

# PassGoodPool: Joint Passengers and Goods Fleet Management With Reinforcement Learning Aided Pricing, Matching, and Route Planning

Kaushik Manchella, Marina Haliem<sup>ID</sup>, Vaneet Aggarwal<sup>ID</sup>, *Senior Member, IEEE*,  
and Bharat Bhargava<sup>ID</sup>, *Life Fellow, IEEE*

**Abstract**—The ubiquitous growth of mobility-on-demand services for passenger and goods delivery has brought various challenges and opportunities within the realm of transportation systems. As a result, intelligent transportation systems are being developed to maximize operational profitability, user convenience, and environmental sustainability. The growth of last mile deliveries alongside ridesharing calls for an efficient and cohesive system that transports both passengers and goods. Existing methods address this using static routing methods considering neither the demands of requests nor the transfer of goods between vehicles during route planning. In this paper, we present a dynamic and demand aware fleet management framework for combined goods and passenger transportation that is capable of (1) Involving both passengers and drivers in the decision-making process by allowing drivers to negotiate to a mutually suitable price, and passengers to accept/reject, (2) Matching of goods to vehicles, and the multi-hop transfer of goods, (3) Dynamically generating optimal routes for each vehicle considering demand along their paths, based on the insertion cost which then determines the matching, (4) Dispatching idle vehicles to areas of anticipated high passenger and goods demand using Deep Reinforcement Learning (RL), (5) Allowing for distributed inference at each vehicle while collectively optimizing fleet objectives. Our proposed model is deployable independently within each vehicle as this minimizes computational costs associated with the growth of distributed systems and democratizes decision-making to each individual. Simulations on a variety of vehicle types, goods, and passenger utility functions show the effectiveness of our approach as compared to other methods that do not consider combined load transportation or dynamic multi-hop route planning. Our proposed method showed improvements over the next best baseline in various aspects including a 15% increase in fleet utilization and a 20% increase in average vehicle profits.

**Index Terms**—Ride-sharing, urban delivery, vehicle dispatch, deep q-network, reinforcement learning, intelligent transportation, fleet management.

## I. INTRODUCTION

### A. Motivation

**R**IDE Sharing market is expected to increase by a 20% compound annual growth rate (CAGR) in the next few years [1]. Further, about 90% of retailers are expected to shift to crowdsourced delivery by 2028 [2]. Platforms have three stakeholders: the offering company, the driver, as well as the customer. Customers seek convenience while drivers and the ride-hailing platforms seek profits [3]. Hence an ideal solution would be one that maximizes benefits for all these parties in providing passenger and goods transportation. Additionally, MoD services have proven potential to improve the efficiency and sustainability with vehicle sharing [4]. Urban traffic congestion, pollution, and energy expenditure are a few of the adverse side effects of transportation systems that stand to be improved by the adoption of intelligent transportation systems [5]. Route planning for mobility-on-demand (MoD) services has also proven to be an opportunity for optimization [6] towards more profitable transportation models. In the development of technology for vehicle networks, another challenge to deal with is the computational challenges that arise due to scaling the number of vehicles [7], [8]. To support quick inferences during driving, vehicles are increasingly being equipped with computational resources for in-house deep learning inference [9]. Considering this, we present a distributed algorithm for fleet management where each vehicle is able to learn its own policies and make its own decisions while contributing to the optimization of fleet objectives. The key aspects of the fleet management system include the passenger-vehicle and goods-vehicle assignment, efficient route for the vehicle to pick up and drop the passengers and goods, efficient pricing to the passengers, and the dispatching of idle vehicles. All these problems are inter-connected and influence each other, and thus cannot be studied independently. Combining passengers and goods in one framework can potentially optimize the fleet utilization; however, they impose significant challenges. As a special case, the route-planning problem (Given a set of vehicles  $V$  and passenger/goods requests  $R$ , designing a route that consists of a sequence of pick-up and drop-off locations for a given vehicle) is NP-hard [10]. Since multiple sub-problems are discrete optimization problems with large number of

Manuscript received October 30, 2020; revised May 3, 2021; accepted November 9, 2021. Date of publication December 17, 2021; date of current version March 29, 2022. This work was supported in part by the National Science Foundation under Grant CNS-1618335 and in part by CISCO Systems, Inc. An earlier version of this paper was presented in part at the NeurIPS Machine Learning for Autonomous Driving Workshop, December 2020. The Associate Editor for this article was K. Savla. (Kaushik Manchella and Marina Haliem contributed equally to this work.) (Corresponding author: Vaneet Aggarwal.)

Kaushik Manchella is with the School of IE, Purdue University, West Lafayette, IN 47907 USA (e-mail: kmanchel@purdue.edu).

Marina Haliem and Bharat Bhargava are with the Department of CS, Purdue University, West Lafayette, IN 47907 USA (e-mail: mwadea@purdue.edu; vaneet@purdue.edu).

Vaneet Aggarwal is with the School of IE and the School of ECE, Purdue University, West Lafayette, IN 47907 USA (e-mail: bbshail@purdue.edu).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TITS.2021.3128877>, provided by the authors.

Digital Object Identifier 10.1109/TITS.2021.3128877

decision variables, a computationally efficient architecture and framework is essential for the joint pricing, matching, route-planning, and dispatching, which is the focus of this paper. This joint framework brings novelty in different components as well as strong integration that can guide future research in terms of the overall system rather than performance of individual sub-systems.

### B. Related Work

1) *Dynamic Ridesharing*: Model-free approaches have been proven to be efficient in the space of urban mobility for vehicle dispatching. The authors of [11] first introduced a distributed approach to vehicle dispatching using Deep Reinforcement learning, where pooling of passengers was not considered. The authors of [12] expanded that to consider pooling which leads to significantly improved fleet performance in terms of accept rate, customer waiting time and fleet utilization. However, all of these approaches did not explicitly model pricing strategies which in fact are key drivers of the MoD marketplace. Additionally, the aforementioned approaches use naive matching approaches and left significant room for improvement in route-planning for vehicles. Efficient matching and pricing in addition to the dispatch were further considered in [10]. While this showed promising results, it did not consider layover-based multi-hop approaches in route planning and the framework only addresses passenger deliveries.

2) *Joint Pricing and Matching*: Regarding the pricing problem, there are various ways to divide the trip costs among the rideshare partners [13]–[15]. However, none of the works in literature combine pricing with matching, they study them as two separate problems. Grouping requests together greatly affects the pricing decisions and vice-versa. Ascertaining grouping together requests that are not going in opposite directions has not been addressed in literature. Most works in literature assume that vehicles are told which passengers go together, by taking near-by pickup locations together (e.g., [12], [16]) and thus are at the risk of grouping rides going in opposite directions. In this paper, we propose a model-free technique for the ride-sharing pricing problem. In contrast to the model-based approaches in literature [17]–[19], our proposed approach can adapt to dynamic distributions of customer and driver preferences. At the same time, we incorporate the pricing aspect to our matching algorithm which would prioritize rides with significant path-intersections to be grouped together. This provides an efficient dynamic pricing-aware matching algorithm for the ridesharing environment by managing the prices for passengers based on the distance traveled due to serving this passenger (i.e., inserting this request in the vehicle’s current route). The intuition is that the insertion cost of a request leading to an opposite direction to any of the previously inserted requests will be very high, and thus will be highly unlikely to be inserted into this specific route. In this manner, the matching, route planning, and pricing components work in tandem with each other to reach a common ground solution that is favorable by and profitable to both parties.

3) *Joint Passenger Goods Ridesharing*: Specific to the passenger-vehicle and goods-vehicle assignment in route planning, it has been shown in the literature that the “insertion

operation” is an effective method to solve NP-hard dynamic assignments [6], [20], [21]. However, these approaches were only limited to two requests per vehicle. Further, these approaches don’t consider goods transportation. Recently, the authors of [10] leveraged the insertion-operation and extended matching to consider vehicle capacity constraints. However, this approach can’t be applied directly since in our case, the goods can reach the destination in multiple hops, and the pricing for goods is based only on distance not the insertion cost. In [22], the authors formulated the freight insertion problem (FIP); however, they don’t consider multi-hop routing that allows for further optimization.

The authors of [23] considered joint dispatch policy for goods and passengers. However, the approach uses greedy matching policies and do not consider any pricing of the passengers. This paper proposes efficient inter-connected decisions of pricing, matching, and route-planning that influences the rewards in the dispatch problem, which is solved using a reinforcement learning approach. Specifically, we propose novel approaches by considering a combined workload of passengers and goods along with an improvement in route-planning with multi-hop considerations. The goal of our approach is to influence the customer and vehicle utility functions, utilizing an optimal dispatching framework, to (i) achieve convenient pricing and matching decisions, (ii) optimize vehicles’ route planning to deliver both passengers and goods using multi-hop routing, and (iii) re-position idle vehicles to areas of predicted high demand.

