Fachhochschule Dortmund
University of Applied Sciences and Arts
Embedded Systems Engineering

Fachhochschule Dortmund

University of Applied Sciences and Arts

Autonomous Transportation-on-Demand: Holistic Autonomous Transportation in Smart Cities

Master Thesis

Author: BRODO, Luca
Matriculation Number: 2780015

Supervisor: Henkler Stefan
Co-Supervisor: Name
Date: 03.2024

Abstract

Declaration

I hereby confirm, that I have written the Master Thesis at hand independently – in case of a group work: my respectively designated part of the work -, that I have not used any sources or materials other than those stated, and that I have highlighted any citations properly.

Contents

List of Figures

List of Tables

Chapter 1

Motivation

The economic impact of congestion in the United States is substantial, amounting to approximately \$121 billion per year or 1% of the country's GDP ([1]). This figure encompasses not only the staggering 5.5 billion hours lost to traffic congestion but also an additional 2.9 billion gallons of fuel burned. Moreover, these estimates overlook the potential costs associated with negative externalities such as vehicular emissions (including greenhouse gases and particulate matter) ([2]), travel-time uncertainty ([3]), and an elevated risk of accidents ([4]).

1. First, Create a simulator using queing theory. Some ideas are

   - Using poisson process, or a distribution that came out from invented data.
   - Use jackson networks and markov process to simulate roads and vehicles.
   - Describe the differences between graph and queing theory.
   - When modeling the probability for the customer to come, we need to do it based on multiple factory. It must be in function of the hours, the weather, the time period -¿ Use neural networks. How to train them?

2. Use the simulation to run the vehicle deplyment solution from the Project Thesis

3. Complete the model for the routing problem

   - Describe differences between static and dynamic
   - Describe the dynamic frameowrk with reuse of optimization.
   - Alternative, use the partially dynamic routing problem where you already have a set of requests and other spawn randomly. Here you can use FCFS and NN

- Include charging and comfort optimization
- The traveling time has in function o multiple elements. We have weather, road conditions, real time traffic information (Distributed system that communicates??), battery cost -¿ Non linear, we want to find a good driving profile that maximises battery performance and, at the same time, improves ride comfort

4. Run it in the simulation

- Use different solutions like tabu search and mixed integer optimization
- using the dynamic metrics, evaluate the algorith performances
- evaluate distance based on the heuristic

Chapter 2

Literature Review

Fehn et al. in [5] modeled a fleet according to energy prices. *Montoya et al.* in [6] approximated nonlinear charging function using linear piece-wise approximations. According to their result, the error was around 1%. Furthermore, they introduced the idea of different type of charging stations. *Froger et al.* in [7] also made use of a piece-wise linear function for their charging regime. Similarly, also *Kancharla et al.* in [8]. Other approaches, such as *Nie et al.* in [9] do not consider, in their model. the possibilites for vehicles to be recharged. Regarding modeling the battery charging profile, multiple approaches have been proposed. For example, *Han et al.* in [10] develop a model based on the internal resistance and voltage of the battery. This model has not been considered in favour of other models, which modeled the battery charging profile according to more fitting factors. *Yu et al.* in [11] proposed a more sophisticated model for the charging constraint in general. Their approach, however, differs from the one proposed in this work as they develop it in order to include charging the vehicles during routing, which is discarded in this work in favour of a different approach. *Lee et al.* in [12] considers a simplified model for the battery which provides states of charge (SOC) in function of the charging current. The author claims his approach to be general enough to work for every charging profile, as long as $SoC(t)$ is a concave and non-decreasing function and $SoC(0) = 0$. Moreover, there exists an inverse function $SoC^{-1}(a)$ for $0 < a < Q$.

For the vehicle dispatching problem, multiple solutions have been already proposed. For example, *Vasco et al.* in [13] model it as a linear minimum cost multicommodity flow problem; *Holzapfel et al.* in [14] model it as a MIP problem. Others, on the other hand, make use of other heuristics. For example, *Levin et al.* in [15] and *Mora et al.* in [16] use nearest neighbors. Recently, to improve real-time capabilites, some approaches have been proposed which make use of deep learning, like for e.g. the one proposed by *Yu et al.* in [17]. On a basic level, the routing problem can be reconducted to the Vehicle Routing Problem (VRP) (a thorough analysis and explanation can be found in [18]). Usually, the VRP is solved and analyzed as a static problem, which implies that

9

requests are known in advance. As a result, origins and destination of each trip is also known a priori. As pointed out by *Frazzoli et al.* in [19], in AToD systems requests are dynamic, meaning they are not known in advance. More specifically, the task of managing an AToD system can be viewed as a specific case of the dynamic one-to-one pickup and delivery problem. As pointed out by *Zhang* in [20],to provide a more detailed characterization of these systems, several additional attributes and constraints must be considered:

- Immediate Service Requests: Requests are made for immediate service rather than being scheduled in advance with specified time windows.

- Stochastic Customer Arrivals and Travel Times: The timing of customer requests and the duration of travel are subject to stochastic variability.

- Multi-Occupancy Vehicles: Contrary to [20], vehicles must be considered to have a capacity, rather than being single-occupancy. This is motivated by the fact that recently developed AV are equipped with more than one seat (see for e.g. [21])

Wollenstein-Betech et al. in [22] propose a solution to the routing-rebalancing problem in mixed traffic situations. They propose an interesting two-graph solutions, i.e. one graph for the AToD system and one for the pedestrians which are interconnected. Furthermore, they also consider private and public vehicles. *Pavone et al.* in [23] also propose a solution for the rebalancing problem. Although they mainly consider MoD systems, their approach can be easily applied to AToD systems. Their solution, moreover, focuses only on the rebalancing issue, which is studied using a steady-state fluid model, as pointed out in [22]. Furthermore, this work does not take into consideration congestions in their model.

