



HAL
open science

Solving the collaborative vehicle routing problem via multi-agent deep reinforcement learning

Ibrahim El Baraghi, Omar Rifki, Hamid Allaoui

► **To cite this version:**

Ibrahim El Baraghi, Omar Rifki, Hamid Allaoui. Solving the collaborative vehicle routing problem via multi-agent deep reinforcement learning. Séminaire Intelligence Artificielle et Optimisation de l'A2U, Jun 2024, Calais, France. hal-04763620

HAL Id: hal-04763620

<https://hal.science/hal-04763620v1>

Submitted on 2 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Solving the collaborative vehicle routing problem via multi-agent deep reinforcement learning

Ibrahim EL BARAGHI¹, Omar Rifki², Hamid Allaoui³

¹LISIC, Université Sidi Mohamed Ben Abdellah, Maroc - ibrahim.el-baraghi@univ-littoral.fr

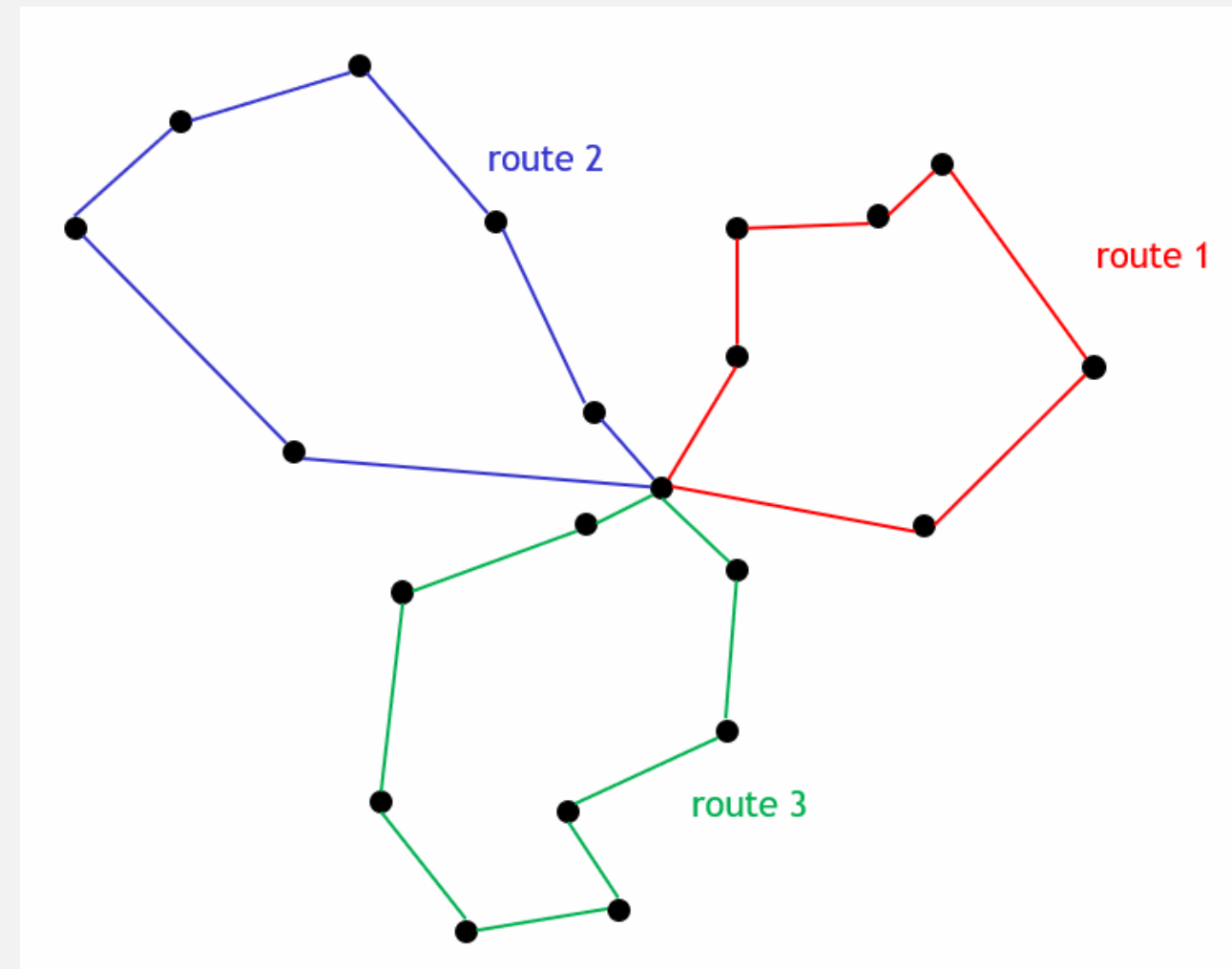
²LISIC, Université du Littoral Côte d'Opale, France - omar.rifki@univ-littoral.fr