### C. Contributions

The key contributions of the paper are as follows:

- 1) This paper provides an integrated passenger and goods transportation system, that allows for pooling capabilities of the passengers, and multi-hop transfer for the goods. The proposed system integrates efficient pricing to the passengers, assignment of passengers and goods to the vehicles, dynamic route planning for vehicles, and dispatching of vehicles when they experience idle times.
- 2) This paper provides a dynamic multi-hop allocation algorithm that decides on where to best drop off goods for layovers along their journey. Our approach aims to minimize the driving time of vehicles while minimizing travel time on delivering goods.
- 3) A distributed pricing approach is proposed for the combined passengers and goods transportation problem. In this algorithm, passengers and vehicles negotiate for the best suitable price based on the preferences and utility of both parties. Further, vehicles are offered a price for goods delivery orders based on the distance gained towards the final destination.
- 4) The proposed approach integrates a demand-aware insertion based route planning algorithm for both goods and passengers that is able to scale up to the maximum capacity of each vehicle. This algorithm takes into account the near-future demand as well as the profit margin associated with each delivery to improve route-planning by preferring rides that are along the vehicles’ current route plans.

5) This paper provides a distributed dispatch policy for joint passengers and goods delivery using reinforcement learning. This approach integrates the reward functions from the pricing, matching, and route planning components, and thus achieves inter-connected decision-making among them. The optimization problem for dispatch aims to maximize vehicle profit and request acceptance while minimizing customer waiting time, idle driving time of vehicles, total number of vehicles on the road (to reduce traffic and congestion), and fuel consumption.

Using real-world taxi-trip, and goods-order datasets, we simulate experiments with our joint approach that shows significant improvements in (i) request accept rate (ii) vehicle profits (iii) idle driving. Our proposed method showed improvements in various aspects including an approximate 10% increase in accept rate, 20% increase in average vehicle profits, and approximately 20% decrease in average idle driving time of the fleet and improving fleet utilization by 15%. We note that each of the components of our framework are crucial to optimizing the fleet objectives as a whole. To the best of our knowledge, this is the first work combining pricing, matching, dispatching and route planning for joint passenger and goods transportation.

#### D. Organization

The rest of the paper is organized as follows. Section II describes the system model. Section II-E provides the architecture and high level algorithm. Section III discusses the route planning methodology including the multi-hop and insertion cost algorithms. Section IV discusses the pricing strategies. Section V describes the Distributed DQN approach for dispatching idle vehicles. Section VI evaluates and discusses the results obtained from the experiment. Section VII concludes this work with discussion on future directions.

#### E. Abbreviations and Acronyms

MoD: Mobility-on-Demand. DARM: Demand Aware Routing and Matching. DPRS: Distributed Pricing Approach for Ride-Sharing. RL: Reinforcement Learning. DQN: Deep Q-Network. Conv-Net: Convolutional Neural Network.

## II. SYSTEM MODEL

In this section, we discuss the key components of our system along with the key objectives. In addition, the model parameters and notations are also provided.

#### A. Proposed Architecture

As seen in Figure 1, the framework consists of 4 primary actors: passengers, goods, vehicles, and the MoD Platform. The MoD platform's main role is to act as an intermediary between customers and vehicles, thereby acting as a marketplace for delivery services. We formulate that the MoD platform is responsible for the following:

1) Matching and Pricing Services: providing the initial matching decisions (based on the proximity of vehicles to ride requests) and facilitate negotiations on pricing.

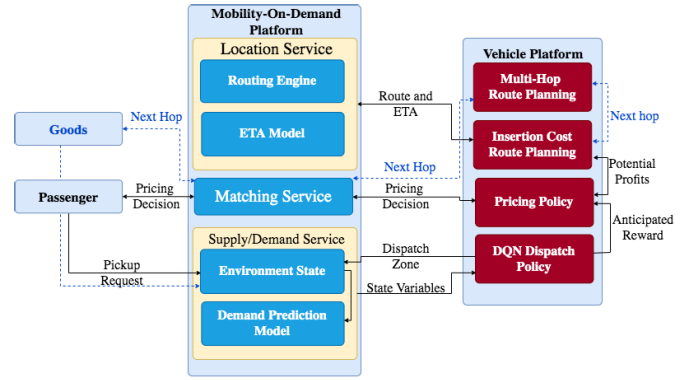


Fig. 1. An illustration of the framework architecture. The image shows the various actors, services, and interactions.

- 2) Supply/Demand Services: This would involve two sub-components: (a) maintaining the states such as current locations, current capacity, destinations, etc., for all vehicles. These states are updated in every time step based on the dispatching and matching decisions, and (b) the (Demand Prediction) model used to calculate the future anticipated demand in all zones.
- 3) Location Services: (a) the estimated time of arrival (ETA) model used to calculate and continuously update the estimated arrival time, and (b) routing engine used to generate vehicle's optimal trajectory to reach a destination.

Demand Prediction has been widely studied in literature (e.g., [11], [24]) we adopt a zone-based demand prediction approach as in [11]. Since they play a crucial role in enabling this framework, we provide details on the Demand Prediction Model and ETA Model in Appendix B.

Each of these services exists to communicate information to customers and vehicles' respective decision-making tools. Similar to a real-world scenario, the workflow involves either the passenger or goods customer requesting the MoD platform for a pickup. This initiates an initial matching which is done greedily by the MoD platform's Matching Service. The Vehicle then receives the request and processes it depending on whether it is a passenger or a good. If assigned a passenger, the system facilitates a pricing decision between the vehicles and passengers based on their respective preferences. Please note that goods are priced using the initial pricing and do not negotiate. On the passenger's end, we model a utility function to indicate the preferences and accept/reject proposed prices. Goods, on the other hand, follow the initial pricing set by the MoD platform. On the vehicle end, we propose a distributed decision-making algorithm that scales with the increasing number of vehicles. Leveraging the distributed compute in the vehicles, each vehicle is equipped with the following policies: 1) Route Planning and Matching Policy, 2) Pricing Policy, and 3) DQN Dispatch Policy. In the vehicle model, each of these policies exchange information with the MoD Platform as well as amongst themselves to make optimal decisions on: 1) Which area of the map to dispatch to, 2) How much to charge the passenger upon initial matching, and 3) Which routes to take in delivering passengers.



This framework involves both passengers and agents in the decision-making process. They learn the best pricing actions based on their utility functions that dynamically change based on each agent's set of preferences and environmental variables. With a DQN dispatch policy, vehicles act as agents that learn the optimal policy in a distributed manner (at each vehicle) through interaction with the environment. We shall refer to vehicles as agents from this point forth. Vehicles learn the best future dispatch action to take at time step  $t$ , taking into consideration the locations of all other nearby vehicles, but without anticipating their future decisions. Note that, vehicles get dispatched to areas of anticipated high-demand either when they just enter the market, or when they spend a long time being idle (searching for a ride). Vehicles' dispatch decisions are made in parallel. Therefore, our algorithm learns the optimal policy for each agent independently as opposed to centralized-based approaches such as in [11].

### B. Objectives

In this section, we detail our system's global reward objective which allows efficient fleet dispatch in fulfilling service workloads of different kinds. This global reward is optimized by our proposed algorithm in a distributed fashion as vehicles solve their own DQN (Deep Q-Network) to maximize rewards.

The goals of the system objective include: 1) satisfying the demand of pick-up orders, thereby minimizing the demand-supply mismatch. 2) minimizing the time taken to pick-up an order (aka pick-up wait time) in tandem to the dispatch time taken for a vehicle to move to a pick-up location. 3) minimizing the additional travel time incurred by orders due to participating in a shared vehicle. 4) minimizing the number of vehicles deployed to minimize fuel consumption and traffic congestion, and 5) maximizing vehicle earnings.

This overall objective is optimized at each vehicle in the distributed transportation network. This optimization includes a route planning and matching policy, a pricing policy, and a dispatching policy, all working in tandem with each other.

### C. Multi-Hop Scenario

In Figure 2, a goods delivery request is to go from zone C to the destination. The goods package can be transported by one vehicle from C to B, and another vehicle from B to the destination. Zone B will serve as a transit location referred to, in this paper, as "hop-zone". This flexibility of changing vehicles is a multi-hop transport of goods. Such multi-hop flexibility improves the packing of goods, as was shown for the case of passengers in [25]. For passenger transfers, it was shown that the multi-hop transfers leads to 30% lower cost and 20% more efficient utilization of fleets, as compared to the ride-sharing algorithms. Even though such transfers may not be convenient for passengers, they can be used for goods which motivates the choice in this paper. In addition to the multi-hop transfer of the goods, pooling capabilities of passengers are also exploited for efficient joint transportation system.

