Learning Profile-Aware Vehicle Routing Problems with Collaborative Attention

Extended abstract submitted for presentation at the 12^{th} Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

Note: This submission is double-blinded. Do not include author names or affiliations.

October 31, 2024

Keywords: neural combinatorial optimization, vehicle routing problems, reinforcement learning, attention network

1 INTRODUCTION

Vehicle Routing Problems (VRPs) optimize delivery routes for vehicle fleets. While the Heterogeneous Capacitated Vehicle Routing Problem (HCVRP) considers vehicles with varying capacities, we propose the Profiled Vehicle Routing Problem (PVRP), which adds vehicle-client-specific operational constraints. Due to its NP-hard nature, large instances rely on heuristic solutions, though these face scalability challenges.

Recent advances in reinforcement learning (RL) offer promising alternatives for combinatorial optimization. RL can learn effective solutions through simulated environments without domain expertise and potentially outperform traditional heuristics in complex scenarios. Most approaches use pointer networks, encoding input data for fast autoregressive decoding. While learning-based methods have succeeded in various VRPs, including multi-agent and heterogeneous agent scenarios, they still have not addressed per-client agent heterogeneity.

We propose the Collaborative Attention Model with Profiles (CAMP), a multi-agent reinforcement learning approach for PVRP. CAMP extends the attention-based encoder-decoder framework by incorporating vehicle-client profiles and enabling collaborative decision-making through specialized communication layers. The model uniquely processes profiled client representations per vehicle and uses a parallel pointer mechanism to evaluate profile-based actions.

Our main contributions are:

- We introduce the Profiled Vehicle Routing Problem (PVRP), a generalization of HCVRP that incorporates vehicle profiles with client-specific preferences and operational constraints.
- We propose CAMP, a novel multi-agent reinforcement learning approach for PVRP that integrates vehicle and client profiles into an attention-based encoder-decoder framework for collaborative decision-making.
- We introduce a specialized attention-based communication architecture that processes profiled client representations per vehicle and enables cooperative decisions through a parallel pointer mechanism.
- We evaluate CAMP on multiple PVRP variants including PVRP with Preferences (PVRP-P) and PVRP with Zone Constraints (PVRP-ZC), demonstrating competitive performance against both classical methods and neural multi-agent models.

2 METHODOLOGY

2.1 Formulation of PVRP

Consider a VRP with node set N including depot (0) and clients, and vehicle set K. Each client i has demand d_i and location s_i, while each vehicle k has capacity Q_k, speed s_k, and a profile parameter vector p_k defining its relationship with each client. The objective minimizes the total travel cost while considering profile-specific preferences. We study two variants: (1) PVRP with Preferences (PVRP-P), where p_{ik} represents preference scores between vehicles and clients, balancing route efficiency with preferences using parameter α; and (2) PVRP with Zone Constraints (PVRP-ZC), where p_{ik} enforces zone-based restrictions by setting $p_{ik} = -\infty$ when vehicle k is prohibited from serving client i. Standard VRP constraints apply, including capacity limits, single-visit requirements, and flow conservation.

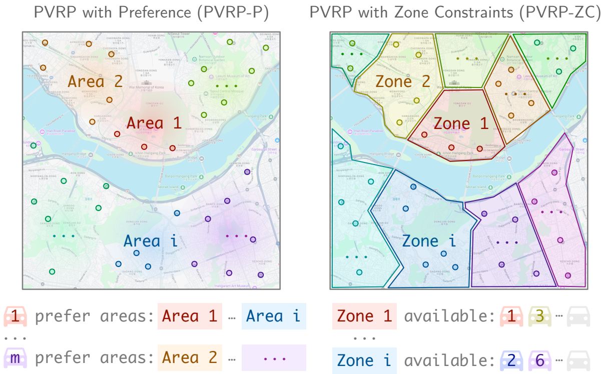
PVRP with Preference (PVRP-P)
PVRP with Zone Constraints (PVRP-ZC)
Area 2
Area 1
Area i
Zone 2
Zone 1
Zone i
1 prefer areas: Area 1 … Area i
m prefer areas: Area 2 … ...
Zone 1 available: 1 3 …
Zone i available: 2 6 …

Figure 1 – *Practical examples of PVRPs.*

2.2 Modeling PVRP with MARL

We formulate PVRP as a Markov Decision Process (MDP) $M = S, A, \tau, r$, where the state $s_t = (V_t, X_t)$ comprises vehicle states V_t and node states X_t. Each vehicle state v_t^k includes remaining capacity o_t^k, travel time T_t^k, partial route G_t^k, and profile parameters p_k. Node states x_t^i contain location s^i and remaining demand d_t^i.

2.2.1 Action Space and State Transitions

Actions $a_t = (v_t^k, x_t^j)$ represent vehicle-node assignments. The state transition function updates vehicle and node states accordingly: $s_{t+1} = (V_{t+1}, X_{t+1}) = \tau(V_t, X_t, a_t)$. For PVRP-P and PVRP-ZC variants, transitions differ in handling profile parameters p_{ik}: PVRP-P: p_{ik} represents preference scores; PVRP-ZC: p_{ik} is binary, masking invalid vehicle-client assignments.