Chapter 3

Preliminaries

Chapter 4

Time-invariant Approach

This chapter focuses on tackling dispatching, routing and rebalancing problems for an AToD system. First, a time-invariant vehicle-centric model for the transportation network is developed. The model presented in this document expands upon the one formulated in [24], using graph theory and a vehicle-centric approach to provide flexibility in handling different vehicle capacities and types. Additionally, this model addresses the combination of mobility of people with goods transportation, allowing for a more comprehensive analysis of transportation networks. This model will then be used to formulate and solve the aforementioned problems in a real-time fashion. Finally, the proposed formualtin is evaluated on a ficticious use-case.

4.1 Vehicle-centric Model

The model provided in this section will expand the model formulated in [24]. It is worth noting that, although the model presented in [24] is specific for the scenario described in previous sections of the same work, it can be expanded to cover more general use cases.

It follows that the transportation network will be modeled using graph theory. Most of the literature makes use of (Eulerian) fluid-dynamic models ([19]), i.e. customers and AVs are represented as (non-integer) flow between nodes, in this work, this model will be used in addition to a vehicle-centric approach, where customers and AVs will be modeled individuall. Contrary to the solutions proposed by most of the literature, this work combines mobility of people with goods transportation, therefore modeling vehicles individually will provided the required flexibility to handle different capacities and the increased complexity given by the plethora and variety of vehicles, as we will see later in this section. The set of vehicles will be refered to as $\mathcal{A}$.

As previously stated, let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ being the directed graph representing corresponding transportation network of the city in question, where $\mathcal{V}$ is the set of vertices and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. Any vertex, or node, $v \in \mathcal{V}$ represents

a location. Following the notation in [24], each node consists of an area of inter-
ests. For the context of this paper, the two terms will be used intercheangeably,
as they ultimately indicate the same thing. An edge $\langle v_i, v_j \rangle \in \mathcal{E}$ represents a
connection, which can consist of a road or a combination of those, linking v_i
to v_j. The level of abstraction is decided by the engineer and can be used to
regulate the granularity of the model. Locations v_i and v_j can indicate low level
key points, such as cross roads or trafficated traffic lights, as well as higher level
areas such as residental areas or leisure centers. Similarly, an edge can contain
multiple crossroads or joints or simply indicating a path from v_i to v_j, regard-
less of additional details. A correct granularity is not trivial to decide a priori
and a general approach is likewise hard to delineate. Determining the level of
abstraction remains, therefore, an engineer's prerogative and highly depends on
the system in question. It is also worth noting that the, for conveniency and to
better reflect the situation in real world applications, the graph is not a bidirect
graph. To model a two-way road, one should simply make use of two edges from
locations v_i and v_j.

As also described in [24], each edge is associated with multiple metrics and in-
formation.

Firstly, at each edge one must attribute a travel time T. T is a function
$T : \mathcal{E} \times \mathcal{A} \to \mathbb{R}_{>0}$, which at each given a vehicle and edge maps a float value
$T_{i,j}^a \in \mathbb{R}_{\geq 0}$ indicating the time required for the vehicle of type a to travel the
path from v_i to v_j. In this work, the unit of $T_{i,j}^a$ is of particular interest, as it
depends on the application.

Secondly, each edge is also associated with a distance $d : \mathcal{E} \to \mathbb{R}_{\geq 0}$, which maps
an edge to a float value $d_{i,j}$ indicating the distance from v_i to v_j.

Another factor which is considered in this work, as stated above, is pollution.
This metric is associated to each edge and is a function $f : \mathcal{E} \times \mathcal{A} \to \mathbb{R}_{\geq 0}$. To
simplify the discussion for this work, f is assumed to produce a certain value
refered to as pollution index. In other words, it is not supposed to represent
a measurable element, such as CO_2 emissions, but rather a value which can
represent multiple quantitative factors and is higher if the combination of path
and vehicle type is highly polluting. While it can be argued that the pollution
can also be simply a function of d and T, this abstraction does not account for
other elements, such as road type, vehicle fuel and road slope. The index f,
therefore, is a mathematical abstraction which allows a more flexible model for
the system. To use this abstraction, however, each edge must include additional
information, including the one mentioned before.

Similarly to what observed by *Zhang et al.* in [25], we can introduce the concept
of congestions in the model by constraining the routing by the capacity of each
road. In other words, we can associate each edge with a capacity $c : \mathcal{E} \to \mathbb{N}_{>0}$ in-
dicating the limit in terms of car occupancy above which the traffic in that edge
slows down eventually reaching a congestion. As pointed out in [25], this sem-
plified model is adequate in this context as well as the aim of this work focusing
solely on the control of vehicles to prevent congestion rather than analyzing their
behavior in congested network. While it is understood that congestions behave
differently in real world scenarios and other, more sophisticated approaches ex-

ist ([26], [27]), for the sake of simplicity, we will assume $T_{ija} = \infty$ if the number of vehicles in that edge from $\langle v_i, v_j \rangle$ to be larger than the capacity c_{ij}.

In addition, each edge $\langle v_i, v_j \rangle$ will also include information regarding traffic limitations, i.e. whether a vehicle type is allowed or not to travel that link. Intuitively, limitations will be represented as a function $s : \mathcal{E} \times \mathcal{A} \to \{0, 1\}$, where 1 implies the vehicle can traver that edge. Following the example in [24], limitations can be because of various factors, like for e.g. weight or height. For this reason, it is convenient to abstract away such details and just indicate wether a link is can be traversed or not. Notably, one could use this representation to transform a bidirect graph G' to be equivalent to G by letting $s(e, a) = 0$ for any one-way road $e = \langle v_i, v_j \rangle$ for all $a \in \mathcal{A}$. For convenience, this function will be incorporated in the definition of the capacity, updating it to become Equation 4.1.

$$c(e = \langle v_i, v_j \rangle) = \begin{cases} 0 & \text{if } s(e) = 0 \\ c(e) & \text{otherwise} \end{cases} \tag{4.1}$$

Following the discussion above, this is equivalent to setting $T_{ija} = \infty$ for all $a \in \mathcal{A}$ according to this model specifications. In this way, we can treat the road as being unaccessible without increasing the number of conditions and decrease the redeability of the model.

