Neural Processing Letters (2024) 56:161
https://doi.org/10.1007/s11063-024-11625-w

A Pilot Study of Observation Poisoning on Selective Reincarnation in Multi-Agent Reinforcement Learning

Harsha Putla[1] · Chanakya Patibandla[1] · Krishna Pratap Singh[1] · P Nagabhushan[2]

Accepted: 11 April 2024 / Published online: 2 May 2024

Abstract

This research explores the vulnerability of selective reincarnation, a concept in Multi-Agent Reinforcement Learning (MARL), in response to observation poisoning attacks. Observation poisoning is an adversarial strategy that subtly manipulates an agent's observation space, potentially leading to a misdirection in its learning process. The primary aim of this paper is to systematically evaluate the robustness of selective reincarnation in MARL systems against the subtle yet potentially debilitating effects of observation poisoning attacks. Through assessing how manipulated observation data influences MARL agents, we seek to highlight potential vulnerabilities and inform the development of more resilient MARL systems. Our experimental testbed was the widely used HalfCheetah environment, utilizing the Independent Deep Deterministic Policy Gradient algorithm within a cooperative MARL setting. We introduced a series of triggers, namely Gaussian noise addition, observation reversal, random shuffling, and scaling, into the teacher dataset of the MARL system provided to the reincarnating agents of HalfCheetah. Here, the "teacher dataset" refers to the stored experiences from previous training sessions used to accelerate the learning of reincarnating agents in MARL. This approach enabled the observation of these triggers' significant impact on reincarnation decisions. Specifically, the reversal technique showed the most pronounced negative effect for maximum returns, with an average decrease of 38.08% in Kendall's tau values across all the agent combinations. With random shuffling, Kendall's tau values decreased by 17.66%. On the other hand, noise addition and scaling aligned with the original ranking by only 21.42% and 32.66%, respectively. The results, quantified by Kendall's tau metric, indicate the fragility of the selective reincarnation process under adversarial observation poisoning. Our findings also reveal that vulnerability to observation poisoning varies significantly among different agent combinations, with some exhibiting markedly higher susceptibility than oth-

✉ Harsha Putla
pis2017002@iiita.ac.in

Chanakya Patibandla
pis2017003@iiita.ac.in

Krishna Pratap Singh
kpsingh@iiita.ac.in

P Nagabhushan
pnagabhushan@iiita.ac.in

[1] Department of Information Technology, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India

[2] Vignan's Foundation for Science, Technology & Research, Guntur, Andhra Pradesh, India

ers. This investigation elucidates our understanding of selective reincarnation's robustness against observation poisoning attacks, which is crucial for developing more secure MARL systems and also for making informed decisions about agent reincarnation.

Keywords Observation poisoning · Selective reincarnation · Multi-agent reinforcement learning · Adversarial attacks · Kendall's tau metric

1 Introduction

Modeling and solving complex problems involving multiple interacting agents is made possible through the powerful paradigm of MARL [1–3]. It is widely applicable in various fields such as autonomous vehicles [4], traffic light control [5], games [6–8] and intelligent energy grids [9], highlighting its immense potential.

Selective reincarnation is a recent advancement in MARL [10], which enables past computations to be reused [11]. This approach reduces overall computational cost and allows the system to adapt to changing environments by leveraging prior knowledge such as model weights, offline datasets, and other computational assets. Reincarnation is selectively performed based on criteria such as maximum and average returns, ensuring the most effective agents are reincarnated. Recent studies utilizing reincarnation have demonstrated improvements in policy selection, integrated learning paradigms, and the utilization of prior computations, effectively enhancing the efficiency and effectiveness of policy gradient algorithms in practical settings [12–14].

Selective reincarnation has brought improvements, but it also introduces new vulnerabilities in the face of adversarial attacks. Observation poisoning is one such attack that can degrade the performance of well-trained neural network policies by perturbing the observation space [15]. This issue extends to crowdsensing systems where false data can be injected to interfere with analysis results [16]. With the increasing prevalence of these attacks, it is urgent to develop robust models that can withstand such threats [17]. The safety of MARL systems is critical for their successful deployment in real-world scenarios like autonomous driving and robotics [18]. Ignoring safety in RL can lead to catastrophic outcomes [19]. Recent studies [20–23] highlight the potential of safe RL to enhance the reliability of AI systems. This is particularly relevant in the context of MARL, where action [24], policy [25], and reward [26] poisoning attacks pose significant threats to system performance. Therefore, it is necessary [27–29] to test and evaluate the susceptibility of selective reincarnation to adversarial attacks, which aids in the essential step of development of robust defenses and resilient algorithms [30–36].