### D. Model Parameters and Notations

In this section, we describe the notations used to represent, the state, action, and reward spaces for our DQN dispatching

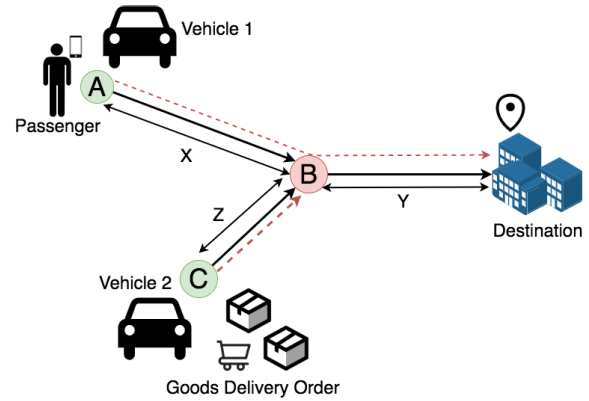


Fig. 2. An illustration of transportation scenario with hybrid delivery workload. Vehicle 1 and 2 are in zones A and C respectively. Zone B represents a hop-zone' where the packages part of the goods order packages transfer over to another vehicle.

algorithm. The optimization of the system is achieved over  $T$  time steps with each step of length  $\Delta t$ . The fleet make decisions on where on the map to go to serve at each time step  $\tau = t_0, t_0 + \Delta t, t_0 + 2(\Delta t), \dots, t_0 + T(\Delta t)$  where  $t_0$  is the start time. The map is split up into a grid with each square being taken as a zone. Zones are represented by  $i \in \{1, 2, 3, \dots, M\}$ . The number of vehicles in the fleet is represented by  $N$ . A vehicle is marked as *available* if there is remaining seating or trunk capacity. Vehicles that are completely full or are not considering taking passengers or goods are marked *unavailable*. Available vehicles in zone  $i$  at time slot  $t$  is denoted  $v_{t,i}$ . Only available vehicles are eligible to be dispatched.

1) *Vehicles' States*: The vector  $X_t$  tracks the vehicle states. For a given vehicle  $V$  we denote seating capacity and trunk space for packages as  $V_{C_{seat}}$  and  $V_{C_{trunk}}$ , respectively. As a result  $X_t$  will track: 1) current zone for vehicle  $V$ , 2) available seats, 3) available trunk space, 4) time at which delivery order is picked up, and 5) destination zone of each delivery order.

2) *Demand*: Pick-up requests at a given zone  $i$  at time slot  $t$ , represents the demand at a given area at that time and is denoted  $d_{t,i}$ . The pick-up request demand in each zone is predicted for,  $T$  steps ahead, through a historical distribution of trips across the zones, and is denoted  $D_{t:T} = (\vec{d}_t, \dots, \vec{d}_{t+T})$ .

3) *Supply*: At each time slot  $t$ , the supply of vehicles for each zone is projected to future time  $\tilde{t}$ . Consequently, the number of vehicles that will become available at  $\tilde{t}$  is denoted by  $\delta_{t,\tilde{t},i}$ . This value is ascertained from an ETA (estimated time of arrival) prediction for all moving vehicles. Consequently, given a set of dispatch actions, we are able to predict the number of vehicles in each zone for  $T$  time slots ahead, denoted by  $V_{t:T}$ . All the data is combined to represent the environment state space  $s_t$  by the tuple  $(X_t, V_{t:T}, D_{t:T})$ . At each assignment of a request to vehicle, the state space tuple is updated with the expected pick-up time, source, and destination data.

### E. Proposed Framework

In addition to the distributed architecture, our key novelties are demonstrated in the policies used for decision making at each agent. The Route Planning and Matching policy which

uses Multi-hop routing, Pricing policy, and DQN Dispatch policy interact with each other and with the MoD platform to optimize for the system objectives. At the agent, the ride requests are input to the system along with the heat map for the supply and demand service of the MoD platform. It is to be noted that our framework assumes passenger requests are constrained by seating capacity and goods delivery requests are constrained by the vehicle's trunk capacity.

After having performed an initial greedy vehicle-passenger(s) matching where one request (or more) gets assigned to the nearest vehicle based on its maximum capacity, the MoD platform initiates an initial price for the service to be provided to the assigned customer. Subsequently, the customer and price are communicated to a given agent for planning and pricing. Each agent having received the list of initial matchings (requests) assigned to it with their corresponding initial prices, performs a hop-zone assignment and an insertion operation to plan its route. In this step, vehicles reach their final matchings list by dealing with their initial matchings list in the order of their proximity to it, performing an insertion operation to its current route plan (as long as this insertion satisfies the capacity, extra waiting time, and additional travel distance constraints to guarantee that serving this request would yield a profit).

The vehicles adopt a dispatching policy using DQN, where they get dispatched to zones with anticipated high demand when they experience large idle duration or when they newly enter the market. If the customer is a passenger, using the expected discounted reward learned from DQN and the ride's destination, vehicles weigh their utility based on the potential hotspot locations and propose a new price for the customer. This takes place on a customer-by-customer basis, where a vehicle upon inserting a customer to its current route plan, proposes to him/her the new price. Then, the customer has the ability to accept or reject based on his/her own independent utility function. Finally, the vehicle upon receiving the customer's decision, either confirms the addition of this customer to its route plan or removes him/her. A Vehicle communicates with the MoD platform, as needed, to request new information on other vehicles (prior to making a dispatch or price decision) or update its own status (after any decision).

We present the proposed framework's logic in Algorithm 1. In line 1, we initialize the states of the vehicles and consequently run the system at each time step which may be discretized to  $t$ . In lines 3-4, the MoD platform initializes the vehicle marketplace and initializes their locations using the dispatching algorithm shown in Algorithm 4. The MoD platform then fetches all requests and currently available vehicles in lines 5-6 respectively. It is to be noted that the iterations across each vehicle shown in lines 7-40 may be run asynchronously at each vehicle. Each vehicle is therefore able to run the services indicated on the Vehicle Platform (see figure 1). In line 8, the MoD greedily assigns the nearest set of requests  $A_j$  to each vehicle using Algorithm 2. For each of these requests, we determine if it is a passenger or goods as in lines 10 and 24, respectively. If a passenger has been assigned, the vehicle retrieves the initial pricing which is determined by Equation (1) as seen on line 14. Lines 15-17 perform

---

### Algorithm 1 Framework Algorithm

---

```

1: Initialize vehicles' states  $X_0$  and  $t_0$ 
2: for  $t \in T$  do
3:   Fetch all vehicles that entered the market in time slot  $t$ ,  $V_{new}$ 
4:   Dispatch  $V_{new}$  to zones with anticipated high demand using
      Algo. 4
5:   Fetch all ride requests at time slot  $t$ ,  $D_t$ 
6:   Fetch all available vehicles at time slot  $t$ ,  $V_t$ 
7:   for each vehicle  $V_j \in V_t$  do
8:     Obtain initial matching  $A_j$  using Algo. 2
9:     for each request  $r_i \in A_j$  do
10:      if request is passenger then:
11:        if vehicle seating capacity  $V_{C_{seat}}^j == 0$  then
12:          Insert  $r_i$  to  $D_{t+1}$ 
13:          Continue
14:          Obtain initial pricing  $P_{init,pass}$  using Eq. (1)
15:          Perform Route planning using Algorithm 5
16:          Obtain  $S'_{V_j}[r_i]$  based on  $cost(V_j, S'_{V_j}[r_i])$ 
17:          Update trip time  $T_i$  based on  $S'_{V_j}[r_i]$  using ETA
      model
18:          Calculate final price  $P(r_i)$  based on  $S'_{V_j}[r_i]$  using
      Eq. 2
19:          Get customer  $i$ 's final decision  $C_d^i$  on  $P(r_i)$  using
      Eq. 3 and Eq. 4
20:          if  $C_d^i == 1$  then
21:            Update  $S_{V_j}$  to  $S'_{V_j}[r_i]$ 
22:          else
23:            Insert  $r_i$  to  $D_{t+1}$ 
24:          else if request is good then:
25:            if vehicle trunk capacity  $V_{C_{trunk}}^j == 0$  then
26:              Insert  $r_i$  to  $D_{t+1}$ 
27:              Break
28:            if vehicle capacity  $V_C^j == 0$  then
29:              Obtain price using  $P_{init,good}$  using Equation (1)
30:              Perform Route planning using Algorithm 5
31:              Obtain  $S'_{V_j}[r_i]$  based on  $cost(V_j, S'_{V_j}[r_i])$ 
32:              Update trip time  $T_i$  based on  $S'_{V_j}[r_i]$  (ETA model)
33:            else
34:              Assign next destination using Algorithm 3
35:              Perform Route planning using Algorithm 5
36:              Obtain  $S'_{V_j}[r_i]$  based on  $cost(V_j, S'_{V_j}[r_i])$ 
37:              Update trip time  $T_i$  based on  $S'_{V_j}[r_i]$  (ETA model)
38:            Update State Vector  $s_t$ 
39:            Retrieve next stop from  $S_{V_j}$ 
40:            Drive to next stop
41:          Fetch all idle vehicles with Idle_duration > 10 minutes,  $V_{idle}$ 
42:          Dispatch  $V_{idle}$  to zones with anticipated high demand using
      Algo. 4
43:          Update State Vector  $s_t$ 

```

---

insertion-cost based route planning as in Algorithm 5. The pricing strategy is executed on line 18 and the accept/reject decision is obtained in line 19 based on the utility function of the passenger. In the event that a goods package has been assigned, we obtain an initial price based on Equation (1) as shown on line 29. We then perform multi-hop assignment in line 30 based on Algorithm 3 followed by insertion-cost-based route planning on line 30. The two combined yield our novel multi-hop route planning for goods. It is to be noted that we place goods in the trunk capacity of a given vehicle, and the passengers in the seating capacity.