As mentioned above, the set of autonomous vehicles is indicated as $\mathcal{A}$ and each vehicle will be modeled as a tuple $\langle \underline{s_a}, \bar{t_a}, S_a, Q_a, I_a^b, R_a^-, R_a^+, \theta_a, B_a(t), \mathcal{R}_a, \mathcal{T}_a \rangle$. $\underline{s_a} \in \mathcal{E}$ and $\bar{t_a} \in \mathcal{E}$ are the starting and terminal node respectively; $Q_a \in \mathbb{R}_{>0}$ will be used to indicate battery capacity and charing rate and discharging rate will be represented as $R_a^+ \in \mathbb{R}_{>0}$ and $R_a^- \in \mathbb{R}_{>0}$ respectively; $\theta_a \in [0, 1]$ is used to model the battery breakpoint and $B_a(t) \in \mathbb{R}_{\geq 0}$ is the state of charge at time t; $\mathcal{R}_a$ is the set of requests assigned to vehicle a and $\mathcal{T}_a$ is the type. For a more detailed understanding of the vehicle type, the reader is encouraged to analyze the discussion in [24]. For the purpose of this work, the vehicle type will be understood as a tuple $\langle P_a, G_a, C_a, F_a \rangle$, where , $G_a \in \mathbb{R}_{\geq 0}$ and $P_a \in \mathbb{R}_{\geq 0}$ are the goods and people capacity and $C_a \in \mathbb{R}_{>0}$ and $F_a \in \mathbb{R}_{>0}$ indicate operational cost and pollution factor respectively. Furthermore, each set of assigned requests $\mathcal{R}_a$ is a subset of the set of all requests in the systems, i.e. $\mathcal{R}_a \subseteq \mathcal{R}$. For the purpose of this model, we will assume each request assignment to be unique, i.e. $\mathcal{R}_a \neq \mathcal{R}_b$ and $\mathcal{R}_a \subseteq \mathcal{R} \setminus \mathcal{R}_b$, for all $(a, b) \in \mathcal{A}$ with $a \neq b$.

The battery will be modeled according to the two operating mode, i.e. charging and discharging. While it is understood that these two operations are highly influenced by multiple factors, it is sensible to make assumptions in order to simplify the model. As it is also widely spread in industry, one can neglect the influence of external factors such as weather condition or intrinsic characteristics of the battery, such as temperature or age. The charging profile will be modelled taking inspiration from the model proposed by *Lee et al.* in [12]. As mentioned above, the vehicles have a charging rate, a battery capacity and a breakpoint, namely R_a^+, Q_a and θ_a respectively. The state of charge at time $t \in \mathbb{R}_{\geq 0}$ of a certain vehicle will be obtained according to the model described in Equation 4.2, which is derived from the CC-CV (Constant current - Constant Voltage)

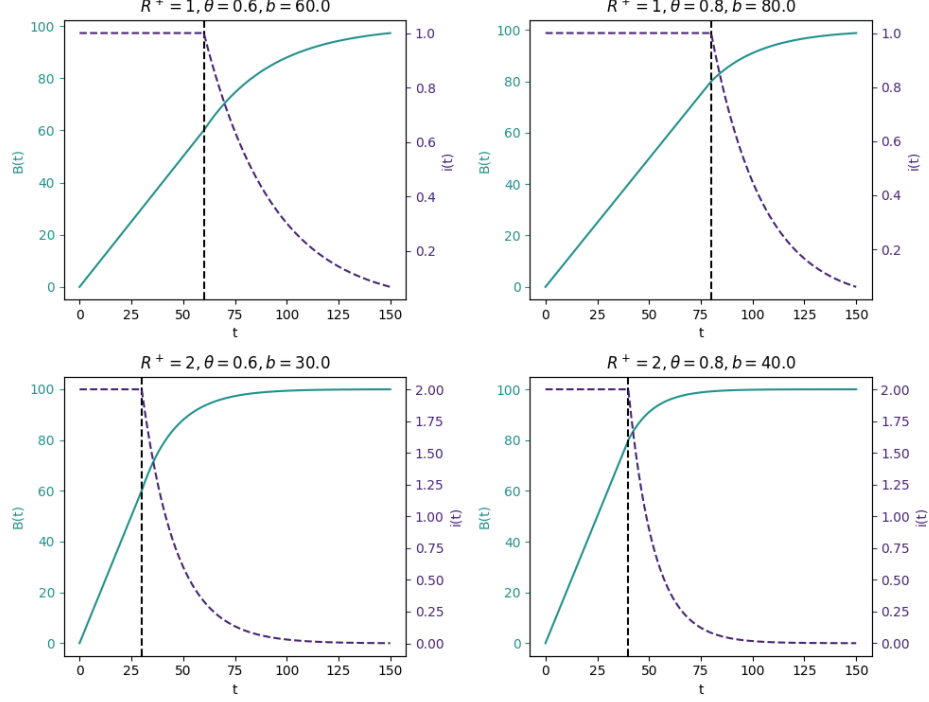scheme. (A more complete explenation can be found in [28]).

Figure 4.1: Different charging profile obtained according to the model developed in Equation 4.2. B is the State of Charge in Percentage, while $i(t)$ in function of the time unit t. The battery have $Q = 100$

$$
B_a(t+1) = \begin{cases} B(t)_a + R_a^+ & \text{if } t \leq b_a \\[2ex] Q_a - \dfrac{Q_a i(t)(1 - \theta_a)}{R_a^+} \end{cases} \tag{4.2}
$$

with $b_a = \dfrac{\theta_a Q_a}{R_a^+}$, $i(t) = \begin{cases} R_a^+ & \text{if } t \leq b_a \\ R_a^+ e^{-(t - b_a)/\tau} \end{cases}$ and $\tau \in \mathbb{R}_{\geq 0}$ being the time constant, representing how quickly the current decreases.

