Repeated Deceptive Path Planning against Learnable Observer

Extended Abstract

Shiyue Cao
School of Artificial Intelligence, University of Chinese Academy of Sciences & Institute of Automation, Chinese Academy of Sciences
Beijing, China
caoshiyue2021@ia.ac.cn

Pei Xu
Institute of Automation, Chinese Academy of Sciences
Beijing, China
pei.xu@ia.ac.cn

Likun Yang
University of Chinese Academy of Sciences
Beijing, China
yanglikun2021@ia.ac.cn

Lei Cui
University of Chinese Academy of Sciences
Beijing, China
cuilei2024@ia.ac.cn

Shizhao Yu
University of Chinese Academy of Sciences
Beijing, China
yushizhao2022@ia.ac.cn

Shiyu Zhang
Institute of Automation, Chinese Academy of Sciences
Beijing, China
shiyu.zhang@ia.ac.cn

Yongjian Ren
University of Chinese Academy of Sciences
Beijing, China
renyongjian2022@ia.ac.cn

Xiaotang Chen
Institute of Automation, Chinese Academy of Sciences
Beijing, China
xtchen@nlpr.ia.ac.cn

Kaiqi Huang
School of Artificial Intelligence, University of Chinese Academy of Sciences & Institute of Automation, Chinese Academy of Sciences
Beijing, China
kaiqi.huang@nlpr.ia.ac.cn

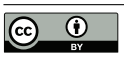
ABSTRACT

We introduce Repeated Deceptive Path Planning (RDPP), a novel setting where an agent must conceal its destination from a learnable observer that can adapt from historical trajectories. We show that existing deceptive planning methods, designed for static observers, fail in RDPP due to accumulated adaptation lag. To address this, we propose Deceptive Meta Planning (DeMP), a two-level optimization framework that anticipates and counteracts observer updates across episodes via meta-level learning. Experiments demonstrate that DeMP significantly outperforms traditional methods, enabling sustained deception against learning adversaries while maintaining efficient path costs.

KEYWORDS

Deceptive Path Planning; Goal Recognition; Reinforcement Learning

ACM Reference Format:
Shiyue Cao, Pei Xu, Likun Yang, Lei Cui, Shizhao Yu, Shiyu Zhang, Yongjian Ren, Xiaotang Chen, and Kaiqi Huang. 2026. Repeated Deceptive Path Planning against Learnable Observer : 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/DCBH3506



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/DCBH3506

1 INTRODUCTION

Deceptive Path Planning (DPP) studies how an agent can reach its true destination while preventing an external observer from correctly inferring its goal, a problem arising in security-sensitive applications such as transportation and military operations [1, 3, 6]. Existing DPP methods typically assume one-shot interactions with static observers [4].

In this work, we consider a more realistic setting, *Repeated Deceptive Path Planning* (RDPP), where an agent interacts repeatedly with a *learnable* observer that improves its goal prediction model using historical trajectories. Such repeated interactions have been shown to significantly degrade the effectiveness of deceptive strategies designed for static observers [5], highlighting the challenge of sustaining deception under repeated interactions with an adaptive adversary.

To address RDPP, we propose **Deceptive Meta Planning (DeMP)**, a meta-level planning framework that explicitly accounts for the observer's learning behavior across episodes, enabling sustained deception against adaptive observers.

2 METHOD

2.1 Modeling Repeated Deceptive Path Planning

We define a Repeated Deceptive Path Planning (RDPP) problem as $M_{RDPP} = (M_{DPP}, \Phi, K)$, where M_{DPP} denotes a deceptive MDP [5], Φ is the parameter space of the observer's learnable model, and K is the number of repeated interactions. In episode k, the agent executes a deceptive policy in M_{DPP}, generating a trajectory $\tau^{(k)} = \{(s_t, a_t)\}_{t=0}^{T_k}$. The observer only observes a partial prefix $\zeta^{(k)} =$

$\tau^{(k)}_{0:\lfloor \alpha T_k \rfloor}$ and produces a predictive distribution over goals

$$O(\zeta^{(k)}; \phi^{(k)}) = P(G \mid \zeta^{(k)}; \phi^{(k)}).$$

After the episode, the agent receives $O(\zeta^{(k)}; \phi^{(k)})$ as feedback, while the observer observes the true goal G^* and the full trajectory $\tau^{(k)}$, and updates its parameters via

$$\phi^{(k+1)} \leftarrow \mathcal{U}(\phi^{(k)}, \tau^{(k)}, G^*),$$

where $\mathcal{U}$ minimizes the negative log-likelihood $-\log P(G^* \mid \tau^{(k)}; \phi)$.

The agent aims to maximize cumulative belief-induced reward depending on the observer's latent belief $b_t^{(k)}$ over repeated episodes,

$$V_\pi = \mathbb{E}\left[\sum_{k=1}^{K}\sum_{t=0}^{T_k} \gamma^t \, \mathcal{L}(s_t^{(k)}, a_t^{(k)}, s_{t+1}^{(k)}, b_t^{(k)})\right].$$

Since the observer's belief is not directly observable during execution, the agent must adapt its policy across episodes using only post-episode predictive feedback, resulting in a non-stationary deceptive planning problem.

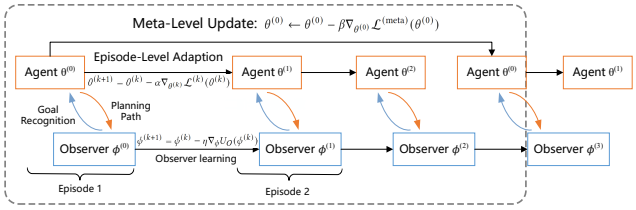


Figure 1: The agent performs episode-level policy adaptation over repeated interactions, while meta-level updates refine the policy initialization using cross-episode feedback.