**Algorithm 2** Greedy Initialization

---

```

1: Inputs: Available Vehicles  $V_t$ , Requests  $D_t$  including origins  $o$ 
   and destinations  $d$ 
2: Outputs: Request to Vehicle Assignments  $A_j$  for each  $V_j \in V_t$ 
3: Initialize  $A_j = []$ ,  $V_{capacity}^j = V_C^j$  for each  $V_j \in V_t$ 
4: for each  $r_i \in D_t$  do
5:   if  $r_i$  is a passenger then  $C = C_{seats}$ 
6:   else if  $r_i$  is a good then  $C = C_{trunk}$ 
7:   Obtain locations of candidate vehicles  $V_{cand}$ , such that:
    $|loc(V_j) - o_i| \leq 5km^2$  and  $(V_C^j + |r_i|) \leq C_{max}^{V_j}$ .
8:   Calculate trip time  $T_{j,i} \in t_{cand,i}$  from  $loc(V_j) \in V_{cand}$  to  $o_i$ 
   using the ETA model.
9:   Pick  $V_j$  whose  $T_{j,i} = argmin(T_{cand,i})$  to serve request  $r_i$ 
10:  Push  $r_i$  to  $A_j$ 
11:  Update  $loc(V_j) \leftarrow o_i$ 
12:  Increment  $V_{capacity}^j \leftarrow V_{capacity}^j + |r_i|$ 
13: Return  $A_t = [A_j, A_{j+1}, \dots, A_n]$ , where  $n = |V_t|$ 

```

---

As we see in the overall Algorithm, the key components include the route planning and matching policy, pricing policy, and the dispatch policy. These will be discussed in the following sections.

### III. MATCHING AND ROUTE PLANNING POLICY

The vehicle-request assignment task, which we refer to as Matching, is performed for both passengers and goods. A passenger request is initially matched by the MoD platform before insertion cost based route planning by the vehicle and a pricing agreement (described in Section IV). A goods delivery request, on the other hand, is also initially matched by the MoD platform before dynamic multi-hop and insertion cost based route planning. In both cases, the overall matching policy consists of the sequences of events described below.

#### A. Initial Matching

The MoD platform initially performs a greedy vehicle-request assignment using Algorithm 2. In this procedure, each request is allocated to the nearest vehicle that minimizes waiting time given all the vehicles' capacity constraints and available status. It is important to note that passengers are allocated to the seating capacity while goods are allocated to the trunk capacity.

In Algorithm 2, we input the set of available vehicles on the map  $V_t$  and the demand backlog  $D_t$ . The algorithm ultimately outputs a set of request-to-vehicle assignments  $A_j$ . In Line 3, we initialize an empty set of assignments for  $A_j$  and fetch the capacities of each of the available vehicles into  $V_{capacity}^j$ . Consequently after line 4, for each request  $r_i$  in the demand backlog  $D_t$ , we first check whether  $r_i$  is a passenger or good and fetch vehicles seating capacity  $C_{seat}$  or trunk capacity  $C_{trunk}$  accordingly (see lines 5-6). Upon obtaining the location of vehicles in line 7, we calculate the trip time in line 8 and select the vehicle  $V_j$  that minimizes the waiting time for request  $r_i$ . We then append the assignment to  $A_j$ .

Given the set of initial assignments  $A_j$ , we proceed to the next steps depending on if an assignment is for a passenger or a good.

**Algorithm 3** Hop-Zone Assignment

---

```

1: Inputs: Goods Request (Origin  $o_i$ , Current Location  $l_i$ , Final
   Destination  $d_i$ ), Route Plan  $S_{V_j}$ , Drop-off Radius  $d_{drop}$ , Min
   Distance Gain  $d_{gain_{min}}$ 
2: Outputs: Next Destination  $h_i$ 
3: Fetch All available hop-zone locations  $H$ 
4: Compute distance from Current Location to Final Destination
    $d_0 = dist(l_i, d_i)$ 
5: if  $d_0 < d_{drop}$  then
6:   Return  $d_i$ 
7: else
8:   Initialize  $h_i$  as  $S_{V_j}[0]$ 
9:   Initialize distance  $d_s = \infty$ 
10:  for stop  $s \in S_{V_j}$  do
11:    Compute distance  $d'_s$  such that  $d'_s = dist(s, d_i)$ 
12:    if  $d'_s < d_s$  then
13:      Update  $d_s \leftarrow d'_s$ 
14:      Update  $h_i \leftarrow s$ 
15:  Initialize distance  $d_h = \infty$ 
16:  for hop-zone  $h \in H$  do
17:    Compute distance  $d'_h$  such that  $d'_h = dist(h, h_i)$ 
18:    if  $d'_h < d_s$  then
19:      Update  $d_h \leftarrow d'_h$ 
20:      Update  $h_i \leftarrow h$ 
21:    Compute distance gain  $d_{gain} = (dist(h, h_i) - d_0)/d_0$ 
22:  if  $d_h < d_{drop}$  |  $d_{gain} < d_{gain_{min}}$  then
23:    Return  $d_i$ 
24:  else
25:    Return  $h_i$ 
26: Return Hop-trips

```

---

#### B. Hop-Zone Assignment for Goods

In transporting goods, we present a route planning framework that is analogous to hitch-hiking. A package is routed through layovers or "hop-zones" (as we will refer to henceforth in this paper), in an iterative fashion such that at each iteration, the system maximizes the distance towards the final destination. This way, at each leg of the package's journey, we allocate the locally optimal hopzone which is explained in Algorithm 3. We refer to this as Multi-hop routing.

To facilitate hop-zones in practice, we consider the suggestion from [26] where each hop-zone is a public point of interest such as a store, post office, or fuel station. We assume incentives for these locations to hold packages that are in transit along their journey to the destination.

Algorithm 3 determines which hop-zone to allocate for a given goods request  $r_i$ . Given an input of the location information contained in  $r_i$ , vehicle  $V_j$ 's current route  $S_{V_j}$ , a drop-off threshold radius  $d_{drop}$ , and a minimum distance gain tolerance  $d_{gain_{min}}$ , the Hop-zone assignment algorithm outputs the "next-destination" for  $r_i$ .

$d_{drop}$  and  $d_{gain_{min}}$  are two hyper parameters that suggest the minimum threshold below which a hop-zone transit is no more desirable. As seen in line 5, if the request is already within the drop-off radius, the next destination assigned is the final destination of the request.

Given the vehicle's current route  $S_{V_j}$ , lines 10-14 first check for the stop that maximizes the distance towards the final destination of the request. Upon finding the most suitable stop, lines 16-21 find the nearest hopzone to the stop. This is deemed the



most suitable hop-zone unless the threshold parameters are not met as checked in lines 22,23.

### C. DARM Distributed Optimization for Route Planning

We use an insertion-cost based route planning for both passenger and goods delivery requests, extending the approach in [10]. If the request is a good, it proceeds to this DARM framework for route planning after hop-zone assignment from Algorithm 3. If the request is a passenger, DARM route planning is performed immediately after initial assignment from Algorithm 2. The details are provided in Appendix C.

## IV. DISTRIBUTED PRICING POLICY

In this section, we will discuss the components for pricing based strategy, that is part of the distributed ride-sharing framework detailed in Section II. When a customer requests a pickup, the MoD platform assigns a vehicle and proposes an initial price for the journey. For this initial price, we build upon what has been proposed in DPRS (Distributed pricing based ride-sharing) [10]. In our simulated environment, we considered various vehicle types: hatch-back, sedan, luxury, and van. Each of which have varying seating capacities, trunk capacities, mileage, and base price for vehicle  $j$  per trip denoted by  $B_j$ .

Given that we consider passengers and goods/packages, it is to be noted that the pricing policy from the perspective of the driver is based on which type of customer has been assigned. For each delivery of goods, vehicles adhere to the initial pricing provided by the MoD platform. On the other hand, the vehicles follow the negotiation-based pricing strategy proposed by DPRS [10] when picking up passengers. In such a case, each passenger expresses their preferences through a defined utility function in the negotiation process. Likewise, vehicles express their preferences based on their anticipated earnings in picking up the passengers. These steps are discussed in this section.