2.2.2 Reward Structure and Optimization

The reward function provides terminal rewards based on variant-specific objectives:

$$R(s_T) = \sum_{k \in K} \sum_{(i,j) \in G_T^k} (\alpha p_{ik} - \frac{c_{ij}}{s_k}) \qquad \text{for PVRP-P} \tag{1}$$

$$R(s_T) = -\sum_{k \in K} \sum_{(i,j) \in G_T^k} \frac{c_{ij}}{s_k} \qquad \text{for PVRP-ZC} \tag{2}$$

Following the autoregressive sequence generation paradigm, we construct solutions by encoding problem instances $\boldsymbol{x}$ as $\boldsymbol{h} = f_\theta(\boldsymbol{x})$ and decoding actions sequentially:

$$\pi_\theta(\boldsymbol{a}|\boldsymbol{x}) = \prod_{t=1}^{T-1} g_\theta(a_t | a_{t-1}, ..., a_0, \boldsymbol{h}) \tag{3}$$

2.3 Collaborative Attention Model with Profiles

CAMP adopts a parallel autoregressive approach for PVRP, where multiple vehicles make decisions simultaneously. The model consists of a profile-aware encoder and a collaborative decoder,

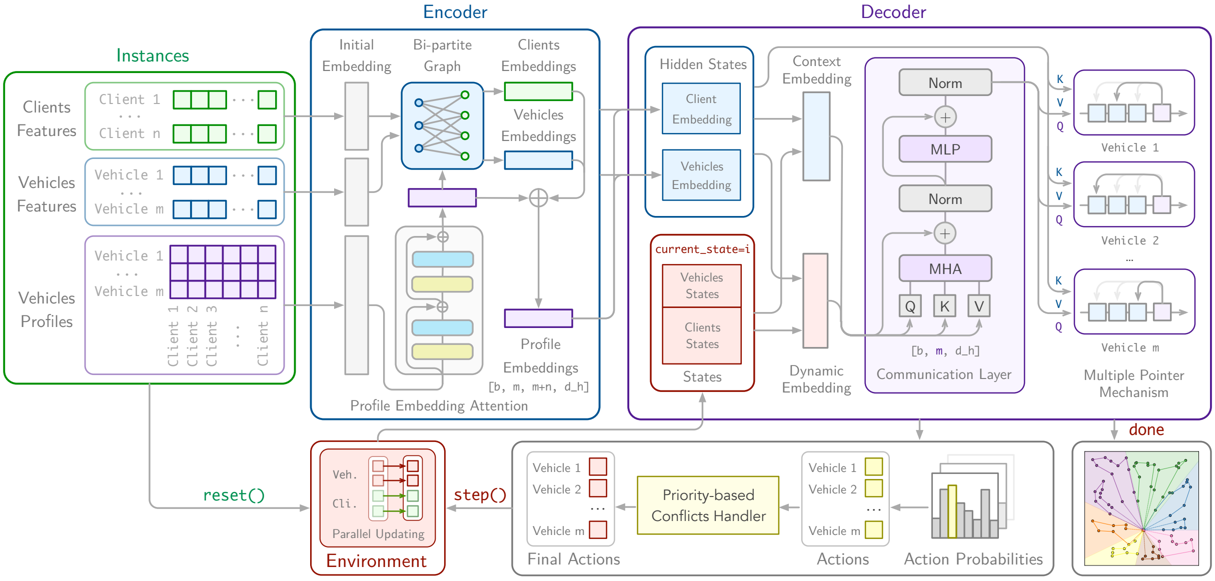

Figure 2 – *Overview of CAMP.*

designed to handle vehicle-specific preferences and constraints while enabling inter-vehicle communication.

The encoder processes vehicle profiles through three components: (1) generating initial embeddings for vehicles, clients, and their relationships using learned transformations to capture essential characteristics; (2) employing multi-head attention to process each vehicle's profile independently, maintaining distinct representations; and (3) utilizing bipartite graph message passing to enable information flow between vehicles and clients, creating interconnected representations while preserving unique features. The decoder operates in parallel through a specialized pointer mechanism: it first generates vehicle-specific queries incorporating both static and dynamic information, then enables inter-vehicle communication through transformer blocks for coordinated decision-making, and finally selects actions for all vehicles simultaneously using a conflict resolution strategy based on probability scores.