Our research explores the impact of observation poisoning on the decisions regarding which agents to selectively reincarnate within a MARL framework. We conducted extensive experiments using the HalfCheetah environment [37] and the IDDPG [38] algorithm, introducing triggers such as Gaussian noise addition, observation reversal, random shuffling, and scaling into the teacher dataset of the MARL system. To focus on selective reincarnation and its susceptibility to poisoning attacks, we exploit the 'Good-Medium' dataset [10]. This dataset comprises about the final 40% of teachers' experiences, stored in the Off-the-Grid MARL framework [10, 39]. We noted significant influences on reincarnation decisions and quantified this influence using Kendall's tau metric [40]. Our study provides valuable insights into the robustness of selective reincarnation in MARL against poisoning attacks, paving the way for developing more secure and reliable systems for real-world application.

Along with the susceptibility of the selective reincarnation, our findings comprehensively assess agent combinations under the various attack scenarios, offering insights into their vulnerability. For instance, the combination back ankle (BA), front ankle (FA), and front knee (FK) is most vulnerable, with an overall vulnerability score of 46%, calculated as the average vulnerability across multiple attack scenarios. In contrast, BA was the only agent to demonstrate resilience, showing a 10% negative vulnerability score, which suggests that it performs better under attack conditions than in the baseline scenario.

In the next section, we discuss prior work that aligns with our research, establishing the significance of studying the vulnerability and resilience of MARL's selective reincarnation against adversarial attacks.

2 Related Work

Our research intersects three primary areas of prior work: "Selective Reincarnation in MARL" [10], "Adversarial Attacks in Deep RL, specifically Observation Poisoning", and "Robustness Evaluation in MARL".

2.1 Selective Reincarnation in MARL

MARL has garnered attention due to its ability to model complex interactions between multiple agents. Recent literature has explored the concept of reincarnation in MARL which involves reusing prior computations based on past performances. This approach has shown significant benefits, such as improved computational efficiency and adaptability, as discussed in a work by [11]. Transfer learning has been another area of interest in MARL. Study [41] introduce an ontology-based approach to facilitate knowledge transfer across agents, which aligns with the broader theme of reusing knowledge. Moreover, [42] proposes methods to transfer knowledge from trained agents to newer ones, resulting in improved training efficiency and performance. Selective reincarnation, a form of reincarnation in which we can select which agents to reincarnate, in MARL has been found to improve learning efficiency by reusing previous computations across selected agents [10]. In a cooperative, heterogeneous HalfCheetah MARL setup, it shows faster convergence and better returns than starting anew or complete reincarnation. However, careful selection of agents to reincarnate is crucial as incorrect selection can yield inferior results. Our research focuses on the unexplored aspect of the robustness of selective reincarnation in MARL against susceptibility to poisoning attacks, specifically observation poisoning.

2.2 Adversarial Attacks in Deep RL: Observation Poisoning

The vulnerability of MARL systems, especially in the face of adversarial attacks, has been a pressing concern. The paper [43] discusses the challenges posed by dynamic environments and the need for continuous coordination among agents. This work underscores the importance of our research, which focuses on the vulnerabilities introduced by observation poisoning.

Observation poisoning, an adversarial tactic, can majorly derail an agent's learning by manipulating its observation space, thereby threatening RL system robustness. Research shows that even slight disruptions can significantly affect Deep RL agents, inducing them to adopt sub-optimal policies [44]. A two-stage optimization-based attack can efficiently

introduce adversarial noise into RL, heavily impacting performance [45]. Backdoors attack using triggers in deep RL agents hamper their performance [46]. Notably, a small amount of poisoned training data can lead to successful backdoor attacks, highlighting system vulnerabilities [47].

2.3 Robustness Evaluation in MARL

Ensuring that MARL systems are robust, especially as they are deployed in diverse and challenging environments, is crucial. Due to the varied landscape of adversarial attacks on MARL, particularly on input observations, it is essential to understand these threats and develop appropriate evaluation metrics and defense strategies.

