SCIENCE CHINA
Information Sciences

• LETTER •

August 2026, Vol. 69, Iss. 8, 189202:1–189202:2
https://doi.org/10.1007/s11432-025-4796-2

DPCN: enabling efficient multi-agent path finding via decentralized-planning-centralized-negotiation paradigm

Qidong LIU[1], Xiaowen LI[1*], Chuang ZHAN[1], Jie ZHANG[2], Cheng LONG[2] & Mingliang XU[1]

[1]*School of Computer Science and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China*
[2]*College of Computing and Data Science, Nanyang Technological University, Singapore 639798, Singapore*

Received 17 April 2025/Revised 8 October 2025/Accepted 24 November 2025/Published online 6 July 2026

Citation Liu Q D, Li X W, Zhan C, et al. DPCN: enabling efficient multi-agent path finding via decentralized-planning-centralized-negotiation paradigm. Sci China Inf Sci, 2026, 69(8): 189202, https://doi.org/10.1007/s11432-025-4796-2

Multi-agent path finding (MAPF) is a computational problem that involves finding collision-free paths for a set of agents operating in a shared environment such that each agent can reach its destination without colliding with other agents or static obstacles. Traditional planners are typically centralized, relying on a central controller to globally coordinate the paths of all agents. This approach results in limited scalability and increased computational complexity when managing large-scale teams.

With the advancement of reinforcement learning (RL), an increasing number of studies have modeled the MAPF problem as a Markov decision process. These methods are typically decentralized, allowing each agent to dynamically interact with the environment and make decisions based on information within its local field of view (FOV). Although this strategy significantly reduces computational complexity, it poses challenges in effectively coordinating the behaviors of agents, which may result in conflicts. To solve conflicts, existing studies [1, 2] suggest prioritizing agents to coordinate their actions. However, these approaches, which usually rely on fixed or heuristic-based priorities, are not only time-consuming but also inflexible.

In this study, we introduce a learnable framework to address the above problems through a decentralized-planning-centralized-negotiation (DPCN) paradigm, as illustrated in Figure 1(a). In the planning phase (details in Appendix A), we model each automated guided vehicle (AGV) as an agent, which determines its intention based solely on its local observation, without considering the implications for other agents. During the negotiation phase (details in Appendix B), we identify all conflict sets and model each one as a super-agent. A policy network is then proposed to output a priority ranking for the agents within the conflict set, taking into account their individual intentions and the real-time environmental conditions.

Nevertheless, conflict sets would vary across timesteps, and the agents within each conflict set differ, leading to dynamically changing super-agents, each with a unique set of elements. This poses two challenges: (1) inconsistent action sets, where the super-agent's action is not drawn from a fixed set but is instead generated by its constituent internal agents; and (2) dynamic super-agent training, where the dynamic evolution of super-agents over time prevents the assignment of a dedicated policy network π_k and action-value function Q_k per super-agent, posing a major training challenge. To address these challenges, we introduce a special-edition pointer network (PNSE), an enhanced model inspired by the pointer network [3], and train it using a customized policy gradient reinforcement learning algorithm.

PNSE. The network architecture of the PNSE is shown in Figure 1(b). Taking the super-agent Ω_k as an example, we assume there are m ($2 \leqslant m \leqslant 5$) agents within it. For each agent i, the local observation $\mathbf{o}_i \in \mathbb{R}^{3\times3\times8}$ is encoded into $\hat{\mathbf{o}}_i \in \mathbb{R}^{d_o}$ by an observation encoder. Furthermore, the intention $\tilde{a}_i \in \mathbb{R}^1$ obtained from Eq. (A2) in Appendix A is embedded into $\hat{\mathbf{a}}_i \in \mathbb{R}^{d_e}$. We then concatenate $\hat{\mathbf{o}}_i$ and $\hat{\mathbf{a}}_i$ for agent i, yielding the feature $\tilde{\mathbf{z}}_i \in \mathbb{R}^{d_o+d_e}$, i.e., $\tilde{\mathbf{z}}_i = Concat(\hat{\mathbf{o}}_i, \hat{\mathbf{a}}_i)$. The joint feature of all agents within the super-agent is denoted by $\tilde{\mathbf{Z}} \in \mathbb{R}^{m\times(d_o+d_e)}$.

