha_1^* \Pi_{11}^* - \sum_{i=1}^{n} \alpha_i^* \Pi_{ni}^* - \beta_0^* + \beta_1^* < 0 \\ \frac{\lambda_c^{(n)} + \lambda_v p_{n-1} q_{n-1}^* \Pi_{nn}^* - \lambda_v p_{n-1} \Pi_{nn}^* + R^*}{\lambda_v p_0 \Pi n n^*} & \alpha_n^* \neq 0 \\ \frac{\mu_c}{\lambda_v^* p_0} & \beta_1^* \neq 0 \\ \zeta_0(R^*, q^*, \Pi^*, \alpha^*, \beta^*, \gamma^*, \omega^*, \mu^*, \nu^*, \delta^*) & Otherwise \end{cases}
$$

$$
q_i^* = \begin{cases} 0 & \alpha_{i+1}^* - \alpha_i^* - \beta_0^* > 0 \\ 1 & \alpha_{i+1}^* - \alpha_i^* - \beta_0^* < 0 \qquad i = 1, \ldots, n-1. \\ \frac{R^* + \lambda_c^{(i)} + \lambda_v \left[\sum_{k=i+2}^{n-1} (p_{k-1} q_{k-1}^* - p_k q_k^*) \Pi_{ki} + (p_{n-1} q_{n-1}^* - p_0 q_0^*) \Pi_{ni} - \sum_{k=i}^{n} p_{k-1} \Pi_{ki} + p_{i-1} q_{i-1}^* \Pi_{ii}^* - p_{i+1} q_{i+1} \Pi_{i+1i+1} \right]}{\lambda_v p_i (\Pi_{ii} - \Pi_{i+1i+1})} & \alpha_i^* \neq 0 \\ \zeta_i(R^*, q^*, \Pi^*, \alpha^*, \beta^*, \gamma^*, \omega^*, \mu^*, \nu^*, \delta^*) & Otherwise \end{cases}
$$

$$
\Pi_{ij}^* = \begin{cases} 0 & \alpha_j^* \lambda_v (p_{i-1} q_{i-1^*} - p_{i-1} - p_i q_i^*) + \delta_i^* > 0 \\ 1 & \alpha_j^* \lambda_v (p_{i-1} q_{i-1^*} - p_{i-1} - p_i q_i^*) + \delta_i^* < 0 \quad i = 1, \ldots, n-1. \\ \zeta_{ij}(R^*, q^*, \Pi^*, \alpha^*, \beta^*, \gamma^*, \omega^*, \mu^*, \nu^*, \delta^*) & Otherwise \end{cases} \tag{13}
$$

the higher the waiting time will be. Moreover, in previous related work [18] [19], we showed that increasing the number of classes n beyond its strict lower bound introduced in Lemma 2 in [18] will damage the system performance and increase the maximum response time.

Fig. 2 compare the maximum expected response time performances against $\sum_{i=1}^{n} \lambda_c^{(i)}$, for different decision approaches namely our derived optimal decisions to the following decisions sets:

1) Optimized charging decisions (i.e. $q_i \ \forall \ i$) with same class dispatching (i.e. $\Pi_{ii} = 1 \ \forall \ i$ and $\Pi_{ij} = 0 \ \forall \ i, j \neq i$)
2) Always partially charge decisions (i.e. $q_i = 0 \ \forall \ i$) with same class dispatching (i.e. $\Pi_{ii} = 1 \ \forall \ i$ and $\Pi_{ij} = 0 \ \forall \ i, j \neq i$)
3) Equal split charging decisions (i.e. $q_i = 0.5 \ \forall \ i$) with same class dispatching (i.e. $\Pi_{ii} = 1 \ \forall \ i$ and $\Pi_{ij} = 0 \ \forall \ i, j \neq i$)
4) Always partially charge decisions (i.e. $q_i = 0 \ \forall \ i$) with proportional sub-classes dispatching decisions (i.e. Π_{ij} proportional to the customers sub-classes needs)
5) Equal split charge decisions (i.e. $q_i = 0 \ \forall \ i$) with proportional sub-classes dispatching decisions (i.e. Π_{ij} proportional to the customers sub-classes needs)

These five schemes represent the possible non-optimized policies, in which each vehicle takes its own fixed decision irrespective of the system parameters. These schemes are possible in case of a non connected and optimized system.