An example of the charging profiles from this model can be seen in Fig 4.1. Since $i(t)$ is assumed to be constant at time $t \leq b_a$, i.e. $i(t) = R_a^+$, the first

condition is derived as follows.

$$B_a(t + \alpha) = B_a(t) + \int_t^{t+\alpha} R_a^+ dt$$

$$B_a(t + \alpha) = B_a(t) + R_a^+ \int_t^{t+\alpha} dt$$

$$B_a(t + \alpha) = B_a(t) + R_a^+ \left. t \right|_t^{t+\alpha}$$

$$B_a(t + \alpha) = B_a(t) + R_a^+(t + \alpha - t)$$

$$B_a(t + \alpha) = B_a(t) + \alpha R_a^+$$

with $\alpha = 1$.

The discharging, on the other hand, will be modeled as it being directly proportional to the travel time T, as described in Equation 4.3.

$$B_a(t + T_{u,v}^a) = B(t)_a - R_a^- T_{u,v}^a \qquad (4.3)$$

The equation describes the relationship between the initial and final battery levels on edge $\langle v_u, v_v \rangle$ during the transit. Following Equation 4.3, one can assign to each edge a rate of battery discharge by defining it as the difference between the state of charge at time t and at time $t + T_{u,v}^a$.

$$D_{u,v}^a = B(t)_a - B_a(t + T_{u,v}^a) = R_a^- T_{u,v}^a \qquad (4.4)$$

Equation 4.4 allows to describe the discharging rate as a function of only the edge and the vehicle.

In terms of operational costs, multiple factors should be considered and the analysis must be extended from the one developed in [24], where the operational cost was only in function of the vehicle type. Similarly to the aforementioned work, this model assumes a certain operational cost depending on the vehicle type, however the discussion is also extended in terms of vehicle charging cost. In other words, we can decouple it from the general concept of the operational cost per vehicle and considering in function of the charging or discharging profiles described above. Furthermore, it also makes sense to assign the cost also depending on the distance traveled, i.e. the value d associated to the edges.

Compared to [24], in this work, the model also extends the information associated to the nodes. Nodes will be of three categories namely (i) normal nodes, (ii) charging stations and (iii) depots, whose sets are denoted as $\mathcal{V}_n$, $\mathcal{V}_c$, $\mathcal{V}_d$, respectively. We can, therefore, conclude that $\mathcal{V} = \mathcal{V}_n \cup \mathcal{V}_c \cup \mathcal{V}_d$. The information associated to the nodes depend from the category each node belongs to. Nodes belonging to $\mathcal{V}_n$ do not possess any specific information, as they are not of high importance for this work. Charging nodes $\mathcal{V}_c$ and depots $\mathcal{V}_d$, on the other hand, have in common the vehicle capacity they can host: charging nodes can only host as many vehicle as charging stations available, while depots can

16

only station as many vehicles as the number of free parking places. Additionally, charging nodes are assumed to have all the same models of charging stations. For simplicity, only three types will be considered, i.e. fast, normal and slow charging, which will be based on the model in Equation 4.2. The choice of parameters is explained in Table 4.1

Charging station	R^+	θ
Slow	1	0.6
Medium	1	0.8
Fast	2	0.8

Table 4.1: Choice of parameters for the charging stations according to the type. For the charging profile, please refer to Fig 4.1

Requests are modeled as tuples $\langle \underline{s}', \bar{t}', Q', P', \lambda, a', b' \rangle$, where $\underline{s}' \in \mathcal{E}, \bar{t}' \in \mathcal{E}$ represent the pickup and delivery point respectively; $G' \in \mathbb{R}_{\geq 0}$ ($P' \in \mathbb{R}_{\geq 0}$) refers to the amount of goods (people) required to transport, and $\lambda \in \mathbb{R}_{>0}$ is the rate of requests, in customers per unit time, which, therefore, makes the requests stationary and deterministic. Additionally, requests must be delivered within a time window within $[a', b']$.

4.1.1 Model Evalutation

Some comments are in order. The model in Section 4.1 is time-invariant. According to the definition presented by *Frazzoli et al.* in [19], time invariance in the context of transportation modeling refers to the assumption that the number of requests remain constant over time allowing for the simplification of temporal dynamics and treating specific time intervals as homogeneous units. This modeling concept is employed when the rate of change in transportation demands is deemed slow compared to the average travel time of individual trips, often observed in stable urban environments ([29]). While this model is indeed applied in a relatively dense urban environment, some integrations are required in order to adapt it to time-varying scenarios. Moreover, customer requests are also assumed to be known. This requirement can be fulfilled in practice with requests made in advance or some techniques to estimate requests throughout the day. It is important to note that request estimation might lead to suboptimal performance.
Secondly, as already mentioned in the previous section, the model used to describe congestions is indeed rather simple and more complex formulation might better capture the phenomena. However, the semplification is considered powerful enough for the purpose of this work and, as pointed out by *Zhang et al.* in [25] as well, more sophisticated models can be used offline using simulation techniques to derive the capacity metric used in this model.
Some comments are in order also regarding vehicles. Firstly, we assume the

vehicles to be autonomous and fault proof. It is outside of the scope of this work to deal malfunctioning vehicles or exceptional situations outside of the normale functioning regime of the system.Additionally, we assume the vehicles to be fully electric. This assumption is widely used in literature and it is motivated, among other aspects, also by the recent trends in industry to transition towards electrical mobility. Furthermore, the model provided for the battery charging and discharging profiles is rather simplistics and more sophisticated approaches exist in literature (see Chapter 2 for more details). However, these models should be seen an addition to the corrent approach and, albeit with some potential modifications, it is plausible they can be integrated in the system model.

While in [24] vehicles have been designed to be capable of transporting a potentially large number of people, some considerations should be made in this regard. While it is sensitive to consider vehicles capable of transporting up to 50 people in terms of environmental, economical and overall transportation efficiency and while it is also expected that the model designed would work with those vehicles as well, in terms of practical use, it might be more appropriate to make use of those vehicles in different ways. Notably, it should be studied whether it is efficient to tailor the routes of those vehicles according to passenger needs. Firstly, if we do not consider ride-sharing possibilities, the scenario of such large group of people traveling together is very unlikely. Additionally, if we consider ride-sharing, the possible route combinations, which increases drastically as the number of requests and nodes in the graph increase, would likely make routing such vehicle relatively inefficient. Motivated also by the fact that passengers might have common stops, it might be more sensible to treat those vehicles as buses are treated nowadays, i.e. with a pre-determined route among stops which are placed according to the most common stops, such as hospitals or train stations. Similarly, if we consider large vehicles to transport goods, such as large trucks, those are not used for home delivery, but rather used to transport goods among specialized centers. It must be noted that both situations do not invalidate this work, or any previous work. It is clear that the system described in this work in combination to the truck for goods transportation. Regarding people mobility, while buses are indeed an already established and efficient transportation mean, this systems aims at filling up the situations where a bus system is indeed lacking, such as transportation of people with special needs or, in general, more tailored to the specific needs of potential customers.