³LGI2A, Université d'Artois, France - hamid.allaoui@univ-artois.fr

1. Context

Context

In logistics and transportation, companies aim to minimize costs and environmental impact but often face challenges due to the unwillingness to share sensitive customer information.



Objective

Reduce transportation and environmental costs for companies that do not wish to share their customer data.

Centralized approaches

- Centralized collaborative planning involves a central authority making decisions with full information, optimizing for the total profit of all participants jointly. (1)

Decentralized approaches

- Decentralized collaborative planning, independent entities collaborate by exchanging subsets of their customers with limited information sharing.
- aiming to maximize their individual profits while maintaining some degree of cooperation. (1)

Our approach

- Centralized planning with decentralized elements.
- Each agent is free to use its heuristic to decide which customers to share, and instead of auctions assigning shared customers, a reinforcement learning (RL) approach with an attention mechanism handles these assignments.

2. Our problem

Exact method

- We propose a model of Collaborative VRP in MILP formulation for benchmark purposes with our proposed multi-agent RL architecture. (2)

$$\min \sum_{i \in V} \sum_{j \in V} C_{ij} \cdot X_{ij}$$

subject to :

$$\sum_{j \in V_a} X_{0j} \leq K_a \quad \forall a \in \{1, \dots, A\}$$

$$\sum_{i \in V_a} X_{iu} = \sum_{j \in V_a} X_{uj} \quad \forall u \in V_a \setminus \{0\} \quad \forall a \in \{1, \dots, A\}$$

$$\sum_{j \in V} X_{ij} = 1 \quad \forall i \in V \setminus \{0\}$$

$$y_j \geq y_i + q_j X_{ij} - Q(1 - X_{ij}) \quad \forall i, j \in V$$

$$d_i \leq y_i \leq Q \quad \forall i \in V$$

$$X_{ij} \in \{0, 1\} \quad \forall i, j \in V$$

Giving :

