# Cover Letter

Dear Editor in Chief and Area Editor

Thank you very much for your decision letter dated 2 February 2025 on our manuscript ID 24-4059 entitled 'Deep reinforcement learning for solving two-echelon capacity vehicle routing problems: An end-to-end method'.

We are grateful for the constructive comments of the review team on our manuscript. We have improved the manuscript according to the reviewers' comments. Our point-by-point responses to the review comments are listed, respectively, in our replies to the reviewers.

The manuscript has been greatly improved by these revisions and we hope that you will now find it suitable for publication in IEEE Robotics and Automation Letters.

Thank you again and kind regards,

Zhi Pei, PhD
Professor
College of Mechanical Engineering, Zhejiang University of Technology

We are truly grateful to the editor and reviewers for the constructive comments and suggestions. We have now revised the paper in light of these comments. The revisions we have made in response to the editor and reviewers' comments are reported below (for easy reference, the original comments are listed in bold font).

=====================================================

### Response to comments from Referee 4

**This paper performs the study of Two-Echelon Capacitated Vehicle Routing Problem (2E-CVRP) by applying the Reinforcement Learning method. The paper uses a Transformer Neural Network Architecture to perform the learning of the optimization problem. The algorithm was tested on synthetic datasets with known distribution and benchmark datasets with unknown distributions from related works in the literature and shows that the solution speed surpasses the commercial solvers and its accuracy is better than the existing methods. While the method is novel in terms of its application to the 2E-CVRP, there are some minor issues that need to be addressed.**

**1. In Section 3B of page 4 of the paper, there is a verse that says each of nodes 2 to 9 is selected ...? Based upon Figure 1,shouldn't it be nodes 3 to 9 is selected in the second action instead of 2 to 9? This needs to be clarified.**

Response from the author

Thanks for the thoughtful suggestions. This reasoning is correct because a satellite node is typically followed by a customer node. We include satellite node 2 as a selectable node because a satellite can act as a "virtual" node. Specifically, "The next station after a satellite can be another satellite, but it will not be included in the cost calculation. This approach ensures that the initial starting point can be any satellite, allowing node 2 to be visited after node 1."

**2.In Section 5A of the paper, before the Table 1, there is a github link to look for the dataset. The link is not working. Please verify that before the submission. Similarly for the benchmark dataset.**

Response from the author

Thanks for the thoughtful suggestions. We apologize for not updating the link in time. It has now been replaced with the correct link.

**3. In Section 5B of page 6 of the paper, it is mentioned that there are 3 different types of DRL-2ECVRP is studied. One amongst them is DRL-2ECVRP(TL). It is not clear how the training is performed using this type. Does this method perform two rounds**

**of training? If so, how long does the overall training process time take before getting evaluated on test instances? This needs to be clarified.**

Response from the author

Thanks for the thoughtful suggestions. For DRL-2ECVRP(TL), the method essentially involves two rounds of training. The first round is problem-specific, while the second round fine-tunes the model weights for a particular instance. The time for the first round of training is provided in Table 1 and is not included in the instance-solving time. In contrast, the time for the second round of fine-tuning is included in the total solving time (see Table 2). In our revised manuscript, we added the following explanation in Section 5B of page 6:

"In other words, the second training phase refines the solution for that particular instance through several iterative steps, and the total solving time includes the duration of this second training phase."

**4. In Section 5C of page 7 of the paper, the benchmark dataset using which this performance analysis is made needs to be referenced in the paper.**

Response from the author

Thanks for the thoughtful suggestions. Thanks for the thoughtful suggestions. We have added references for both the benchmark datasets and the source of the best-known solutions. The revised content in Section 5C on page 7 is as follows:

"we still test our model on several benchmark datasets (Set2 [13], Set3, Set4b [13],Set5 [15]) using DRL-2ECVRP(TL)."

"The experimental results compared with the best-known solutions (BKS) [19] can be found at: https://github.com/sunweice/DRL-TECVRP."

**5. In Conclusion section of page 8 of the paper, while mentioning about future work, it would be better if the paper touches upon performing a study with a real life scenario in the future.**

Response from the author

Thanks for the thoughtful suggestions.We have added the following content to the Conclusion section.