Like other domains in machine learning [48–50], one standard attack on MARL systems is the Gaussian noise addition (GNA), which introduces subtle yet effective adversarial strategies by adding noise to agents' observations. Attackers can mislead agents and adversely affect their learning trajectories through this manipulation. The significance of defending against Gaussian noise addition is emphasized in research such as [51], showcasing the profound impact of such a seemingly simple attack on MARL systems.

Shuffling and reversal attacks are also potent adversarial tactics that can drastically alter an agent's perception of the environment without changing the actual state of the environment. These manipulations lead to sub-optimal learning outcomes. Multiple works, such as [16, 45, 52–56], highlight the importance of understanding and mitigating the risks associated with these shuffling-based attacks in MARL.

Although scaling attacks have been extensively studied in broader machine learning contexts [45, 48, 57–59], their impact on MARL systems remains less explored. These attacks manipulate the magnitude of agents' observations, leading to skewed perceptions and decisions. Our work assesses the robustness of selective reincarnation in the face of diverse poisoning attacks in MARL, including scaling attacks. We focus on understanding how observation poisoning affects agent performance within a selectively reincarnated HalfCheetah framework and its implications for selecting agents for reincarnation. Our exploration sheds light on potential vulnerabilities, offering valuable insights into the challenges and susceptibility of selective reincarnation in MARL to poisoning attacks.

In summary, it is crucial to defend against various poisoning attacks on MARL systems, including Gaussian noise addition, shuffling and reversal attacks, and scaling attacks. Our in-depth analysis provides valuable findings into the robustness of selective reincarnation against these diverse poisoning attacks in MARL. These insights not only highlight potential vulnerabilities but also serve as a foundational resource for researchers exploring defense mechanisms.

3 Methodological Foundations and Evaluation Metrics

In order to provide a comprehensive understanding of the impact of observation poisoning on MARL systems, this section introduces the key methodologies and statistical framework we employ. From the foundational principles of agent behavior to the specialized metric like Kendall's Tau for assessing ranking changes and the "Overall Vulnerability" for assessing the vulnerability of agent combinations, we lay down the theoretical groundwork for our experimental setup and analysis.

3.1 Independent Deep Deterministic Policy Gradient (IDDPG)

In the field of MARL, especially in the context of fully cooperative MARL with shared rewards, agents frequently operate within a framework known as the Decentralized Partially Observable Markov Decision Process (Dec-POMDP) [60]. This framework is described by a tuple $M = (N, S, \{Ai\}, \{Oi\}, P, E, \rho_0, r, \gamma)$. Here, N signifies the set of n agents, and $s \in S$ represents the true state of the environment. Each agent i is privy to partial observations from the environment, determined by an emission function $E(o_t|s_t, i)$. At every timestep t, agents receive local observations o_i^t and decide on actions a_i^t, culminating in a joint action a_t. Agents maintain an observation history $o_i^{0:t}$, which influences their policy $\mu_i(a_i^t|o_i^{0:t})$. The environment then transitions based on $P(s_{t+1}|s_t, a_t)$ and allocates a shared reward via $r(s, a)$.

Given the Dec-POMDP framework, the inherent challenge lies in the decentralized nature of the environment, where each agent only has access to partial observations. Thus, the primary objective is to find a joint policy $\pi = \{\pi_1, \pi_2, ..., \pi_n\}$ for all agents $i \in N$, such that the expected return $G_i = \sum_{t=0}^{T} \gamma^t r_i^t$ for each agent i, following its policy π_i, is maximized, taking into account the policies π_j of all other agents $j \neq i$.