Next, the self-attention mechanism is used to enable agents to perceive each other, producing the feature $\mathbf{Z}'$. To stabilize training and expedite convergence, we introduce the residual connections and layer normalization. The final feature $\mathbf{Z}$ is computed as

$$\mathbf{Z} = f^{LN}(f^{LN}(\tilde{\mathbf{Z}} + \mathbf{Z}') + f^{MLP}(f^{LN}(\tilde{\mathbf{Z}} + \mathbf{Z}'))), \tag{1}$$

where f^{LN} represents layer normalization and f^{MLP} represents the MLP.

Centralized negotiation involves analyzing the potential impact of agents' intentions within the conflict set based on the current state, and then selecting a winner to execute its intention. Inspired by the pointer network, we get a probability distribution ψ:

$$\psi = softmax\frac{\mathbf{G}\mathbf{W}_{\mathbf{G}}(\mathbf{Z}\mathbf{W}_{\mathbf{Z}})^{\mathrm{T}}}{\sqrt{d_z}}, \tag{2}$$

where $\mathbf{W}_{\mathbf{G}} \in \mathbb{R}^{d_z\times d_z}$ and $\mathbf{W}_{\mathbf{Z}} \in \mathbb{R}^{d_z\times d_z}$ are learnable weight matrices. $\mathbf{G} \in \mathbb{R}^{1\times d_z}$ is the global state representation derived from all agents' local observations, computed as

$$\mathbf{G} = f^{AP}(f^{MLP}(Concat(c_1, \cdots, c_n))). \tag{3}$$

Here, $c_i \in \mathbb{R}^{d_g}$ is obtained by Eq. (A1) in Appendix A, representing the feature of the local state observed by agent i. f^{AP} represents the average pooling layer.

Training. For training this reinforcement learning task, we employ the policy gradient method to maximize the objective function. At timestep t, we consider a game involving K super-agents,

* Corresponding author (email: xiaowli@gs.zzu.edu.cn)