### A. Initial Pricing

When first matched, the MoD platform proposes the initial price taking into consideration the following factors:

- *Trip Distance*: This includes the distance till pickup as well as the distance from pickup to drop off. As shown in Algorithm 5, this distance is computed using the weights of the  $n$  edges that constitute the vehicle's optimal route from its current location to origin  $o_i$  and to destination  $d_i$ . This route is obtained through the insertion-operation, the route which minimizes the DARM cost function.
- *Customer/Package Waiting Time*: The time elapsed from request till pickup for a trip  $i$  is denoted as  $T_i$ .
- *Number of Shared Entities*: Number of other passengers or goods who share travelling a trip distance. This is determined from the vehicle's route considering whether all or part of the trip is shared. A vehicle  $j$ 's capacity is defined as the combined capacity  $V_j^C = V_j^{C_{trunk}} + V_j^{C_{seats}}$ . The capacity of vehicle  $j$  when it reaches the origin  $o_i$  location of this request  $r_i$  subtracted from its capacity when it reaches the destination  $d_i$ ,  $V_j^C[d_i]$ . We therefore define  $V_j^C[r_i] = |V_j^C[d_i] - V_j^C[o_i]|$ .

- *Fuel Consumption*: The cost for fuel consumption associated with this trip, denoted by  $D_i * (P_{gas}/M_j)$ , where  $P_{gas}$  represents the average gas price, and  $M_j^V$  denotes the mileage for vehicle  $j$  assigned to trip  $i$ . We note that  $D_i = S_{V_j}[r_i]$  initially, but then as vehicle  $j$ 's route gets updated to  $S'_{V_j}[r_i]$  every time a customer is added to its route,  $D_i$  becomes  $S'_{V_j}[r_i]$  to calculate the fuel consumption of the updated route.

The overall price initialization equation for request  $r_i$  is:

$$\begin{aligned}
 P_{init}[r_i] &= B_j + \left[ \omega^1 * \frac{cost(V_j, S_{V_j}[r_i])}{V_j^C[r_i]} \right] \\
 &+ \left[ \omega^2 * \left( \frac{cost(V_j, S_{V_j}[r_i])}{V_j^C[r_i]} \right) * \frac{P_{gas}}{M_j^C} \right] - \left[ \omega^3 * T_i \right]
 \end{aligned} \tag{1}$$

where  $\omega^1$ ,  $\omega^2$ , and  $\omega^3$  are the weights associated with each of the factors affecting the price calculation.  $\omega^1$  is the price per mile distance according to the vehicle type.  $\omega^2$  is set to 1 as it doesn't change across vehicles, what changes is the mileage in this factor. Finally,  $\omega^3$  is the price per waiting minute that is influenced by the vehicle type, it is negative here as we want to minimize the waiting time for the customer.

We note that, for ride-sharing customers: the trip distance is calculated as the DARM insertion cost  $S_{V_j}[r_i]$  obtained from Algorithm 5. This initial pricing gets updated for on-board passengers whenever a vehicle picks up an additional customer (as the  $V_j^C$  will now increase), where all the overlapping (shared) distances are taken into consideration (when constructing  $S'_{V_j}[r_i]$ ) and thus price may get reduced. Our proposed algorithm will first use the initial price and notify the driver, who will then modify the pricing based on the Q-values of the driver's dispatch-to location.

### B. Vehicle Proposed Pricing

The core intuition behind assessing the cost/benefit of picking up a passenger is in having knowledge over the supply-demand distribution over the city. This supply-demand distribution is learned by each vehicle through the DQN dispatch policy. The intent of the dispatch policy is to provide a given delivery vehicle with the best action to take which is predicted after weighing the expected discounted rewards (Q-values) associated with each possible move on the map using DQN (described in Section VI). We note that, with the profits term being part of the reward function, the rewards (Q-values) have embedded information about the travel distance and fuel consumption. Thus, Q-values are sufficient to reflect the profitability of serving at a given zone over the map. With the dispatch policy running every so often (set to 5 minutes in our simulation), the vehicle is able to gain the necessary insight on the supply-demand distribution (as well as all other factors embedded in the reward function) and make informed decisions on the pricing strategy. This decision making sequence is captured in the following steps: With the knowledge on expected discounted sum of rewards (Q-values) across the map, the vehicle:

- Ranks regions on the map by expected discounted sum of rewards, which may be obtained using the DQN's Q values. We denote this rank as  $\alpha$ .
- Maintain a set of  $\lambda$  highest ranked regions as  $L$ .
- Upon Route-planning, re-compute initial pricing  $P_{init}(r_i)$  using the updated route  $S'_{V_j}$  using Equation (1). This way, the vehicle takes detours required from its current path into account as it serves incoming request  $r_i$ , thus encouraging requests going in the same direction to be matched together.
- If  $loc(r_i)$  is in  $L$ , propose  $P_{init}(r_i)$ , otherwise the vehicle will propose a higher cost to make up for the cost of mobilizing to a region of low demand. In such a case, the proposed price  $P(r_i)$  is computed using Eq. (2).

$$P(r_i) = P_{init}(r_i) + [P_{init}(r_i) * \frac{\alpha_{loc(r_i)}}{2} * B_j] \quad (2)$$

### C. Passenger Decision Function

Upon a vehicle's proposed price, the passenger may accept or reject the price. This is achieved through the passenger's utility function as follows:

$$U_i = [\omega^4 * \frac{1}{V_C^j}] + [\omega^5 * \frac{1}{T_i}] + [\omega^6 * \frac{1}{V_T^j}] \quad (3)$$

In Equation (3) above, we assign weights to the passenger preferences and consider 3 components: (i) preference in sharing the vehicle expressed by  $\omega^4$ , this is captured in the utility equation based on the current capacity of vehicle  $j$  assigned to trip  $i$ , denoted by  $V_C^j$ . (ii) tolerance in waiting time  $T_i$  expressed by  $\omega^5$  (iii) preference in vehicle type  $V_T^j$  expressed by  $\omega^6$ , where  $\omega^4$ ,  $\omega^5$ , and  $\omega^6$  are the weights associated with each of the factors affecting the customer's overall utility. These values are set at random for each customer to represent a variety of customer preferences that can be accommodated by our system. The decision on whether or not a passenger accepts a proposed price is determined by the value of the utility function. We introduce a flexibility factor  $\delta_i$  to represent how much the customer  $i$  is willing to compromise in the decision-making process. If the value of the utility is  $\geq$  the price multiplied by  $\delta_i$ , then the customer accepts as follows:

$$C_d^i = \begin{cases} 1, & \text{if } U_i \geq P(r_i)\delta_i. \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

## V. DISTRIBUTED DQN DISPATCH POLICY

We utilize a distributed DQN dispatch policy to rebalance idle vehicles to areas of anticipated demand and profit. We utilize a reinforcement learning framework, with which we can learn the probabilistic dependence between vehicle actions and the reward function thereby optimizing our objective function. At each time step  $t$ , idle vehicles observe the state of the environment, perform inference on their trained DQN to dispatch to regions on the map where they can anticipate a reward. This ultimately improves fleet utilization. Algorithm 4 describes the dispatching of vehicles by running inference on the trained DQN. Line 5 of the algorithm specifically describes

---

### Algorithm 4 Dispatching Using DQN

---

- 1: **Input:**  $X_t, V_{t:T}, D_{t:T}$
  - 2: **Output:** Dispatch Locations
  - 3: **for** each vehicle **do**
  - 4:   **Construct** a state vector  $s_{t,n} = (X_t, V_{t:T}, D_{t:T})$
  - 5:   **Push** state vector to Q-Network
  - 6:   **Get** best action  $a_{t,n} = \text{argmax}[Q(s_{t,n}, a, \theta)]$  which represents dispatch zone
  - 7:   **Update** Dispatch Locations for vehicle
  - 8: **Return** Dispatch Locations
- 

the best action that a given vehicle  $V_n$  infers from the trained Q-network given the state  $s_{t,n}$  and set of possible actions  $a_{t,n}$ .

The fleet of autonomous vehicles was trained in a virtual Spatio-temporal environment that simulates urban traffic and routing. In our simulator, we used the road network of the New York City Metropolitan area along with a realistic simulation of taxi pick-ups and package delivery requests. This simulator hosts each deep reinforcement learning agent which acts as a delivery vehicle in the New York City area that is looking to maximize its reward (Eq. (5)). In our algorithm, we define the reward  $r_{t,n}$  as a weighted sum of different performance components that reflect the objectives of our DQN agent. We present the state, action, and reward for the dispatch policy:

*State:* The state variables defined in this framework capture the environment status and thus influence the reward feedback to the agents' actions. We discretize the map of our urban area into a grid of length  $x$  and height  $y$ ; resulting in a total of  $x*y$  zones. This discretization prevents our state and action space from exploding thereby making implementation feasible. The state at time  $t$  is captured by following tuple:  $(X_t, V_{t:T}, D_{t:T})$ . These elements are combined and represented in one vector denoted as  $s_t$ . When a set of new ride requests are generated, the simulator engine updates its own data to keep track of the environment status. Since the state space is large, we don't use the full representation of  $s_t$ , instead a map-based input is used to alleviate this massive computing. The three-tuple state variables in  $s_t$  are passed as an input to the DQN input layer which consequently outputs the best action to be taken.