In this model, we also made certain assumptions on the nodes. First of all, we assumed the charging stations to be all of the same types. It might be argued that some stations might have different types of chargers. This characteristics can be reflected by the model simply by 'splitting' them. A station having, for instance, two types of chargers can be represented by two equivalent nodes in the graph having different category. Moreover, while in Table 4.1 we only considered three modes, this can be easily extended.
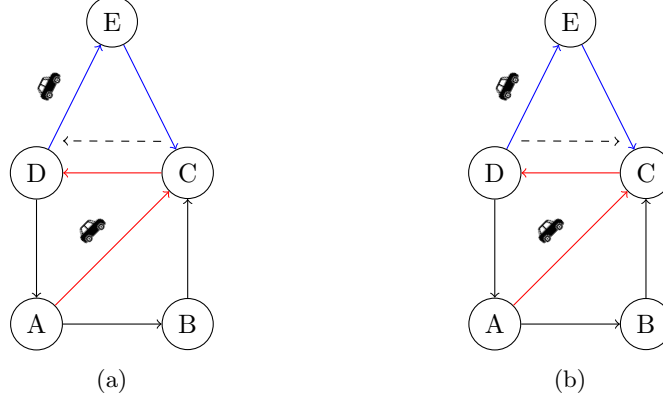
Figure 4.2: Fig 4.2a and Fig 4.2b show a simplified example of a sensible request assignment. In red and blue are the paths the two AVs can traverse, while the dashed blue two different requests (In Fig 4.2a the custoer asks to go from C to D, while in Fig 4.2b the customer asks to go from D to C). In the case of Fig 4.2a, it is more sensible to assign the request to the red AV, while in the case of Fig 4.2b, the blue AV is a better choice.

4.2 Problem Formulation

WE ASSUME THE CHARGE OF THE VEHICLE TO BE ENOUGH FOR ALL REQUESTS, CHARGIN CAN ONLY BE DONE BEFORE REBALANC-ING, THEREFORE WE ARE GOING TO LEAVE THIS PROBLEM TO AS-SIGNMENT.

4.2.1 Dispatching

Informally, the dispatching problem can be defined as the task of assigning requests to the most suitable vehicle. There exist already multiple solutions proposed for the problem and we refer to Chapter 2 for a more thorough analysis. Dispatching is critical for the overall system performance and must be done in a way that can further facilitate the next steps. Clearly, dispatching can not be decoupled and solved as a stand-alone problem. For example, Fig 4.2 shows a simplified situation where a sensible dispatching, which depends on a posterior step, will improve system performance. Nevertheless, during the dispatching problem, some additional elements must be considered as well.

Since the model proposed in Section 4.1 is a vehicle centric model, we can model the dispatching problem with the help of a binary variable x_{ar} defined in Equation 4.2.1.

$$x_{a,r} = \begin{cases} 1 & \text{if } r \text{ is assigned to } a \in \mathcal{A} \\ 0 \end{cases} \qquad \forall r \in \mathcal{R}$$

According to the model, each vehicle has a capacity which must not be exceeded. Such constraint can be expressed as follows for both people and goods.

$$\sum_{r \in \mathcal{R}} P'_r \cdot x_{a,r} \leq P_a \quad \forall a \in \mathcal{A} \tag{4.5}$$

$$\sum_{r \in \mathcal{R}} G'_r \cdot x_{a,r} \leq G_a \quad \forall a \in \mathcal{A} \tag{4.6}$$

Furthermore, if a vehicle is already on the move, the request can be picked up only if the vehicle's charge is enough to satisfy such request as well. In this case, such requirement can be expressed as follows.

$$\sum_{r \in \mathcal{R}} e(\underline{s'_r}, \bar{t'_r}) x_{a,r} \leq B_a \quad \forall a \in \mathcal{A} \tag{4.7}$$

where $e : \mathcal{E} \times \mathcal{E} \to \mathbb{R}_{>0}$ is a function expressing the required energy to go from $\underline{s'_r}$ to $\bar{t'_r}$.

Accordingly, using Equation 4.5, 4.6 and Equation 4.7, one can simply formulate it as an integer programming problem by finding the appropriate cost function to minimize, such as minimizing waiting times. Thanks to this formulation, one can also make sure that each request has been served at most once, i.e. Equation 4.8.

$$\sum_{a \in \mathcal{A}} x_{a,r} \leq 1 \quad \forall r \in \mathcal{R} \tag{4.8}$$

Furthermore, we must ensure that each request must be assigned to only one vehicle. This is insured by Equation 4.9.

$$\sum_{r \in \mathcal{R}} x_{a,r} = 1 \quad \forall a \in \mathcal{A} \tag{4.9}$$

The above mentioned equations are based upon the work presented by *Hyland et al.* in [30].

This way allows to define the following cost function which expresses the number of served requests.

$$\mathcal{J}_u = \sum_{r \in \mathcal{R}} (1 - \min_{a \in \mathcal{A}}(x_{a,r}, 1)) \tag{4.10}$$

The main strength of this approach is that can be integrated naturally in the formulation for the other steps, like for e.g. the one in Section 4.2.2.

Alternatively, the dispatching problem can be solved independently without being integrated in other steps. For example, in Section 2, some works are mentioned that make use of heuristics such as nearest neighbours. On the one hand, these approaches are known to obtain sub-optimal solutions for the problem; on the other hand, they provide flexibility and might result in less time or space complexity.