"In future work, we plan to validate this DRL-based method in real-world scenarios with actual logistics data. This will further explore the model's robustness and adaptability, paving the way for broader practical applications in logistics and transportation."

==================================================================

## Response to comments from Referee 9

**This paper focuses on the Two-Echelon Vehicle Routing Problem (2E-VRP). It particularly examines the issue of computational overhead, where the time required for computation grows exponentially as the size of the problem increases in 2E-VRP. To address this, the paper introduces a solution using a deep reinforcement learning (DRL)-based approach for the 2E-VRP.**

**The formulation of the Markov Decision Process (MDP) appears to be robust and well-founded.**

**This represents the first application of a reinforcement learning (RL)-based method to solve the Two-Echelon Vehicle Routing Problem (2E-VRP). The outcomes are significant, demonstrating a considerable reduction in computation time while preserving high accuracy.**

**1. Further details could be beneficial in elucidating how transfer learning is employed to enhance the quality of the solution.**

Response from the author

Thanks for the thoughtful suggestions. For transfer learning, we simply use trained model weights—whether from a model of the same scale or a different scale—and then perform a few steps of fine-tuning on a single instance. More detailed explanations are provided in the paper.

"For the DRL-2ECVRP(TL), we use a pre-trained model of the same scale (Note: It's possible to use a model of a different scale,but this will result in some loss of accuracy.For example, you can use a 50-node model and perform transfer learning on a 100-node instance) as the test instance and perform an additional round of training during the solution process, In other words, the second training phase refines the solution for that particular instance through several iterative steps, and the total solving time includes the duration of this second training phase. "

**2. It is recommended that the evaluation includes a more diverse range of satellite quantities to enhance the robustness of the results.**

Response from the author

Thanks for the thoughtful suggestions. We conducted experiments by varying the number of satellites. Specifically, for the 51-node instance, we tested configurations with 3 and 7 satellites, while for the 100-node instance, we tested configurations with 8 and 12 satellites. The results(Table 1,2,3,4) demonstrate that our method remains effective with different numbers of satellites.Due to the length of the paper, This result are provided at the following link: https://github.com/sunweice/DRL-TECVRP.

Table 1: Results for the 2E-CVRP instances of Random Set(51 nodes and 3 satellites).

| Instance | Gurobi(10m) | | | DRL-2ECVRP(TL) | | |
|---|---|---|---|---|---|---|
| | B.O | T(s) | Gap | B.O | T(s) | Gap(%) |
| 51c_3s_0.pkl | 10.57 | 600.00 | 0.00 | 10.81 | 34.88 | 2.23 |
| 51c_3s__1.pkl | 8.83 | 600.00 | 0.00 | 8.97 | 33.48 | 1.64 |
| 51c_3s__2.pkl | 12.24 | 600.00 | 0.00 | 12.31 | 33.16 | 0.59 |
| 51c_3s__3.pkl | 9.21 | 600.00 | 0.00 | 9.44 | 34.57 | 2.45 |
| 51c_3s__4.pkl | 9.16 | 600.00 | 0.00 | 9.23 | 32.88 | 0.78 |
| 51c_3s__5.pkl | 11.49 | 600.00 | 0.00 | 11.57 | 33.43 | 0.65 |
| 51c_3s__6.pkl | 11.36 | 600.00 | 0.00 | 11.59 | 34.40 | 2.02 |
| 51c_3s__7.pkl | 11.92 | 600.00 | 0.00 | 12.05 | 33.82 | 1.12 |
| 51c_3s__8.pkl | 14.27 | 600.00 | 0.00 | 14.43 | 32.56 | 1.12 |
| 51c_3s__9.pkl | 10.29 | 600.00 | 0.00 | 10.30 | 34.61 | 0.13 |

Table 2: Results for the 2E-CVRP instances of Random Set(51 nodes and 7 satellites).