Fig. 2 compares these approaches with a decreasing SoC distribution. The figure clearly show superior performances for our derived optimal policy compared to the other policies, especially as $\sum_{i=1}^{n} \lambda_c^{(i)}$ gets closer to λ_v, which are the most properly engineers scenarios (as large differences between these two quantities results in very low utilization), This approves the expression found in lemma 1. A Gains of 49.3%, 69.8%, 93.22%, 86.7% and 94.4% in the performances, can be noticed compared to the previously stated policies respectively.

Fig. 3 shows the study of the resilience requirements for our considered model in the critical scenarios of sudden reduction in the number of charging sources within the zone. This reduction may occur due to either natural (e.g., typical failures of one or more stations) or intentional (e.g., a malicious attack on the fog controller blocking its access to these sources). The resilience measure that the fog controller can take in these scenarios is to notify its customers of a transient increase in the vehicles' response times given the available vehicles in the zone. For this, we are only comparing the new proposed model to our previously proposed model. The figures shows clearly the advantage brought by the sub-class dispatching model. The gain gets higher in critical scenarios and reaches up to 65% with very acceptable maximum response time even is very low energy resources. This demonstrates the importance of our proposed scheme in achieving better customer satisfaction.

VI. Conclusion

In this paper, we proposed solutions to the computational and charging bottlenecks threatening the success of AEMoD systems. The computational bottleneck can be resolved by employing a fog-based architecture to distribute the optimization loads over different service zones, reduce communication delays, and matches the nature of dispatching and charging processes of AEMoD vehicles. We also proposed a multi-class dispatching and charging scheme and developed its

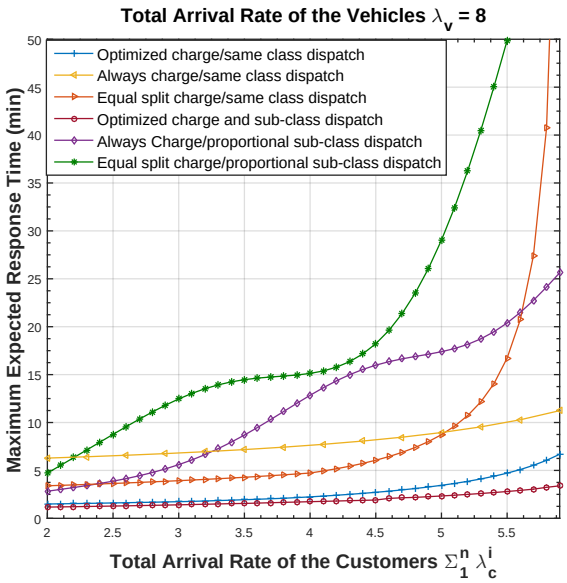

Fig. 2. Comparison to non-optimized policies for Decreasing SoC distribution.

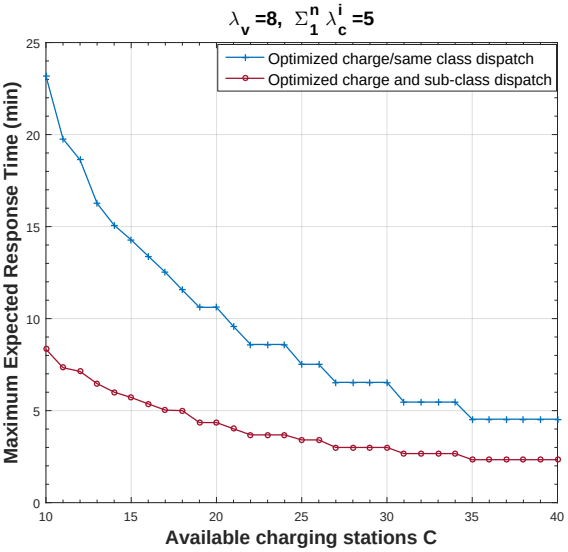

Fig. 3. Effect of varying charging points availability.