4.2.2 Routing

After being assigned to incoming requests, vehicles must be routed in such a way that can reach all customers and therefore satisfy all the requests. In other words, the routing problem consists of determining paths, i.e. a series of edges, each vehicle must travel in the graph to fulfill the requests while respecting all the requirements and minimizing some metrics, i.e. a cost function. Since this problem can be reconduced to the vehicle routing problem, in particular dynamic pickup and delivery problems (reviewd by *Toth et al.* in [31] and *Laporte et al.* in [32]), the formulation used in this work will be based on this family of problems.

Within our model, binary flow variables will be used to identify whether a vehicle should traverse a link. Formally, this can be expressed as

$$
V_{u,v}^a = \begin{cases} 1 & \text{if } a \text{ traverses } (u,v) \in \mathcal{E} \\ \\ 0 \end{cases} \qquad \forall a \in \mathcal{A}, \forall u, v \in \mathcal{V}
$$

Furthermore, another variable will be needed to deal with time-related requirements. This concept is based on the work in [33]. The decision variable s_u^a indicates the service time of vehicle a at node u. This variable will be relevant only for nodes labeled as being terminal stations for the requests, i.e. $\underline{t'a}$ and will be irrelevant for other nodes.

Accordingly, one can express the abovementioned cost function using this variable. Classical examples of cost functions are for e.g. travel time and travel distance. These are frequently used in literature ([34]), as they are general, in a sense that many other metrics could be reconduced to them. For example, residual charging or operational costs are directly influenced by the two. Furthermore, the formualtion of the cost function derived from these metrics is rather trivial and intuitive, while at the same time producing desirable results in practice. The formulation is described in equation Equation 4.11 and 4.12.

$$
\mathcal{J}_T = \sum_{a \in \mathcal{A}} \sum_{(u,v) \in \mathcal{E}} T_{u,v}^a V_{u,v}^a \tag{4.11}
$$

$$
\mathcal{J}_d = \sum_{a \in \mathcal{A}} \sum_{(u,v) \in \mathcal{E}} d_{u,v} V_{u,v}^a \tag{4.12}
$$

Since we are trying to limit the environmental impact the AToD system will have, we can leverage the pollution index we described in previous sections and develop the cost function as described in Equation 4.13.

$$
\mathcal{J}_f = \sum_{a \in \mathcal{A}} \sum_{(u,v) \in \mathcal{E}} f_{u,v}^a V_{u,v}^a \tag{4.13}
$$

Up to this point, the metrics under consideration have been geared towards minimization. Put differently, the goal is to reduce travel, encompassing both distance and time, to enhance system performance. Likewise, minimizing environmental impact is crucial in this scenario. However, there are instances where maintaining certain metrics at higher levels is preferable. For instance, closely tied to operational costs and environmental impact, it is advantageous to keep the state of charge at its maximum. Hence, it is imperative to maximize the cost function in Equation 4.14.

$$\mathcal{J}_B = \sum_{a \in \mathcal{A}} B_a \tag{4.14}$$

Finally, while it should also be explored whether the combination of those could improve the overall performance of the system. For this purpose, the cost functions can be combined using weights as follows.

$$\mathcal{J}_{tot} = \lambda_T \mathcal{J}_T + \lambda_d \mathcal{J}d + \lambda_f \mathcal{J}_f + \lambda_B \mathcal{J}_B \tag{4.15}$$

Before formulating the routing problem formally, a small consideration must be made. For convenience, the set of nodes recheable from node u by traversing a single edge $\mathcal{N}_u^+$, i.e. the outgoing neighbours of u, and $\mathcal{N}_u^-$ as the set of ingoing neighbours of u.
Additionally, just to ease the notation, we will indicate with S'_a as being the set containing all the initial stations of each request $r \in R_a$. Likewise, will be the set of terminal stations of each request $r \in R_a$.

Unconstrained Version

On a basic level, i.e. without constraints, one can formulate the *Unconstrained Routing Problem (URP)* as follows. Given a transportation network $\mathcal{G}$, a set of vehicles $\mathcal{A}$ and a set of requests $\mathcal{R}$, defined within the description in section Section 4.1, solve:

min (4.11), (4.12), (4.13) or (4.15)

s.t.

$$\sum_{u \in \mathcal{V}} V_{u,v}^a - \sum_{w \in \mathcal{V}} V_{v,w}^a = 0 \quad \forall a \in \mathcal{A}, v \in \mathcal{V} \setminus \{\underline{s_a}, \bar{t_a}\} \tag{4.16}$$

$$\sum_{u \in \mathcal{N}_{\underline{s_a}}^+} V_{\underline{s_a},u}^a = 1 \quad \forall a \in \mathcal{A} \tag{4.17}$$

$$\sum_{u \in \mathcal{N}_{t_a}^+} V_{u,\bar{t_a}}^a = 1 \quad \forall a \in \mathcal{A} \tag{4.18}$$

$$\sum_{u \in \mathcal{N}_{t_a'}^+} V_{u,\bar{t_a'}}^a = 1 \quad \forall \bar{t_a'} \in \bar{T}_a', \forall a \in \mathcal{A} \tag{4.19}$$

$$\sum_{u \in \mathcal{N}_{\underline{s_a'}}^+} V_{u,\underline{s_a'}}^a = 1 \quad \forall \underline{s_a'} \in \underline{S}_a', \forall a \in \mathcal{A} \tag{4.20}$$

(4.16) insures that for all edges which do not lead to a source or destination, if a reaches v from a road, an incoming flow will lead to an outgoing one. In other words, they guarantee connections between roads. (4.17) - (4.18) insure that each universal source and destination is reached ones. Similarly, (4.17) - (4.20) achieves the same result, but for each requests. It should be mentioned that to guarantee that those special nodes are reached at most once one can simply change from equalities to inequalites, like for e.g. $\sum_{u \in \mathcal{N}_{\underline{s_a}}^+} V_{\underline{s_a},u}^a \geq 1$.

In order to find the minimum number of vehicles required, one can use the method described *Brodo* in [24].