$$A = \text{number of agents}$$

$$V = \text{nodes including the depot}$$

$$= V_1 \cup V_2 \cup \dots \cup V_A$$

$$\{0\} = \text{one depot}$$

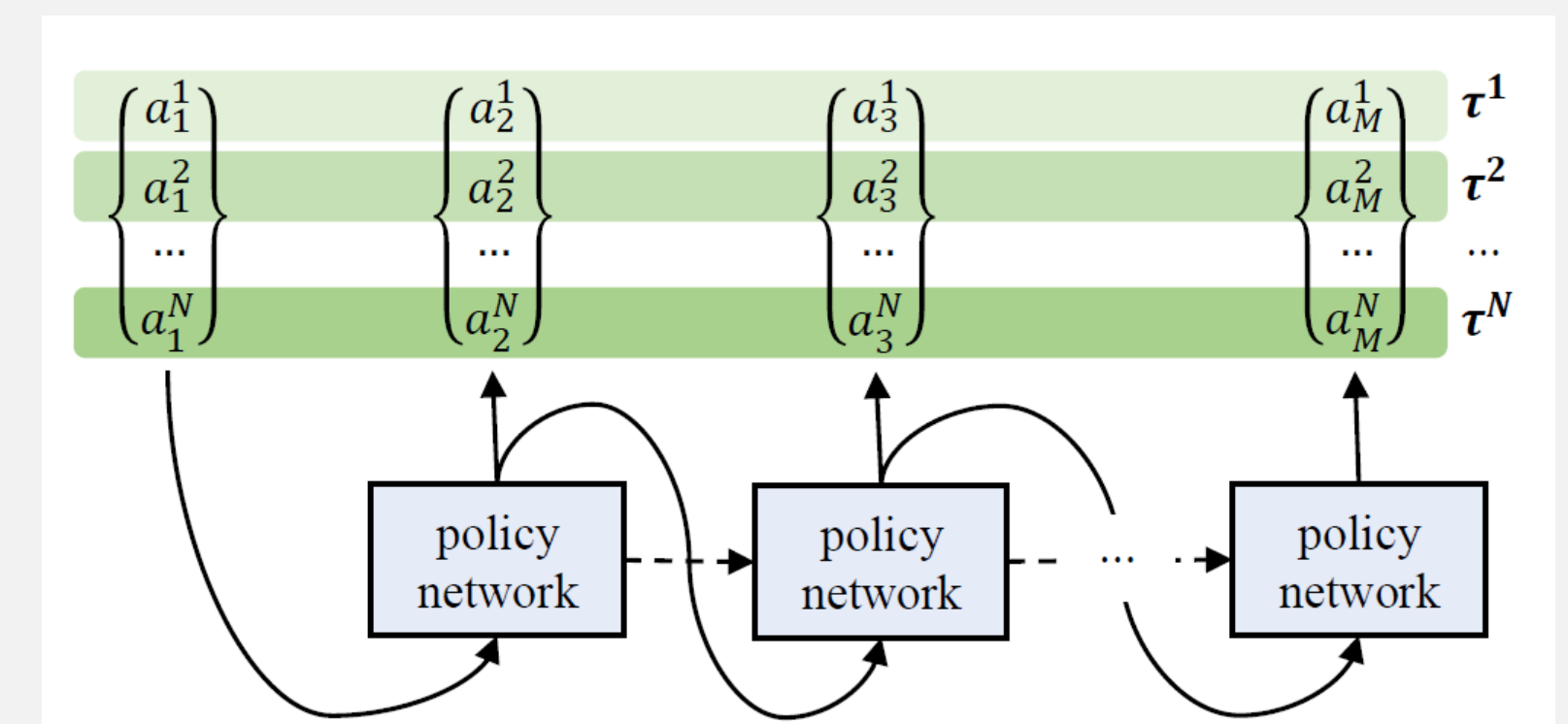
4. Perspectives

- Validate the effectiveness of this architecture with random test instances and literature benchmarks.

3. Learning approach via centralized multi-agent Deep Reinforcement Learning

Base model

- An attention-based model with encoder-decoder networks.
- We choose the *Kown* (3) approach as it evolves several solutions in parallel.
- In our approach each solution corresponds to an agent.



Node selection criteria

- We force all agents to start at the depot.
- For the first move, an arbitrary customer is selected for each agent.
- After we use the decoder to select the best next_node to visit by providing it with the last visited node, remaining load, and a mask of the previously visited nodes.

$$\left\{ \begin{array}{ll} \forall \text{ agent } start_{node} = depot & t = 0 \\ \forall \text{ agent } next_{node} = select_random_node & t = 1 \\ \forall \text{ agent } next_{node} = select_from_decoder() & otherwise \end{array} \right.$$

Masking mechanism

- Hide all customers not belonging to each agent.
- Hide a shared customer when it is visited to other agents.

$$\left\{ \begin{array}{ll} \forall \text{ agent } hide(\{selected_{node}, agent\}) & selected_{node} \in S_{nodes} \\ hide(\{selected_{node}, agent_i\}) & otherwise \end{array} \right.$$

Giving: $S_{nodes} = \text{Set of shared nodes}$

- Hide customers with a demand greater than the remaining load.

Problem instance generation

- Customer nodes locations and the depot node are sampled uniformly from the square.
- A customer node i has a demand δ/D_i , where δ is sampled uniformly from a discrete set $\{1, 2, \dots, 9\}$ and D is set according to the number of customers.
- We introduced an additional variable called the customer allocation vector, which assigns customer nodes to agents, and sets shared customers to the value 0.

5. Main References

- Margaretha Gansterer, Richard F. Hartl. Collaborative vehicle routing : A survey. European Journal of Operational Research 268, 2018
- Pedro Munari, Twan Dollevoet, Remy Spliet. A generalized formulation for vehicle routing problems. arXiv preprint arXiv:1606.01935v2, 2017
- Yeong-Dae Kwon, Jinho Choo, Byoungjip Kim, Iljoo Yoon, Seungjai Min, and Youngjune Gwon. POMO: Policy Optimization with Multiple Optima for Reinforcement Learning. In NeurIPS 2020.