| Instance | Gurobi(10m) | | | DRL-2ECVRP(TL) | | |
|---|---|---|---|---|---|---|
| | B.O | T(s) | Gap(%) | B.O | T(s) | Gap(%) |
| 51c_7s_0.pkl | 9.09 | 600.00 | 1.25 | 8.98 | 34.88 | 0.00 |
| 51c_7s__1.pkl | 9.29 | 600.00 | 0.00 | 9.36 | 33.48 | 0.71 |
| 51c_7s__2.pkl | 9.87 | 600.00 | 0.00 | 10.07 | 33.16 | 2.00 |
| 51c_7s__3.pkl | 8.50 | 600.00 | 0.00 | 8.64 | 34.57 | 1.60 |
| 51c_7s__4.pkl | 8.20 | 600.00 | 1.22 | 8.10 | 32.88 | 0.00 |
| 51c_7s__5.pkl | 8.12 | 600.00 | 0.00 | 8.20 | 33.43 | 1.00 |
| 51c_7s__6.pkl | 10.05 | 600.00 | 0.16 | 10.03 | 34.40 | 0.00 |
| 51c_7s__7.pkl | 9.28 | 600.00 | 0.00 | 9.52 | 33.82 | 2.56 |
| 51c_7s__8.pkl | 8.92 | 600.00 | 1.29 | 8.81 | 32.56 | 0.00 |
| 51c_7s__9.pkl | 11.12 | 600.00 | 5.01 | 10.59 | 34.61 | 0.00 |

Table 3: Results for the 2E-CVRP instances of Random Set(100 nodes and 8 satellites).

| Instance | Gurobi(10m) | | | DRL-2ECVRP(TL) | | |
|---|---|---|---|---|---|---|
| | B.O | T(s) | Gap(%) | B.O | T(s) | Gap(%) |
| 100c_8s_0.pkl | 15.95 | 600.00 | 7.51 | 14.84 | 129.60 | 0.00 |
| 100c_8s__1.pkl | 16.04 | 600.00 | 5.18 | 15.25 | 130.70 | 0.00 |
| 100c_8s__2.pkl | 16.16 | 600.00 | 6.60 | 15.16 | 144.11 | 0.00 |
| 100c_8s__3.pkl | 16.04 | 600.00 | 4.21 | 15.39 | 138.02 | 0.00 |
| 100c_8s__4.pkl | 14.27 | 600.00 | 0.00 | 14.27 | 136.84 | 0.00 |
| 100c_8s__5.pkl | 16.16 | 600.00 | 5.00 | 15.39 | 142.18 | 0.00 |
| 100c_8s__6.pkl | 16.55 | 600.00 | 13.94 | 14.53 | 149.85 | 0.00 |
| 100c_8s__7.pkl | 16.18 | 600.00 | 8.00 | 14.98 | 148.60 | 0.00 |
| 100c_8s__8.pkl | 15.49 | 600.00 | 1.11 | 15.32 | 147.50 | 0.00 |
| 100c_8s__9.pkl | 14.65 | 600.00 | 8.35 | 13.52 | 136.90 | 0.00 |

Table 4: Results for the 2E-CVRP instances of Random Set(100 nodes and 12 satellites).

| Instance | Gurobi(10m) | | | DRL-2ECVRP(TL) | | |
|---|---|---|---|---|---|---|
| | B.O | T(s) | Gap(%) | B.O | T(s) | Gap(%) |
| 100c_12s_0.pkl | 16.36 | 600.00 | 25.57 | 13.03 | 137.43 | 0.00 |
| 100c_12s__1.pkl | 18.29 | 600.00 | 9.98 | 16.63 | 139.78 | 0.00 |
| 100c_12s__2.pkl | 14.59 | 600.00 | 6.89 | 13.65 | 141.92 | 0.00 |
| 100c_12s__3.pkl | 16.77 | 600.00 | 12.35 | 14.93 | 143.99 | 0.00 |
| 100c_12s__4.pkl | 14.43 | 600.00 | 4.83 | 13.76 | 142.15 | 0.00 |
| 100c_12s__5.pkl | 14.61 | 600.00 | 14.48 | 12.76 | 142.53 | 0.00 |
| 100c_12s__6.pkl | 14.32 | 600.00 | 6.07 | 13.50 | 139.51 | 0.00 |
| 100c_12s__7.pkl | 15.28 | 600.00 | 12.42 | 13.59 | 137.54 | 0.00 |
| 100c_12s__8.pkl | 15.98 | 600.00 | 9.25 | 14.63 | 134.87 | 0.00 |
| 100c_12s__9.pkl | 15.73 | 600.00 | 9.56 | 14.36 | 135.59 | 0.00 |