2.2 Deceptive Meta Planning

To solve RDPP, we propose **Deceptive Meta Planning (DeMP)**, a two-level optimization framework illustrated in Figure 1. DeMP explicitly separates short-term adaptation within episodes from long-term learning across episodes.

At the *episode level*, the agent updates its policy parameters to counter the observer's current prediction model. Starting from an initialization $\theta^{(0)}$, the agent performs gradient-based adaptation after each episode:

$$\theta^{(k+1)} = \theta^{(k)} - \alpha \nabla_{\theta^{(k)}} \mathcal{L}^{(k)}(\theta^{(k)}),$$

where $\mathcal{L}^{(k)}$ is a deception-aware loss constructed from the observer's predictive distribution $O(\zeta^{(k)}; \phi^{(k)})$.

After M episode-level updates, DeMP performs a *meta-level* update that optimizes the policy initialization for future interactions. The meta-objective is defined on the final adapted parameters:

$$\mathcal{L}^{(\text{meta})}(\theta^{(0)}) = \mathcal{L}^{(M)}\big(\theta^{(M)}(\theta^{(0)})\big),$$

and the initialization is updated via

$$\theta^{(0)} \leftarrow \theta^{(0)} - \beta \nabla_{\theta^{(0)}} \mathcal{L}^{(\text{meta})}(\theta^{(0)}).$$

By optimizing the initialization to anticipate how the observer's predictions evolve across episodes, DeMP mitigates the accumulation of adaptation lag inherent in purely reactive updates. This two-level optimization enables sustained deception against learnable observers in repeated interactions.

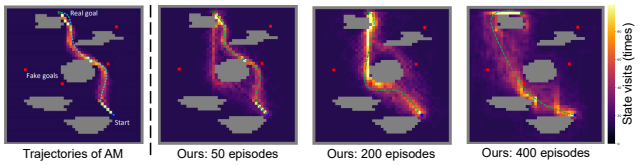


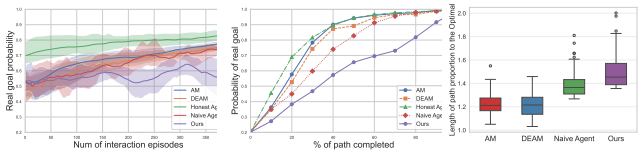
Figure 2: The AM [2] method follows a fixed path, while DeMP continuously adapts its trajectory distribution across episodes, preventing the observer from exploiting a single motion pattern.

3 EXPERIMENTS

We evaluate Deceptive Meta Planning (DeMP) in repeated deceptive path planning (RDPP), focusing on its ability to sustain deception against a learnable observer over long-term interactions. Experiments are conducted on grid-world navigation tasks with multiple candidate goals and an adaptive observer trained for goal recognition.

As shown in Fig. 3, DeMP consistently maintains the lowest predicted probability of the true goal over repeated interactions, while all baselines degrade as the observer adapts. Methods without explicit cross-episode adaptation become exploitable due to repeated exposure to similar trajectory prefixes, enabling early goal inference. In contrast, DeMP sustains deception by continually adapting its strategy to the observer's learning dynamics, at the cost of slightly increased path length.

Figure 2 qualitatively illustrates this effect: baseline methods converge to static or weakly varying trajectories, whereas DeMP exhibits persistent trajectory diversification, preventing stable early-goal associations and enabling long-term deceptive performance.



(a) (b) (c)

Figure 3: Performance of Deceptiveness and cost. (a) Deceptiveness across 400 episodes. (b) Deceptiveness in last episode. (c) Path costs.

4 CONCLUSION

We introduced repeated deceptive path planning (RDPP), a setting that captures deception under repeated interactions with a learnable observer. To address this challenge, we proposed Deceptive Meta Planning (DeMP), a two-level optimization framework combining episode-level adaptation with meta-level updates. Experiments demonstrate that DeMP sustains deception over long-term interactions, outperforming existing methods that degrade under observer adaptation. Future work will explore extending DeMP to more complex environments, richer observer models, and multi-agent deceptive interactions.

ACKNOWLEDGMENTS

This work was supported by the National Science and Technology Major Project under Grant No. 2022ZD0116403, and in part by the Beijing Natural Science Foundation under Grant No. 4264131.

REFERENCES

[1] Jonathan Bell. 2003. Toward a Theory of Deception. *International Journal of Intelligence and CounterIntelligence* 16 (2003), 244 – 279.

[2] Zhengshang Liu, Yue Yang, Tim Miller, and Peta Masters. 2021. Deceptive Reinforcement Learning for Privacy-Preserving Planning. In *Adaptive Agents and Multi-Agent Systems*.

[3] Junren Luo, Wanpeng Zhang, Wei Gao, Zhiyong Liao, Xiang Ji, and Xueqiang Gu. 2019. Opponent-aware planning with admissible privacy preserving for UGV security patrol under contested environment. *Electronics* 9, 1 (2019), 5.

[4] Peta Masters and Sebastian Sardina. 2017. Deceptive Path-Planning.. In *IJCAI*. 4368–4375.

[5] Melkior Ornik and Ufuk Topcu. 2018. Deception in optimal control. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 821–828.

[6] Qingfeng Xu, Yingnan Shi, Junchao Wang, Tim Miller, Hangding Xu, Tianmu Wang, Hongbin Lin, Xin-Jun Liu, and Zhenguo Nie. 2022. Path Planning and Information Protection of Mobile Robots Based on Deceptive Reinforcement Learning. In *International Conference on Mechanism and Machine Science*. Springer, 2271–2284.