- 1)  $X_t$  will track vehicle seating capacity and trunk space for goods/packages.  $V_{C_{seat}}$  and  $V_{C_{trunk}}$  respectively. As a result  $X_t$  will track: *current zone of vehicle  $v$ , available seats, available trunk space, pick-up time of delivery order, destination zone of each order.*
- 2)  $V_{t:T}$  is a prediction of number of available vehicles at each zone for  $T$  time slots ahead.
- 3)  $D_{t:T}$  has a term  $d_{t:T}$  that predicts joint demand of passengers & goods delivery orders at each zone for  $T$  time slots ahead.

*Action:* The action of vehicle  $n$  is denoted by  $a_{t,n}$  decides the zone it should be dispatched to at time slot  $t$ , which is given as  $Z_{t,i}$ . In our simulator, the vehicle can move (vertically or horizontally) at most 7 cells, and hence the action space is limited to these cells. A vehicle can move to any of the 14 vertical (7 up and 7 down) and 14 horizontal (7 left and 7 right). This results in a  $15 \times 15$  action space  $a_{t,n}$  for each vehicle as a vehicle can move to any of these cells or it can



remain in its own cell. If vehicle  $n$  is full it can not serve any additional customer. Alternatively, if a vehicle decides to serve current customers, the shortest route is used along the road network to pickup the assigned orders.

*Reward:* The reward function which drives the dispatch policy learner's objectives, is shaped in a manner which (1) minimize the supply-demand mismatch for passengers and goods: ( $diff_t$ ), (2) minimize the dispatch time:  $T_t^D$  (i.e., the expected travel time of vehicle  $V_j$  to go zone  $m$  at time step  $t$ ), (3) minimize the extra travel time a vehicle takes for car-pooling compared to serving one customer:  $\Delta_t$ , (4) maximize the fleet profits  $P_t$ , and (5) minimize the number of utilized vehicles:  $e_t$ . Below is the reward function  $r_{t,n} = r(s_{t,n}, a_{t,n})$  which is used at each agent  $n$  of the distributed system at time  $t$ . The weights shown in Equation (5) are used only in instances when an agent is not fully occupied and is eligible to pickup additional orders.

$$r_{t,n} = \beta_1(b_{t,n} + p_{t,n}) - \beta_2c_{t,n} - \beta_3 \sum_{u=1}^{U_n} \delta_{t,n,u} - \beta_4 \mathbb{P}_{t,n} + \beta_5 \max(e_{t,n} - e_{t-1,n}, 0) \quad (5)$$

The reward function as a linear combination of the following terms along with their respective weights  $\beta_i$ : (1) total number of customer orders picked up  $b_{t,n}$  denotes the number of passengers served by vehicle  $n$  at time  $t$  and  $p_{t,n}$  denotes the number of packages being carried in the trunk of vehicle  $n$  at time  $t$ . (2)  $c_{t,n}$  denotes the time taken by vehicle  $n$  at time  $t$  to hop or take detours to pick up extra orders. This term discourages the agent from picking up additional orders without considering the delay in current passengers/goods orders. (3)  $\sum_{u=1}^{U_n} \omega_u \cdot \delta_{t,n,u}$  denotes the sum of additional time vehicle  $n$  is incurring at time  $t$  to serve additional passengers or packages. (4)  $\mathbb{P}_{t,n}$  denotes the profit, which is calculated by subtracting cost of fuel consumption from revenue. (5)  $\max(e_{t,n} - e_{t-1,n}, 0)$  addresses the objective of minimizing the the number of vehicles at  $t$  to improve vehicle utilization.

While the primary role of the DQN is to act as a means of dispatching idle vehicles, it contains useful signals on future anticipated demand that is utilized by other components of our method including the Demand Aware Route Planning (DARM), and Pricing Policy for Passengers. The profits component  $P_t$  of the reward function encodes the expected discounted rewards (Q-values) corresponding to each position on the map with the expected earnings at these locations. As a result, vehicles now have a readily available representation of the supply-demand distribution of the city. This representation is essential in informing decisions for route planning and pricing proposals which are described in Section III and Section IV respectively.

## VI. EXPERIMENTS

In this section, we describe the methodology of our experimentation and detail the evaluation of the proposed framework against baselines. The overall objective of our experiments was to observe the key performance metrics of a fleet in an environment that is representative of a real-world use case.

### A. Methodology

1) *Simulator:* Our simulator set up in the following experiments mimicked the MoD architecture as illustrated in Figure 1. The customer, MoD Platform, and vehicles were implemented along with the services and policies specified. We developed a discrete event simulator using Python to spawn customers based on a realistic passenger and goods request dataset which is described in detail in Appendix A. Specifically, we consider the New York Metropolitan area and utilize the Open Source Routing Engine (OSRM) as the backend for the Routing Engine of the MoD platform. For each trip, we obtain the pick-up time, passenger count, origin location, and drop-off location. We use this trip information to construct travel requests demand prediction model as well.

We initialize the simulation by populating vehicles over the map constrained around the major boroughs of New York City. Given we are considering different vehicle types, we randomly assign a type and initial location to each vehicle. Each vehicle type is assigned characteristic features as follows: mileage, maximum seating capacity, maximum trunk capacity, and price rates (per mile of travel distance  $\omega^1$ , and per minute of waiting  $\omega^2$ ). We set the number of fleet vehicles to 15000 to serve the passenger and goods workload. It is to be noted that not all vehicles are populated at once. Vehicles are deployed incrementally into the market at each time step  $t$ .

We also defined a reject radius threshold for a customer request. Specifically, when a request is created by a customer and there is no vehicle within a radius of 5km, it is rejected.

2) *Multi-Hop:* Our simulator was configured with a total of 222 hop-zones evenly distributed across the New York metropolitan area. Each hopzone was initialized with a maximum holding capacity of 1000 packages. These hop-zone locations are provided to Algorithm 3 to be considered in routing for goods.

For our experiments, we set the drop-off radius  $d_{drop} = 2km$ . Additionally, we defined a minimum distance gain threshold to be  $d_{gain_{min}} = 0$ . This way, the algorithm is able to avoid assigning hop-zones which take the package further away from its destination.

3) *DQN Training and Testing:* We use the data of June 2016 for training, and one week from July 2016 for evaluations. We trained our DQN neural networks using the data from June 2016 for 10000 epochs and used the most recent 5000 experiences as a replay memory.

For each experiment, we ran training using data from the month of May for a total of 14 days. Upon saving Q-network weights, we retrieve the weights to run testing on an additional 7 days from the month of June.

4) *Evaluation Metrics:* Tying back to the objective of our system, an effective algorithm is able to perform well in maximizing the number of customers delivered, maximize the profits for vehicles, minimize the idle driving time (thereby improving fleet utilization), and minimize the customer waiting time. Taking these into consideration, for each of the considered algorithms, we evaluate the following metrics to observe the performance of the fleet as a whole.

- *Accepted Requests:* The total number of customers served indicates how effectively the algorithm is able

to minimize the supply demand gap and fulfill delivery requests. We investigate both the accepted requests for goods and passengers separately.

- *Rejected Requests*: The total number of customers rejected. We also show this metric for rejected goods' requests and rejected passengers' requests separately.
- *Occupied Vehicles*: The total number of vehicles currently carrying passengers or goods. We further show the average occupied seating capacity by passengers and the average occupied trunk capacity by goods.
- *Waiting Time*: This captures the time that passengers or goods had to wait to be picked up. We note that wait time is an important metric for customer convenience with mobility-on-demand services. So, we show this for passengers and goods separately.
- *Profits*: This is the average amount of profits accumulated by a vehicle over the course of a day.
- *Cruising Time*: This represents the time at which a vehicle is neither occupied nor gaining profit but still incurring gasoline cost. Lower cruising times therefore suggests a cost effective policy.
- *Travel Distance*: The total amount of distance during which fuel was being consumed.
- *Occupancy Rate*: This is defined as the percentage of time where vehicles are occupied out of their total working time.

