RBC: Retroactive Belief State Compensation for Multi-Agent Collaboration Under Information Delay

Extended Abstract

Dongkun Huo
School of Computer Science and Technology, Huazhong University of Science And Technology
Wuhan, China
dongkunhuo@hust.edu.cn

Hongbo Liu
School of Computer Science and Technology, Huazhong University of Science And Technology
Wuhan, China
loren5555@hust.edu.cn

Shu Yin
School of Computer Science and Technology, Huazhong University of Science And Technology
Wuhan, China
shuyin@hust.edu.cn

Yixue Hao †
School of Computer Science and Technology, Huazhong University of Science And Technology
Wuhan, China
yixuehao@hust.edu.cn

Long Hu
School of Computer Science and Technology, Huazhong University of Science And Technology
Wuhan, China
hulong@hust.edu.cn

Rui Wang
School of Computer Science and Technology, Huazhong University of Science And Technology
Wuhan, China
ruiwang2020@hust.edu.cn

Min Chen
School of Computer Science and Engineering, South China University of Technology
Guangzhou, China
minchen@ieee.org

ABSTRACT

Real-time information is usually not satisfied in real world due to communication or observation delay. Although existing works address individual delay, they do not fully consider the complex effects of composite delay, denoted as "Information Delay", which severely reduce the efficiency of these methods. To address information delay, we propose **R**etroactive **B**elief state **C**ompensation (RBC), a multi-agent framework with enhanced robustness and collaboration. Specifically, we design a multi-step reconstruction model that retroactively rebuilds agents' belief states starting from the generation time of the information. This process corrects the accumulated deviation in the current belief state caused by information delay. Moreover, to enhance proactive collaboration, we introduce an intent inference module. This module enables agents to generate intents, which represent short-term action plans, as content of communication. By aggregating intents from teammates, agents will choose more coherent and synchronized joint actions. To evaluate the performance of RBC, we design scenarios with multiple levels of observation, communication, and composite delays. Experimental results demonstrate that RBC outperforms the baselines in all scenarios with delays.

Yixue Hao is corresponding author. Email: yixuehao@hust.edu.cn.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/MFLP4403

KEYWORDS

Multi-Agent; Reinforcement Learning; Information Delay; Collaboration

ACM Reference Format:
Dongkun Huo, Hongbo Liu, Shu Yin, Yixue Hao †, Long Hu, Rui Wang, and Min Chen. 2026. RBC: Retroactive Belief State Compensation for Multi-Agent Collaboration Under Information Delay: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/MFLP4403

1 INTRODUCTION

Multi-Agent Reinforcement Learning (MARL) shows potential [7, 9, 17] but struggles with partial observability [2, 8, 14, 15]. While belief states mitigate this, real-world delays render information outdated, violating the Markov property and causing learning instability [3]. Existing works address observation [3, 13] and communication delays [11, 16] in isolation, overlooking their compound effects. We define this holistic challenge as "Information Delay".

To address this, we propose **R**etroactive **B**elief state **C**ompensation (RBC). RBC utilizes a multi-step reconstruction module to retroactively refine latent trajectories, ensuring agents operate on up-to-date belief states. Additionally, we introduce an intent inference module to deduce teammates' current intentions from delayed messages. Experiments on SMAC and SMACv2 demonstrate that RBC significantly outperforms baselines under various delay conditions.

2 METHOD

We propose the **R**etroactive **B**elief state **C**ompensation (RBC) framework to mitigate coordination degradation caused by information delay. RBC reconstructs belief states by retroactively refining delayed observations and inferring teammates' intents.

2.1 Retroactive Belief Compensation

Let agent i's input at time t be $x_{i,t} = \{o_{i,t}, m_{ji,t}\}_{j \in \mathcal{N}}$. Due to asynchronous transmission, the received input is delayed as $\tilde{x}_{i,t}$. RBC comprises three components:

Variational Encoder: To encode $\tilde{x}_{i,t}$ into a compact belief $b_{i,t} = (h_{i,t}, z_{i,t})$, we employ a Recurrent State-Space Model [5]. A GRU updates the deterministic state $h_{i,t} = f_{GRU}(h_{i,t-1}, z_{i,t-1}, a_{i,t-1})$ [4]. A VAE then samples a stochastic latent variable $z_{i,t} \sim q_\phi(z_{i,t}|h_{i,t}, x_{i,t})$ [6]. The encoder is trained via reconstruction loss:

$$\mathcal{L}_{recon} = \mathbb{E}_{q_\phi}\left[-\log p_\theta(x_{i,t}|h_{i,t}, z_{i,t})\right]. \tag{1}$$

Retroactive Cell: To handle delayed inputs, we introduce a latent projector $q_\phi^{proj}(z_{i,t}|h_{i,t})$ to estimate the current state from history. It minimizes the KL divergence with the posterior:

$$\mathcal{L}_{proj} = \mathbb{E}\left[D_{KL}\left(q_\phi(\cdot|h_{i,t}, x_{i,t})||q_\phi^{proj}(\cdot|h_{i,t})\right)\right]. \tag{2}$$

During execution, if $x_{i,t}$ is delayed, the agent samples $\hat{z}_{i,t} \sim q_\phi^{proj}$ to form an estimated belief $\hat{b}_{i,t}$.