queuing model and stability conditions. We then formulated the problem of optimizing the proportions of vehicles of each class that will partially/fully charge or directly serve customers of same class or any lower sub-class as an optimization problem, in order to minimize the maximum expected system response time while respecting the system stability constraints. The optimal decisions and corresponding maximum response time were analytically derived. Simulation results demonstrated both the merits of our proposed optimal decision scheme compared to typical non-optimized schemes and previously optimized scheme, and its performance for different distributions of vehicle SoC and customer trip distances.

REFERENCES

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

[2] D. Schrank, B. Eisele, and T. Lomax, "TTIs 2012 Urban Mobility Report, " *Texas A&M Transportation Institute*, Texas, USA.2012.

[3] U. N. E. Programme, "The Emissions Gap Report 2013 - UNEP, " *Tech. Rep.*, 2013.

[4] U. E. P. Agency, "Greenhouse Gas Equivalencies Calculator, " *Tech.Rep.*, 2014. [Online]: http://www.epa.gov/cleanenergy/energy-resources/refs.html

[5] "IoT And Smart Cars: Changing The World For The Better, " *Digitalist Magazine*, August 30, 2016. [Online]: http://www.digitalistmag.com/iot/2016/08/30/iot-smart-connected-cars-willchange-world-04422640

[6] "Transportation Outlook: 2025 to 2050, " *Navigant Research, Q216*, 2016. [Online]: http://www.navigantresearch.com/research/transportation-outlook-2025-to-2050.

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

[8] "Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are, " *Cisco White Paper*, 2015. [Online]:http://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf

[9] R. Zhang, K. Spieser, E. Frazzoli, and M. Pavone, "Models, Algorithms, and Evaluation for Autonomous Mobility-On-Demand Systems, " *in Proc. of American Control Conf.*, Chicago, Illinois, 2015.

[10] R. Zhang, F. Rossi, and M. Pavone, "Model Predictive Control of Autonomous Mobility-on-Demand Systems, " *in Proc. IEEE Conf. on Robotics and Automation*, Stockholm, Sweden, 2016.

[11] S. Belakaria, M. Ammous, S. Sorour, and A. Abdel-Rahim, "Multi-Class Management with Sub-Class Service for Autonomous Electric Mobility On-Demand Systems", 2018. [Online] https://arxiv.org/abs/1804.11328

[12] H. Liang, I. Sharma, W. Zhuang, and K. Bhattacharya, "Plug-in Electric Vehicle Charging Demand Estimation based on Queueing Network Analysis, " *IEEE Power and Energy Society General Meeting*, 2014.

[13] 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. of IEEE International Conference on Communications (ICC16)*, 2016.

[14] S. Boyd and L. Vandenberghe, "Convex Optimization", 1st ed. *Cambridge: Cambridge University Press*, 2015.

[15] A. Papoulis and S. Pillai, Probability, Random Variables, and Stochastic Processes, 4th ed. International Edition: McGraw-Hill, 2002.

[16] A.L. Garcia, Probability, Statistics, and Random Processes for Electrical Engineering, 3rd ed., Prentice Hall, 2008.

[17] S. Boyd and J. Park, "General Heuristics for Nonconvex Quadratically Constrained Quadratic Programming'. *Stanford University*, 2017.

[18] S. Belakaria, M. Ammous, S. Sorour, and A. Abdel-Rahim, "A Multi-Class Dispatching and Charging Scheme for Autonomous Electric Mobility On-Demand, " *IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017.

[19] S. Belakaria, M. Ammous, S. Sorour, and A. Abdel-Rahim, "Optimal Vehicle Dimensioning for Multi-Class Autonomous Electric Mobility On-Demand Systems, " *IEEE International Communication Conference (ICC)*, 2018.

[20] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective, " *IEEE communication surveys & tutorials, VOL. 19, NO. 4, FOURTH QUARTER,* 2017.