Constrained Version

While it could bring some interesting insights, the URP does not contain the necessary information to provide efficient routing for the scenario considered in this work. The goal is to identify the best possible path that (i) satisfies all the requests and ii is congestion free. The *Holistic Congestion-Free Routing Problem (HCRR)* is formally defined as follows.

Given a transportation network $\mathcal{G}$, a set of vehicles $\mathcal{A}$ and a set of requests $\mathcal{R}$, defined within the description in section Section 4.1, solve:

$$\min \ (4.11), (4.12), (4.13) \text{ or } (4.15)$$

s.t.

$$(4.16)\text{-}(4.20)$$

$$\sum_{a \in \mathcal{A}} V_{u,v}^a \le c_{u,v} \quad \forall (u,v) \in \mathcal{E} \tag{4.21}$$

$$V_{u,v}^a \cdot (s_u^a + T_{u,v}^a - s_v^a) \le 0 \quad \forall (u,v) \in \mathcal{E}, \forall a \in \mathcal{A} \tag{4.22}$$

$$a_v' \le s_v^a \le b_v' \quad \forall v \in \mathcal{E}, \forall a \in \mathcal{A} \tag{4.23}$$

$$\sum_{(u,v) \in \mathcal{E}} D_{u,v}^a \cdot V_{u,v}^a \le B_a(0) \quad \forall a \in \mathcal{A} \tag{4.24}$$

(4.21) insures the number of vehicles in the link (u,v) do not exceed the capacity of that link. (4.22) establishes the relationship between the service time of each node, implying that the service time of a predecessor must be lower than the successor. (4.23) establish the time window constraints, indicating that it must be within the interval of the request. For nodes which are not associated with a termination node of a request, one will simply set $a_v' = 0$ and $b_a' = \infty$, also insuring that $s_v^a \ge 0$. (4.24) assures that the vehicle charge is enough to cover all path. $B_a(0)$ can be assumed to be 100, i.e. that the batteries are full at the beginning of service.

Combination with Dispatching

Multiple solutions have already been proposed in literature which solve the dispatching and routing problem in the same work.

In this work, however, in order to combine dispatching and routing into the same formulation, one must adapt some of the conditions specified in previous sections above. More specifically, each equation related to the requests must be changed to accomodate the fact that requests have not been previously assigned.

Accordingly, the *Holistic Routing and Dispatching Problem (HRDR)* can be formulated as

$$\min \ (4.10), (4.11), (4.12), (4.13) \text{ or } (4.15)$$

s.t.

$$(4.5), (4.6)$$
$$(4.8), (4.9)$$
$$(4.16)\text{-}(4.18)$$
$$(4.21)\text{-}(4.24)$$

$$\sum_{u \in \mathcal{N}^+_{\underline{t}'_r}} V^a_{u, \bar{t}'_r} \cdot x_{a,r} = 1 \quad \bar{t}'_r \in r, \forall r \in \mathcal{R}, \forall a \in \mathcal{A} \tag{4.25}$$

$$\sum_{u \in \mathcal{N}^+_{\underline{s}'_r}} V^a_{u, \underline{s}'_r} \cdot x_{a,r} = 1 \quad \underline{s}'_r \in r, \forall r \in \mathcal{R}, \forall a \in \mathcal{A} \tag{4.26}$$

$$x_{a,r} \cdot (s^a_{\underline{s}'_r} - s^a_{\bar{t}'_r}) \leq 0 \quad \forall r \in \mathcal{R}, \forall a \in \mathcal{A} \tag{4.27}$$

(4.25) guarantees that when a request r is allocated to vehicle a, the vehicle must reach the terminal station of the request at least once. Similarly, equation (4.26) ensures that the vehicle traverses the starting station associated with the assigned request. Equation (4.27) serves the purpose of ensuring that the starting station is visited before the terminal station.

4.2.3 Rebalancing

In simpler terms, the rebalancing problem revolves around efficiently redistributing autonomous vehicles (AVs) to optimize their responsiveness to new ride requests while minimizing any existing imbalances in the system. The goal is to fine-tune the positioning of AVs, ensuring they are strategically placed to promptly meet user demands and address any inherent irregularities in the distribution of service requests. This challenge is particularly crucial in ride-sharing systems and transportation systems alike, where the dynamic nature of user requests and varying demand across different locations can lead to imbalances in the fleet's distribution. Effectively tackling the rebalancing problem enhances the overall efficiency of the system, providing users with quicker response times and a more evenly distributed service, ultimately contributing to a smoother and more reliable autonomous transportation network.

4.3 Real-time Formulation

4.4 Simulation

4.5 TO-DO

- Make the battery model in terms of the links. The time is, basically, the velocity of that link over the distance.

- The time must be adapted according to the link travelling time. I.e., each request has a time window. The time from the inbound link must be within the time window.

- Create a simulation using some fake data

Chapter 5

Model Predictive Control Approach

5.1 Linear-Discrete Time Model

Chapter 6

Use Case Simulation

Chapter 7

Summary and Outlook

In this work, we only considered one solution to solve the dispatching problem. While the solution found is guaranteed to be optimal, other algorithms could be used as well. For example, one possible approach is to develop an ad-hoc algorithm for it. For instance, requests could be assigned greedily to vehicles according to some heuristics, such as nearest neighbours, or according to the vehicle status, i.e. state of charge, capacity or traveling path. Requests should be assigned in such a way that are still in line with the previous path (see Fig 4.2). One could, for example, consider as input the previously calculated path for it.

Bibliography

[1] D. Schrank, B. Eisele, and T. Lomax, "Texas transportation institute 2012 urban mobility report," tech. rep., Texas Transportation Institute, Texas A&M University, College Station, TX, 2012.

[2] P. Pant and R. M. Harrison, "Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: a review," *Atmospheric Environment*, vol. 77, pp. 78–97, 2013.

[3] C. Carrion and D. Levinson, "Value of travel time reliability: a review of current evidence," *Transportation Research Part A: Policy and Practice*, vol. 46, pp. 720–741, 2012.