Retroactive Reconstruction: Upon receiving delayed information $\tilde{x}_{i,t}$ with max delay d, RBC refines speculative beliefs. The module reverts the belief to $t-d$ and recursively re-simulates the trajectory to $t-1$. At each step k, the state is updated as:

$$\begin{aligned} h'_{i,k+1} &= f_{GRU}\left(h'_{i,k},\, z'_{i,k},\, a_{i,k}\right), \\ z'_{i,k+1} &\sim \begin{cases} q_\phi\left(z \mid h'_{i,k+1}, x_{i,k+1}\right), & x_{i,k+1} \text{ arrived}, \\ q_\phi^{proj}\left(z \mid h'_{i,k+1}\right), & x_{i,k+1} \text{ delayed}. \end{cases} \end{aligned} \tag{3}$$

This process yields an updated belief that integrates both delayed ground truth and current estimates, reducing error accumulation.

2.2 Collaboration Enhancement via Intent

To facilitate foresighted cooperation, agents exchange high-level intents derived from their belief states.

Intent Generation: We define an Intent Encoder $p_\varphi(e_{i,t}|b_{i,t})$ to generate intent $e_{i,t}$. To ensure relevance to actions, an Intent Inference module $q_\epsilon(e_{i,t}|b_{i,t}, a_{i,t})$ infers intent from the taken action $a_{i,t}$. Following the Information Bottleneck principle [1], we optimize:

$$\mathcal{L}_{intent} = \mathbb{E}\left[D_{KL}\left(q_\epsilon(e_{i,t}|b_{i,t}, a_{i,t})||p_\varphi(e_{i,t}|b_{i,t})\right)\right]. \tag{4}$$

Message Aggregation: Agents aggregate teammates' intents using an attention mechanism where local intent $e_{i,t}$ serves as the query. The aggregated intent $e_{i,t}^{aggr}$ is computed as:

$$e_{i,t}^{aggr} = \sum_{j \neq i} \text{softmax}_j\left(\frac{(W_Q e_{i,t})^T (W_K e'_{j,t})}{\sqrt{d_k}}\right)(W_V e'_{j,t}). \tag{5}$$

The final belief state, $b_{i,t} = \text{concat}(h_{i,t}, z_{i,t}, e_{i,t} + e_{i,t}^{aggr})$, combines perception and collaborative intent to guide the policy π_i.

2.3 Overall Optimization Objective

RBC is trained under the CTDE paradigm. The basic network minimizes the standard TD loss $\mathcal{L}_{TD}$ [12] using a mixing network (e.g., QMIX [10]). The total objective combines all losses with weights λ:

$$\mathcal{L}_{total} = \mathcal{L}_{TD} + \lambda_o \mathcal{L}_{recon} + \lambda_p \mathcal{L}_{proj} + \lambda_e \mathcal{L}_{intent}. \tag{6}$$

2.4 Experimental Analysis

We evaluate RBC on representative maps 5m_vs_6m, MMM2, and Protoss_5_vs_5. In delay-free environments, RBC matches or surpasses leading baselines, exhibiting faster convergence and reduced variance. This success is attributed to the RSSM-based latent encoding and intent inference, which effectively capture environment dynamics and promote consistent coordination. We further assess performance under severe Information Delay, characterized by observation delay $\mathcal{N}(1,1)$ and communication delay $\mathcal{N}(5,2)$. While baseline methods suffer significant performance degradation and widened uncertainty bands due to asynchronous inputs, RBC demonstrates strong robustness with only slight declines in win rates. The framework's retroactive reconstruction and intent aggregation mechanisms work in synergy to maintain accurate, real-time belief states despite severe lags. Consequently, RBC effectively preserves formation control and target selection, achieving superior performance compared to baselines in these challenging, highly asynchronous scenarios.

Table 1: Performance under Information Delay

Delay	Algo	5m_vs_6m	MMM2	P_5_vs_5
No Delay	MAIC	81.04 ± 4.3*	94.48 ± 2.6*	43.82 ± 5.0
	T2MAC	43.75 ± 6.1	0.27 ± 0.5	38.78 ± 3.6
	CACOM	75.69 ± 4.5	90.73 ± 2.8	47.53 ± 5.8†
	RBC	80.1 ± 4.3†	92.88 ± 5.1†	48.67 ± 9.4*
Delayed	MAIC	0.97 ± 1.0	28.09 ± 5.2†	41.56 ± 4.5
	T2MAC	0.42 ± 0.7	0.0 ± 0.0	37.58 ± 4.9
	CACOM	8.85 ± 3.4†	17.08 ± 4.0	44.86 ± 4.4†
	RBC	9.33 ± 3.1*	62.21 ± 7.3*	46.72 ± 6.5*

3 CONCLUSION

We propose Retroactive Belief State Compensation (RBC) to address the degradation of belief state accuracy caused by information delay. RBC utilizes a recurrent state space model to retroactively reconstruct belief states from historical trajectories. To enhance collaboration, we introduce an intent inference module that enables agents to perceive teammates' short-term plans from received messages. Experiments demonstrate RBC's superior robustness across various delay settings, offering a reliable solution for asynchronous multi-agent learning.