**3. It is recommended to consider expanding the study to include more than 100 customers. If this expansion is not feasible, an explanation of the constraints should be provided.**

Response from the author Thanks for the thoughtful suggestions. Large-scale instances consume significant training resources and time. However, our model's transferability allows us to use the pre-trained model from the 100-node scale to address these larger instances. We conducted additional experiments with 200 and 300 customers by directly applying the 100-node pre-trained model for transfer learning on these new instances. The results(Table 5,6) demonstrate that our method remains effective with different numbers of satellites.Moreover, as the number of nodes increases, our method's advantage over the solver becomes more pronounced. Due to the length of the paper, This results are provided at the following link: https://github.com/sunweice/DRL-TECVRP

Table 5: Results for the 2E-CVRP instances of Random Set(200 nodes and 10 satellites).

| Instance | Gurobi(10m) | | | DRL-2ECVRP(TL) | | |
|---|---|---|---|---|---|---|
| | B.O | T(s) | Gap(%) | B.O | T(s) | Gap(%) |
| 200c_10s_0.pkl | 190.38 | 600.00 | 656.68 | 25.16 | 255.94 | 0.00 |
| 200c_10s__1.pkl | 121.44 | 600.00 | 390.29 | 24.77 | 259.81 | 0.00 |
| 200c_10s__2.pkl | 26.96 | 600.00 | 3.00 | 26.17 | 254.29 | 0.00 |
| 200c_10s__3.pkl | 29.41 | 600.00 | 22.44 | 24.02 | 253.86 | 0.00 |
| 200c_10s__4.pkl | 29.97 | 600.00 | 9.79 | 27.30 | 253.34 | 0.00 |
| 200c_10s__5.pkl | 26.22 | 600.00 | 13.07 | 23.19 | 250.75 | 0.00 |
| 200c_10s__6.pkl | 24.59 | 600.00 | 0.00 | 25.26 | 252.26 | 2.72 |
| 200c_10s__7.pkl | 34.31 | 600.00 | 25.46 | 27.35 | 256.50 | 0.00 |
| 200c_10s__8.pkl | 30.60 | 600.00 | 4.79 | 29.20 | 254.55 | 0.00 |
| 200c_10s__9.pkl | 27.58 | 600.00 | 8.49 | 25.42 | 256.15 | 0.00 |

Table 6: Results for the 2E-CVRP instances of Random Set(300 nodes and 15 satellites).

| Instance | Gurobi(10m) | | | DRL-2ECVRP(TL) | | |
|---|---|---|---|---|---|---|
| | B.O | T(s) | Gap(%) | B.O | T(s) | Gap(%) |
| 300c_15s_0.pkl | - | 600.00 | - | 37.65 | 600.00 | 0.00 |
| 300c_15s__1.pkl | - | 600.00 | - | 33.95 | 600.00 | 0.00 |
| 300c_15s__2.pkl | - | 600.00 | - | 36.26 | 600.00 | 0.00 |
| 300c_15s__3.pkl | - | 600.00 | - | 36.14 | 600.00 | 0.00 |
| 300c_15s__4.pkl | - | 600.00 | - | 33.70 | 600.00 | 0.00 |
| 300c_15s__5.pkl | 201.48 | 600.00 | 406.88 | 39.75 | 600.00 | 0.00 |
| 300c_15s__6.pkl | - | 600.00 | - | 33.17 | 600.00 | 0.00 |
| 300c_15s__7.pkl | 275.33 | 600.00 | 632.83 | 37.57 | 600.00 | 0.00 |
| 300c_15s__8.pkl | - | 600.00 | - | 34.88 | 600.00 | 0.00 |
| 300c_15s__9.pkl | 224.15 | 600.00 | 518.68 | 36.23 | 600.00 | 0.00 |

Once again, we appreciate very much the referees valuable comments and suggestions, and sincerely hope that you find this new version of the manuscript satisfactory.