[4] D. A. Hennessy and D. L. Wiesenthal, "Traffic congestion, driver stress, and driver aggression," *Aggressive Behavior*, vol. 25, pp. 409–423, 1999.

[5] F. Fehn, F. Noack, and F. Busch, "Modeling of mobility on-demand fleet operations based on dynamic electricity pricing," in *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 1–6, 2019.

[6] A. Montoya, C. Guéret, J. E. Mendoza, and J. G. Villegas, "The electric vehicle routing problem with nonlinear charging function," *Transportation Research Part B: Methodological*, vol. 103, pp. 87–110, 2017. Green Urban Transportation.

[7] A. Froger, O. Jabali, J. E. Mendoza, and G. Laporte, "The electric vehicle routing problem with capacitated charging stations," *Transportation Science*, vol. 56, no. 2, pp. 460–482, 2022.

[8] S. R. Kancharla and G. Ramadurai, "Electric vehicle routing problem with non-linear charging and load-dependent discharging," *Expert Systems With Applications*, vol. 160, p. 113714, 2020.

[9] Z.-H. Nie, Q. Yang, E. Zhang, D. Liu, and J. Zhang, "Ant colony optimization for electric vehicle routing problem with capacity and charging time constraints," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 480–485, 2022.

[10] K. Han, T. W. Nguyen, and K. Nam, "Battery energy management of autonomous electric vehicles using computationally inexpensive model predictive control," *Electronics*, vol. 9, no. 8, 2020.

[11] J. J. Q. Yu and A. Y. S. Lam, "Autonomous vehicle logistic system: Joint routing and charging strategy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 7, pp. 2175–2187, 2018.

[12] C. Lee, "An exact algorithm for the electric-vehicle routing problem with nonlinear charging time," *Journal of the Operational Research Society*, vol. 72, pp. 1–24, 03 2020.

[13] "The dynamic vehicle allocation problem with application in trucking companies in brazil," *Computers and Operations Research*, vol. 76, pp. 118–133, 2016.

[14] A. Holzapfel, H. Kuhn, and M. G. Sternbeck, "Product allocation to different types of distribution center in retail logistics networks," *European Journal of Operational Research*, vol. 264, no. 3, pp. 948–966, 2018.

[15] M. W. Levin, K. M. Kockelman, S. D. Boyles, and T. Li, "A general framework for modeling shared autonomous vehicles with dynamic network-loading and dynamic ride-sharing application," *Computers, Environment and Urban Systems*, vol. 64, pp. 373–383, 2017.

[16] J. Alonso-Mora, S. Samaranayake, A. Wallar, E. Frazzoli, and D. Rus, "On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment," *Proceedings of the National Academy of Sciences*, vol. 114, no. 3, pp. 462–467, 2017.

[17] J. J. Q. Yu, W. Yu, and J. Gu, "Online vehicle routing with neural combinatorial optimization and deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3806–3817, 2019.

[18] P. Toth and D. Vigo, *The Vehicle Routing Problem.* Society for Industrial and Applied Mathematics, 2002.

[19] G. Zardini, N. Lanzetti, M. Pavone, and E. Frazzoli, "Analysis and control of autonomous mobility-on-demand systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, no. 1, pp. 633–658, 2022.

[20] R. Zhang, *Title of the Thesis.* Ph.d. thesis, Stanford University, Department of Aeronautics and Astronautics, 2016. Submitted to the Department of Aeronautics and Astronautics.

[21] German Aerospace Center (DLR), "Nemo Bili - Mobility from the First to the Last Mile," Year of access. Accessed on: 04.01.2024.

[22] S. Wollenstein-Betech, A. Houshmand, M. Salazar, M. Pavone, C. G. Cassandras, and I. C. Paschalidis, "Congestion-aware routing and rebalancing of autonomous mobility-on-demand systems in mixed traffic," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–7, 2020.

[23] M. Pavone, S. L. Smith, E. Frazzoli, and D. Rus, "Robotic load balancing for mobility-on-demand systems," *The International Journal of Robotics Research*, vol. 31, no. 7, pp. 839–854, 2012.

[24] L. Brodo, "Autonomous mobility-on-demand: A novel framework to study amod in smart cities." Paper not yet submitted for publication, 2023.

[25] F. Rossi, R. Zhang, and M. Pavone, "Autonomous vehicle routing in congested road networks," *CoRR*, vol. abs/1603.00939, 2016.

[26] C. R. Lindsey and E. T. Verhoef, "Congestion modelling," tech. rep., Tinbergen Institute Discussion Paper, 1999.

[27] E. T. Verhoef, "Time, speeds, flows and densities in static models of road traffic congestion and congestion pricing," *Regional Science and Urban Economics*, vol. 29, no. 3, pp. 341–369, 1999.

[28] H. Liu, I. H. Naqvi, F. Li, C. Liu, N. Shafiei, Y. Li, and M. Pecht, "An analytical model for the cc-cv charge of li-ion batteries with application to degradation analysis," *Journal of Energy Storage*, vol. 29, p. 101342, 2020.

[29] H. Neuburger, "The economics of heavily congested roads," *Transportation Research*, vol. 5, no. 4, pp. 283–293, 1971.

[30] M. Hyland and H. S. Mahmassani, "Dynamic autonomous vehicle fleet operations: optimization-based strategies to assign avs to immediate traveler demand requests," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 278–297, 2018.

[31] P. Toth and D. Vigo, *Vehicle Routing: Problems, Methods, and Applications*. Philadelphia: Society for Industrial and Applied Mathematics, 2014.

[32] G. Laporte, "Fifty years of vehicle routing," *Transportation Science*, vol. 43, p. 408–416, 2009.

[33] B. Kallehauge, J. Larsen, O. Madsen, and M. Solomon, *Vehicle Routing Problem with Time Windows*, pp. 67–98. 03 2006.

[34] A. Abdulaal, M. H. Cintuglu, S. Asfour, and O. A. Mohammed, "Solving the multivariant ev routing problem incorporating v2g and g2v options," *IEEE Transactions on Transportation Electrification*, vol. 3, no. 1, pp. 238–248, 2017.