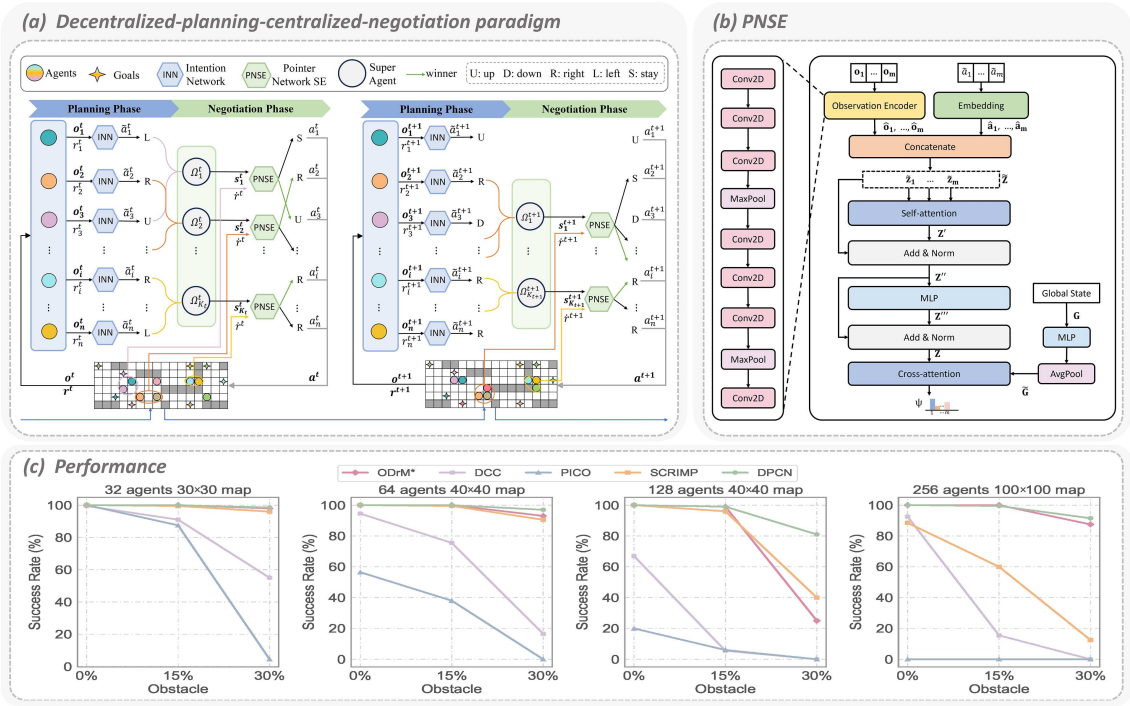

Figure 1 (Color online) Overview of DPCN. (a) DPCN comprises two phases: a planning phase where each agent defines its intention based on local observations, and a negotiation phase where PNSE resolves conflicts; (b) PNSE is proposed to output a priority ranking for the agents within the conflict set, taking into account their individual intentions and the real-time environmental conditions; (c) extensive experimental evaluation demonstrates that our method can efficiently resolve conflicts, thereby improving the planner's execution efficiency and solution quality.

each with policies parameterized by $\theta = \{\theta_1, \cdots, \theta_K\}$, and let $\pi = \{\pi_1, \cdots, \pi_K\}$ denote the set of all agents' policies. The gradient of the expected return for super-agent k, $J(\theta_k) = \mathbb{E}[R_k]$, can be expressed as

$$\nabla_\theta J(\theta) = \mathbb{E}_{v_k \sim \pi}\left[\nabla_\theta \log \pi(v_k|s_k)\frac{1}{K}\sum_{t'=t}^{T}\left(\gamma^{t'-t} r_{all}^{t'}\right)\right]. \quad (4)$$

The parameters θ can be updated using the gradient ascent rule $\theta \leftarrow \theta + \alpha\nabla_\theta J(\theta)$, where α is the learning rate. The proof of Eq. (4) can be found in Appendix C.

Experiment. From the experimental results presented in Figure 1(c), we draw the following conclusion. (1) Centralized planners (e.g., ODrM*) perform well with sparsely distributed agents but degrade in dense/obstacle-rich environments, owing to increased conflicts and excessive computational load on the central controller. (2) RL-based methods (e.g., SCRIMP [4], DCC [2], and PICO [5]) typically rely on fixed or heuristic-based priorities, which are both time-intensive and rigid, ultimately resulting in performance degradation. (3) DPCN sustains consistently high performance across all scenarios, owing to its robust inter-agent coordination mechanism and PNSE-enabled conflict resolution strategy. Details of the experimental setup and further experimental results can be found in Appendix D.

Conclusion. In this study, we introduce a decentralized-planning-centralized-negotiation paradigm that uses neural networks to automatically learn priorities for conflict resolution. Our approach significantly enhances the robustness of the MAPF planner by efficiently resolving conflicts among agents. Specifically, we model each conflict set as a super-agent, and propose the PNSE as the policy network for selecting the winning agent, thereby resolving the conflict. Furthermore, we develop a customized RL training scheme tailored to dynamically evolving super-agents. Notably, our PNSE effectively tackles the challenge of inconsistent action sets, presenting a novel approach for MAPF problems characterized by heterogeneous action sets. In the future, we aim to investigate the potential applications of this algorithm in the realm of heterogeneous MAPF.

Acknowledgements This work was supported by Natural Science Foundation of Henan (Grant No. 232300421095), Foundation of Henan Educational Committee (Grant No. 25HASTIT034), and National Natural Science Foundation of China (Grant Nos. 62276238, U24A20326, 62325602, 62036010).

Supporting information Appendixes A–D. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Gao J, Li Y, Ye Z, et al. PCE: multi-agent path finding via priority-aware communication & experience learning. IEEE Trans Intell Veh, 2024, doi: 10.1109/TIV.2024.3363179
2 Ma Z, Luo Y, Pan J. Learning selective communication for multi-agent path finding. IEEE Robot Autom Lett, 2021, 7: 1455–1462
3 Vinyals O, Fortunato M, Jaitly N. Pointer networks. In: Proceedings of Advances in Neural Information Processing Systems, 2015. 28
4 Wang Y, Xiang B, Huang S, et al. Scrimp: scalable communication for reinforcement- and imitation-learning-based multi-agent pathfinding. In: Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, 2023. 2598–2600
5 Li W, Chen H, Jin B, et al. Multi-agent path finding with prioritized communication learning. In: Proceedings of International Conference on Robotics and Automation (ICRA), 2022. 10695–10701