We train CAMP using REINFORCE with a shared baseline to optimize the model's performance across multiple vehicles. For PVRP-P, we introduce reward balancing to handle varying preference distributions, ensuring fair learning across different scenarios. This approach normalizes rewards dynamically during training, preventing bias towards specific preference patterns.

$$\nabla_\theta \mathcal{L} \approx \frac{1}{B \cdot L} \sum_{i=1}^{B} \sum_{j=1}^{L} G_{ij} \nabla_\theta \log p_\theta(A_{ij}|\boldsymbol{x}_i) \tag{4}$$

3 RESULTS AND DISCUSSION

We evaluate CAMP on both PVRP-P and PVRP-ZC variants across different preference distributions (Random, Angle, Cluster, and Zone) with varying numbers of vehicles and clients. We compare against state-of-the-art classical solvers PyVRP (Wouda *et al.*, 2024) and neural baselines: ET (Son *et al.*, 2024), a sequential action generator for multi-agent TSP; DPN (Zheng *et al.*, 2024), which enhances ET with improved route partitioning; 2D-Ptr (Liu *et al.*, 2024), using a dual-encoder system for HCVRP; and PARCO (Berto *et al.*, 2024), which enables fast parallel decision-making. Classical solvers run on 16 CPU cores with time limits of 5-10 seconds per instance, while neural methods are trained on a single RTX 4090 GPU for 100 epochs.

Fig. 3 illustrates the performance between routing costs and preference satisfaction across different models. For all preference distributions (Random, Angle, Cluster), CAMP consistently

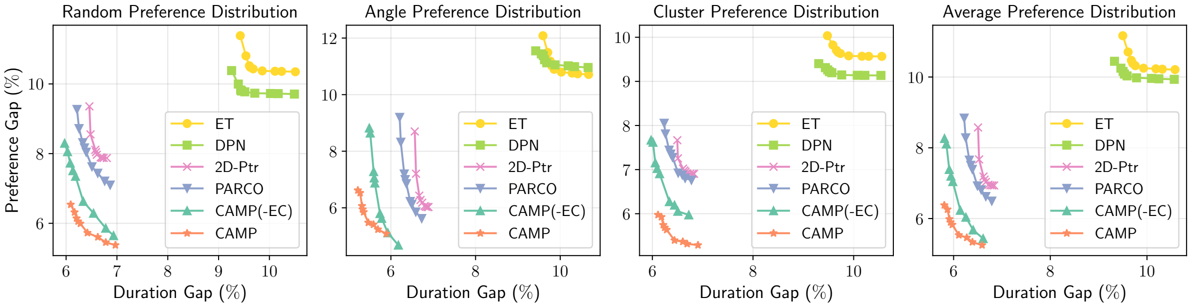

Figure 3 – *Pareto front of the cost of VRP-P against different values of α for different preference matrix distribution. The left bottom is better.*

achieves better Pareto frontiers compared to baselines, demonstrating superior optimization in both objectives. As the preference weight α increases from 0 to 0.2, CAMP maintains lower routing costs while achieving higher preference scores. This indicates our model's ability to effectively balance operational efficiency with preference satisfaction. Our model demonstrates particular strength in the high-α regime, where it achieves up to 15% lower costs than the best baseline while maintaining comparable preference scores. The performance advantage is most pronounced in structured preferences (Angle, Cluster), where CAMP's profile-aware architecture effectively captures and utilizes the underlying preference patterns. These results demonstrate that CAMP not only achieves state-of-the-art performance but also offers better scalability in handling complex preference structures.

4 CONCLUSION

We introduced the Profiled Vehicle Routing Problem (PVRP), extending HCVRP to incorporate client-specific preferences and operational constraints. Our proposed solution, the Collaborative Attention Model with Profiles (CAMP), leverages multi-agent reinforcement learning with profile-aware attention mechanisms to enable collaborative decision-making among heterogeneous vehicles. Experimental results on PVRP variants (PVRP-P and PVRP-ZC) demonstrate that CAMP consistently outperforms both traditional heuristics and neural baselines while maintaining computational efficiency, making it a practical tool for complex routing problems.

References

Berto, Federico, Hua, Chuanbo, Luttmann, Laurin, Son, Jiwoo, Park, Junyoung, Ahn, Kyuree, Kwon, Changhyun, Xie, Lin, & Park, Jinkyoo. 2024. PARCO: Learning Parallel Autoregressive Policies for Efficient Multi-Agent Combinatorial Optimization. *arXiv preprint arXiv:2409.03811*. https://github.com/ai4co/parco.

Liu, Qidong, Liu, Chaoyue, Niu, Shaoyao, Long, Cheng, Zhang, Jie, & Xu, Mingliang. 2024. 2D-Ptr: 2D Array Pointer Network for Solving the Heterogeneous Capacitated Vehicle Routing Problem. *Pages 1238–1246 of: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems.*

Son, Jiwoo, Kim, Minsu, Choi, Sanghyeok, Kim, Hyeonah, & Park, Jinkyoo. 2024. Equity-Transformer: Solving NP-Hard Min-Max Routing Problems as Sequential Generation with Equity Context. *Pages 20265–20273 of: Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38.

Wouda, Niels A, Lan, Leon, & Kool, Wouter. 2024. PyVRP: A high-performance VRP solver package. *INFORMS Journal on Computing.*

Zheng, Zhi, Yao, Shunyu, Wang, Zhenkun, Xialiang, Tong, Yuan, Mingxuan, & Tang, Ke. 2024. DPN: Decoupling Partition and Navigation for Neural Solvers of Min-max Vehicle Routing Problems. *In: Forty-first International Conference on Machine Learning.*