## B. Baselines

Given our proposed novelties of multi-hop route planning, and combined goods+passenger workload transportation, we formulate the following baselines to evaluate the performance of the novelties relative to existing methods:

- 1) *Combined Load (DARM + DPRS + DMH)*: This is our proposed method where we extend DARM + DPRS with Dynamic Multi-hop (DMH) route planning for goods delivery. Here we consider a combined load where each vehicle transports a combination of goods and passengers.
- 2) *Combined Load (DARM + DPRS)*: This is the baseline with an insertion cost route planning and matching along with distributed pricing policy. Here we consider a combined load where each vehicle transports a combination of goods and passengers. Note that Dynamic Multi-hop routing has been omitted.
- 3) *Independent Load (DARM + DPRS + DMH)*: This is the baseline with an insertion cost and multi-hop route planning and matching along with distributed pricing policy. Here we consider an independent load where each vehicle transports either goods or passengers exclusively.
- 4) *Independent Load (DARM + DPRS)*: This is the baseline with an insertion cost route planning and matching along with distributed pricing policy as in [10]. Here we consider an independent load where each vehicle transports either goods or passengers exclusively. Note that Dynamic Multi-hop routing has been omitted.

The proposed baselines aim to evaluate the effectiveness of our dynamic multihop route planning for goods, as well as

the effectiveness of combined load transportation (goods + passengers) as opposed to independent load transportation. In addition, we provide a sensitivity analysis (Appendix D) and an ablation study showing the effectiveness of each component of our framework in Appendix E.

Our proposed method, *Combined Load (DARM + DPRS + DMH)*, incorporates both multihop route planning and combined load delivery with pricing strategies. As compared to *Combined Load (DARM + DPRS)* baseline, we hypothesize that the multihop route planning would be a more effective approach. Given that the core intuition of multihop route planning is to break trips into smaller segments and delivery only to hop locations that are convenient for the vehicle, we expect to see improvements in the number of requests served, the number of fleet vehicles deployed, profits, cruising time, travel distance, and occupancy rate.

Likewise, we have included baselines that do not consider combined load transportation to observe the effectiveness of the combined load transportation. In these scenarios, we formulate our hypothesis that a combined load transportation scenario improves fleet utilization due to the more efficient packing of the vehicle. As opposed to independent load scenarios, in the combined load scenario, we utilize the vehicle trunk capacity. Additionally, the flexibility a vehicle has in transporting both passengers and goods should provide improved profits in the combined scenario.

The combination of these two strategies was expected to yield the most improved fleet utilization and profits. In addition, we verify the effectiveness of the DPRS framework, where drivers negotiate the price with passengers, by comparing against a baseline that does not adopt DPRS in Fig. 3. Note that, for goods, only the initial pricing is used and no negotiation takes place. Further, we verify the multi-hop routing for goods (DMH) by comparing against a baseline that doesn't adopt multi-hops (in Fig. 4), instead, this baseline only drops the goods at their final destination. Further, we see in Appendix D that the dispatch decision takes  $< 0.2$  seconds. Impact of reject radius and drop-off radius is also investigated in Appendix D.

## C. Results Discussion

From our simulations, we observe that the hypothesis for each baseline comparison has been supported for the most part by our experimental results.

Fig. 3 shows that the distributed pricing framework (DPRS) improves the number of accepted passengers' requests by 10 – 30%, decreases the average travel distance of the fleet by about 10 – 25%, and saves up to 800 vehicles from the fleet to serve other requests (i.e., the occupancy rate of the fleet) when compared to a non-pricing baseline. This proves that DPRS framework is able to reach a common ground between passengers and drivers to accommodate both their preferences. In addition, Fig. 4 shows that the dynamic multi-hop approach (DMH) increases the accept rate of goods' requests by 15 – 20%, decreases the travel distance of the fleet by 10 – 20%, and the occupancy rate by about 500 vehicles on average, when compared to a non multi-hop baseline. This proves the efficiency of the DMH approach where it serves

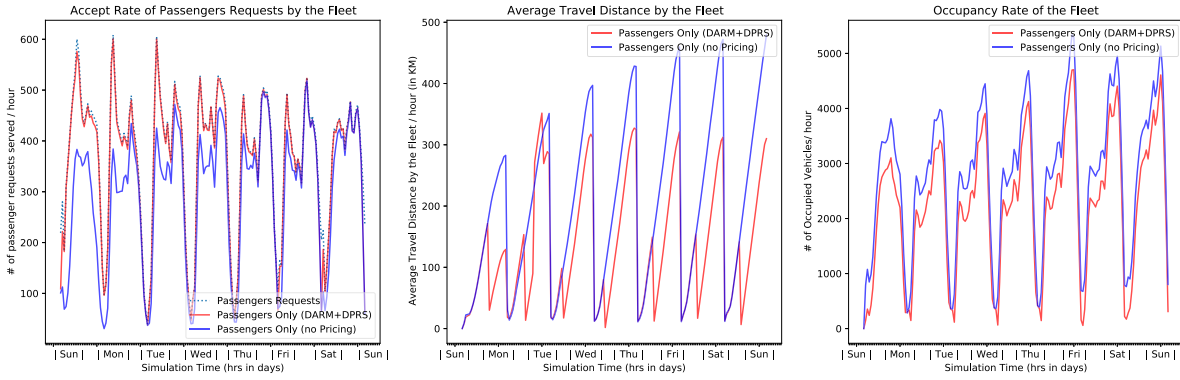


Fig. 3. Performance metrics for an Independent Load (DARM+DPRS) of Passengers Vs. a baseline with no pricing.

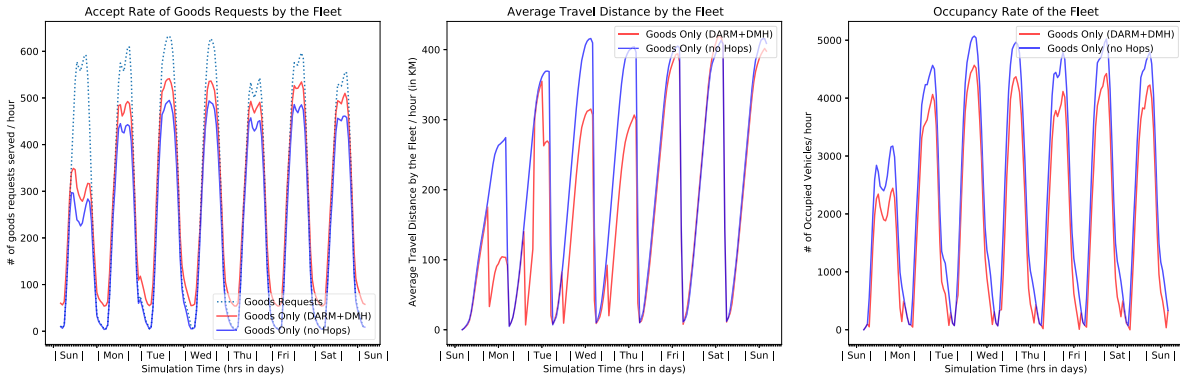


Fig. 4. Performance metrics for an Independent Load (DARM+DMH) of Goods Vs. a baseline with no multi-hops.

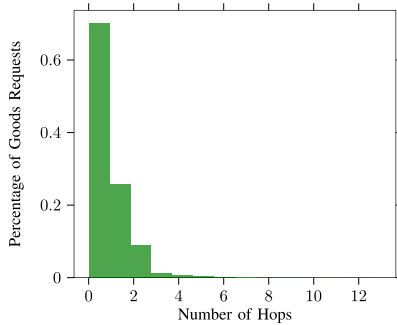


Fig. 5. Histogram of number of hops for all goods deliveries.

more requests with a lower number of vehicles and without increasing their average travel distance.

On the proposed method, we observe the number of hops taken by goods on Figure 5 where 25% of requests had one stop and 8% had two stops. Looking at Figure 13 and Fig. 14 (given in Appendix E), we notice that our proposed method shows the highest number of requests accepted over the course of the simulation (serving  $\approx 97\%$  of the total goods' requests and up to 99% of the total passengers requests). Figures 15 - 18 (in Appendix E) show the performance metrics for each one of the four baselines mentioned above. We note that following our proposed approach, *Combined Load (DARM + DPRS)* comes second, while the two baselines which do not consider a combined load come slightly behind. What is most noteworthy, however, is that the proposed

method is able to deliver more requests while having the fewest number of occupied vehicles. We notice approximately 50% improvement relative to *Independent Load (DARM + DPRS)*, approximately 30% improvement from *Independent Load (DARM + DPRS + DMH)*, and about 10% improvement to *Combined Load (DARM + DPRS)*. Further, Fig. 12 (in Appendix E) shows that the proposed method improves the utilization of both the seating and trunk capacities, with about 22% occupied trunk capacity and 40% occupied seating capacity on average, which leaves enough room for serving future requests, owing in part to the use of multihop route planning for goods. In addition, it shows that the average waiting time for goods' doesn't exceed 20 seconds; whereas for passengers, it is between 10 – 15 seconds. Further analysis of the baselines shown in these figures will be discussed in Appendix E as we investigate the effectiveness of each component of our framework.

In summary, our proposed method combines two strategies that have demonstrated a more profitable and efficient fleet usage: multihop route planning, and combined load transportation. With the addition of insertion based route planning, and demand aware pricing strategies proposed in [10], we demonstrate effective strategies for the efficient management of delivery fleet of vehicles for both goods and passengers.