ACKNOWLEDGMENTS

This work is supported by Wuhan Natural Science Foundation Exploratory Program (Chenguang Program) under Grant 20240408-01020212.

REFERENCES

[1] Alexander A Alemi, Ian Fischer, Joshua V Dillon, and Kevin Murphy. 2017. Deep Variational Information Bottleneck. In *International Conference on Learning Representations*. ICLR.

[2] Yanwen Ba, Xuan Liu, Xinning Chen, Hao Wang, Yang Xu, Kenli Li, and Shigeng Zhang. 2024. Cautiously-Optimistic Knowledge Sharing for Cooperative Multi-Agent Reinforcement Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. AAAI, 17299–17307.

[3] Baiming Chen, Mengdi Xu, Zuxin Liu, Liang Li, and Ding Zhao. 2020. Delay-aware Multi-Agent Reinforcement Learning for Cooperative and Competitive Environments. *arXiv preprint arXiv:2005.05441* (2020).

[4] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, Dekai Wu, Marine Carpuat, Xavier Carreras, and Eva Maria Vecchi (Eds.). Association for Computational Linguistics, Doha, Qatar, 103–111. https://doi.org/10.3115/v1/W14-4012

[5] Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. 2019. Learning Latent Dynamics for Planning from Pixels. In *International conference on machine learning*. PMLR, 2555–2565.

[6] Diederik P Kingma and Max Welling. 2013. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114* (2013).

[7] Meng Li, Zehong Cao, and Zhibin Li. 2021. A Reinforcement Learning-based Vehicle Platoon Control Strategy for Reducing Energy Consumption in Traffic Oscillations. *IEEE Transactions on Neural Networks and Learning Systems* 32, 12 (2021), 5309–5322.

[8] Yiming Li, Shunli Ren, Pengxiang Wu, Siheng Chen, Chen Feng, and Wenjun Zhang. 2021. Learning Distilled Collaboration Graph for Multi-Agent Perception. In *In Proceedings of Advances in Neural Information Processing Systems*, Vol. 34. NIPS, 29541–29552.

[9] Yuanguo Lin, Yong Liu, Fan Lin, Lixin Zou, Pengcheng Wu, Wenhua Zeng, Huanhuan Chen, and Chunyan Miao. 2023. A Survey on Reinforcement Learning for Recommender Systems. *IEEE Transactions on Neural Networks and Learning Systems* 35, 10 (2023), 13164–13184.

[10] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2020. Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. *Journal of Machine Learning Research* 21, 178 (2020), 1–51.

[11] Shoucheng Song, Youfang Lin, Sheng Han, Chang Yao, Hao Wu, Shuo Wang, and Kai Lv. 2025. CoDe: Communication Delay-Tolerant Multi-Agent Collaboration via Dual Alignment of Intent and Timeliness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 23304–23312.

[12] Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. 1999. Policy Gradient Methods for Reinforcement Learning with Function Approximation. In *Advances in Neural Information Processing Systems*, S. Solla, T. Leen, and K. Müller (Eds.), Vol. 12. MIT Press. https://proceedings.neurips.cc/paper_files/paper/1999/file/464d828b85b0bed98e80ade0a5c43b0f-Paper.pdf

[13] Thomas J Walsh, Ali Nouri, Lihong Li, and Michael L Littman. 2009. Learning and Planning in Environments with Delayed Feedback. *Autonomous Agents and Multi-Agent Systems* 18, 1 (2009), 83–105.

[14] Tonghan Wang, Heng Dong, Victor Lesser, and Chongjie Zhang. 2020. ROMA: Multi-Agent Reinforcement Learning with Emergent Roles. In *International Conference on Machine Learning*. PMLR, 9876–9886.

[15] Muning Wen, Jakub Kuba, Runji Lin, Weinan Zhang, Ying Wen, Jun Wang, and Yaodong Yang. 2022. Multi-Agent Reinforcement Learning is a Sequence Modeling Problem. In *In Proceedings of Advances in Neural Information Processing Systems*, Vol. 35. NIPS, 16509–16521.

[16] Tingting Yuan, Hwei-Ming Chung, Jie Yuan, and Xiaoming Fu. 2023. DACOM: Learning Delay-aware Communication for Multi-Agent Reinforcement Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 11763–11771.

[17] Weijia Zhang, Hao Liu, Hui Xiong, Tong Xu, Fan Wang, Haoran Xin, and Hua Wu. 2022. RLCharge: Imitative Multi-Agent Spatiotemporal Reinforcement Learning for Electric Vehicle Charging Station Recommendation. *IEEE Transactions on Knowledge and Data Engineering* 35, 6 (2022), 6290–6304.