## VII. CONCLUSION AND FUTURE WORK

In this paper, we proposed a framework to enable distributed decision making for fleet vehicles with passenger and goods



delivery tasks with the objective of maximizing operational profitability and customer convenience. The system includes a novel route-planning and matching algorithm, pricing strategies, and a Deep Reinforcement Learning based dispatch algorithm. Multi-hop and insertion-cost considerations in our route planning enable vehicles to generate ideal routes on-the-fly. Our pricing strategy which is part of the planning mechanism involves both customers as well as drivers in the decision-making processes. The agents are informed by utility functions that aim to achieve a maximum monetary profit for both drivers and customers. We integrate these algorithms with a distributed DQN approach which dispatches idle vehicles with the objective of anticipating profits from potential customers across the map thus maximizing profit while minimizing idle driving and fuel expenses. Evaluation in a realistic discrete-event simulator demonstrates promising results for the inclusion of multi-hop considerations in route planning. Most notably, we observed a significant reduction in idle driving time which saves on fuel expenses and improved fleet utilization by approximately 15%. The 20% improvement in profits also points to great potential for adoption in a profit-driven society. It is also worth noting that our approach can easily be scaled up to a large number of fleet vehicles and readily deployable locally within each vehicle to enable distributed inference.

In regards to future research directions, we propose developing systems that can incorporate deadline-based constraints to enable transportation of highly urgent goods such as medicines or perishable items. Additionally, it would be interesting to see the improvement of the proposed approach in less densely populated locations than New York City. Extensions to account for non-stationarity [27], [28] and use of intelligent option policies [29] in this framework are important research directions.

## REFERENCES

- [1] S. Stasha. (2021). *Ride-Sharing Industry Statistics to Get You Going in 2021*. Accessed: May 10, 2021. [Online]. Available: <https://policyadvice.net/insurance/insights/ride-sharing-industry-statistics/>
- [2] Zebra. (2019). *Reinventing the Supply Chain: The Future of Fulfillment Vision Study*. Accessed: May 10, 2021. [Online]. Available: [https://www.zebra.com/content/dam/zebra\\_new\\_ia/en-us/solutions-verticals/vertical-solutions/retail/vision-study/fulfillment-vision-study-report-en-us.pdf](https://www.zebra.com/content/dam/zebra_new_ia/en-us/solutions-verticals/vertical-solutions/retail/vision-study/fulfillment-vision-study-report-en-us.pdf)
- [3] R. Hahn and R. Metcalfe, "The ridesharing revolution: Economic survey and synthesis," in *More Equal by Design: Economic Design Responses to Inequality*, vol. 4. Oxford, U.K.: Oxford Univ. Press, 2017.
- [4] S. Shaheen, "Shared mobility: The potential of ridehailing and pooling," in *Three Revolutions*. Washington, DC, USA: Island Press, 2018, pp. 55–76.
- [5] D. Metz, "Developing policy for urban autonomous vehicles: Impact on congestion," *Urban Sci.*, vol. 2, no. 2, p. 33, Apr. 2018.
- [6] X. Bei and S. Zhang, "Algorithms for trip-vehicle assignment in ride-sharing," in *Proc. AAAI*, vol. 18, 2018, pp. 3–9.
- [7] F. Jameel, Z. Chang, J. Huang, and T. Ristaniemi, "Internet of autonomous vehicles: Architecture, features, and socio-technological challenges," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 21–29, Aug. 2019.
- [8] Z. Lei, X. Qian, and S. V. Ukkusuri, "Efficient proactive vehicle relocation for on-demand mobility service with recurrent neural networks," *Transp. Res. C, Emerg. Technol.*, vol. 117, Aug. 2020, Art. no. 102678.
- [9] R. F. Molanes, K. Amarasinghe, J. Rodriguez-Andina, and M. Manic, "Deep learning and reconfigurable platforms in the Internet of Things: Challenges and opportunities in algorithms and hardware," *IEEE Ind. Electron. Mag.*, vol. 12, no. 2, pp. 36–49, Jun. 2018.
- [10] M. Haliem, G. Mani, V. Aggarwal, and B. Bhargava, "A distributed model-free ride-sharing approach for joint matching, pricing, and dispatching using deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7931–7942, Dec. 2021.
- [11] T. Oda and C. Joe-Wong, "MOVI: A model-free approach to dynamic fleet management," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, Apr. 2018, pp. 2708–2716.
- [12] A. O. Al-Abbasi, A. Ghosh, and V. Aggarwal, "DeepPool: Distributed model-free algorithm for ride-sharing using deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4714–4727, Dec. 2019.
- [13] N. A. H. Agatz, A. L. Erera, M. W. P. Savelsbergh, and X. Wang, "Dynamic ride-sharing: A simulation study in metro Atlanta," *Transp. Res. B, Methodol.*, vol. 45, no. 9, pp. 1450–1464, Nov. 2011.
- [14] H. Wang and H. Yang, "Ridesourcing systems: A framework and review," *Transp. Res. B, Methodol.*, vol. 129, pp. 122–155, Nov. 2019.
- [15] C. Zhang, J. Xie, F. Wu, X. Gao, and G. Chen, "Pricing and allocation algorithm designs in dynamic ridesharing system," *Theor. Comput. Sci.*, vol. 803, pp. 94–104, Jan. 2020.
- [16] Y. Xu, Y. Tong, Y. Shi, Q. Tao, K. Xu, and W. Li, "An efficient insertion operator in dynamic ridesharing services," *IEEE Trans. Knowl. Data Eng.*, early access, Sep. 28, 2020, doi: [10.1109/TKDE.2020.3027200](https://doi.org/10.1109/TKDE.2020.3027200).
- [17] 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, Jan. 2016.
- [18] S. Ma, Y. Zheng, and O. Wolfson, "Real-time city-scale taxi ridesharing," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 7, pp. 1782–1795, Jul. 2014.
- [19] A. Kleiner, B. Nebel, and V. A. Ziparo, "A mechanism for dynamic ride sharing based on parallel auctions," in *Proc. 22nd Int. Conf. Artif. Intell.*, 2011.
- [20] H. Zheng and J. Wu, "Online to offline business: Urban taxi dispatching with passenger-driver matching stability," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 816–825.
- [21] J. Wang *et al.*, "Demand-aware route planning for shared mobility services," *Proc. VLDB Endowment*, vol. 13, no. 7, pp. 979–991, Mar. 2020.
- [22] B. Li, D. Krushinsky, H. A. Reijers, and T. Van Woensel, "The share-a-ride problem: People and parcels sharing taxis," *Eur. J. Oper. Res.*, vol. 238, no. 1, pp. 31–40, Oct. 2014.
- [23] K. Manchella, A. K. Umrawal, and V. Aggarwal, "Flexpool: A distributed model-free deep reinforcement learning algorithm for joint passengers and goods transportation," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 2035–2047, Apr. 2021.
- [24] J. Ke, X. Qin, H. Yang, Z. Zheng, Z. Zhu, and J. Ye, "Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network," *Transp. Res. C, Emerg. Technol.*, vol. 122, Jan. 2021, Art. no. 102858.
- [25] A. Singh, A. O. Al-Abbasi, and V. Aggarwal, "A distributed model-free algorithm for multi-hop ride-sharing using deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, early access, Jul. 8, 2021, doi: [10.1109/TITS.2021.3083740](https://doi.org/10.1109/TITS.2021.3083740).
- [26] C. Chen *et al.*, "Crowddeliver: Planning city-wide package delivery paths leveraging the crowd of taxis," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1478–1496, Jun. 2017.
- [27] M. Haliem, V. Aggarwal, and B. Bhargava, "AdaPool: A diurnal-adaptive fleet management framework using model-free deep reinforcement learning and change point detection," *IEEE Trans. Intell. Transp. Syst.*, early access, doi: [10.1109/TITS.2021.3109611](https://doi.org/10.1109/TITS.2021.3109611).
- [28] X. Qian and S. V. Ukkusuri, "Time-of-day pricing in taxi markets," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1610–1622, Jun. 2017.
- [29] X. Qian, S. Guo, and V. Aggarwal, "DROP: Deep relocating option policy for optimal ride-hailing vehicle repositioning," 2021, *arXiv:2109.04149*.
- [30] Y. Tong, Y. Zeng, Z. Zhou, L. Chen, J. Ye, and K. Xu, "A unified approach to route planning for shared mobility," *Proc. VLDB Endowment*, vol. 11, no. 11, p. 1633, 2018.
- [31] TLC. *NYC Taxi Limousine Commission-Trip Record Data*. Accessed: 2018. [Online]. Available: <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>
- [32] C.-C. Tao, "Dynamic taxi-sharing service using intelligent transportation system technologies," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2007, pp. 3209–3212.