To address this challenge, the IDDPG [38, 61] methodology is introduced. For each agent i, a Q-function $Q_i^\theta(o_i^{0:t}, a_i^t)$ is established, conditioned on its observation history and action. Concurrently, a policy network $\mu_i^\phi(o_i^t)$ is set up for each agent, mapping observations to actions. The Q-function is trained to minimize the temporal difference (TD) loss:

$$L_Q(D_i, \theta_i) = E_{o_i^t, a_i^t, r_t, o_i^{t+1} \sim D}\left[(Q_i^{\theta_i}(o_i, a_i) - r_t - \gamma \hat{Q}^{\theta_i}(o_i^{t+1}, \hat{\mu}^\phi(o_t + 1)))^2\right]$$

where, $\hat{Q}^\theta$ and $\hat{\mu}^\phi$ are target versions of the Q-network and policy network, respectively. The variable D_i represents the experience replay buffer for each agent i, which stores past experiences in the form of tuples $(o_i^t, a_i^t, r_t, o_i^{t+1})$. These stored experiences are sampled to train the Q-function and the policy network, thereby minimizing the temporal difference (TD) loss and policy loss, respectively. Simultaneously, the policy network is optimized to predict the action that maximizes the Q-function, resulting in the minimization of the policy loss:

$$L_\mu(D_i, \phi_i) = E_{o_i^t \sim D}\left[-Q_i^{\theta_i}(o_i^t, \mu_i^{\phi_i}(o_i^t))\right]$$

IDDPG is fundamental to our analysis, setting the stage for evaluating observation poisoning within MARL systems, with practical implementation details to follow in 4.2.

3.2 Observation Poisoning Techniques

In our HalfCheetah MARL framework [37], agents acquire observations from the environment, represented as a $d \times d$ matrix, s, with $d = 10$, defining the dimensionality of the observations. The following perturbation techniques have been applied to these observations in the teacher dataset. Despite some techniques causing minor perturbations, their impact on the learning process was found to be significant:

Gaussian Noise Addition: Gaussian noise is added to the observations to induce randomness and foster diverse learning experiences. Each element of the observation matrix s is independently perturbed:

$$s'_{ij} = s_{ij} + \epsilon_{ij}; \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \ \sigma = 0.01$$

The perturbations, despite being minimal, have significant effects on the learning process.
Observation Reversal: The order of rows in the matrix s is inverted, disrupting the potential learning from crucial environmental state information:

$$s'_{ij} = s_{d-i+1,j}, \quad \forall i, j \in \{1, 2, \ldots, d\}$$

Random Shuffling: Rows of s are randomly shuffled, causing disarray similar to the observation reversal technique:

$$s'_{ij} = s_{\pi(i),j}, \quad \pi \in \{\pi : \{1, 2, \ldots, d\} \to \{1, 2, \ldots, d\} \mid \pi \text{ is bijective}\}$$

Scaling: Elements of s are scaled by a constant factor α to subtly modify the observation values, potentially affecting agent learning:

$$s'_{ij} = \alpha \times s_{ij}; \quad \alpha = 1.1$$

Upon implementing these poisoning techniques on the teacher dataset used for training the reincarnating agents, the reincarnating agents will receive these perturbed observations, which will affect the learning process that is implemented using IDDPG. This affects their learning process, which is implemented using IDDPG.

3.3 Rank Correlation Analysis

We evaluated the impact of observation poisoning on the performance of reincarnating agents using Kendall's tau correlation coefficient [40], which assesses the association between pre- and post-poisoning performance rankings. The performance measure can be either average returns or maximum returns.

We begin by defining the sign function $\text{sgn}(z)$ as follows:

$$\text{sgn}(z) = \begin{cases} +1 & \text{if } z > 0 \\ -1 & \text{if } z < 0 \\ 0 & \text{if } z = 0 \end{cases} \tag{1}$$

Where z represents the difference in ranks between two data points. Using Eq. 1, we define Kendall's τ as shown in Eq. 2:

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} \text{sgn}(x_i - x_j) \cdot \text{sgn}(y_i - y_j) \tag{2}$$

In Eq. 2, n is the total number of ranked data points, while x_i and y_i represent the ranks of individual data points in the pre-and post-poisoning rankings, respectively.

A τ value close to 1 signifies minimal ranking changes due to poisoning, while a value near -1 implies significant ranking disruption. This measure offers a quantitative assessment of the influence of poisoning techniques, validating the statistical significance of performance changes caused by poisoning.

3.4 Quantifying Vulnerability of Agent Combinations

In addressing the challenge of adversarial threats in selective reincarnation in MARL systems, this section introduces a mathematical quantification method to assess the 'Overall Vulnerability' of agent combinations to observation poisoning attacks, with a focus on evaluating the vulnerability of reincarnating agent combinations.

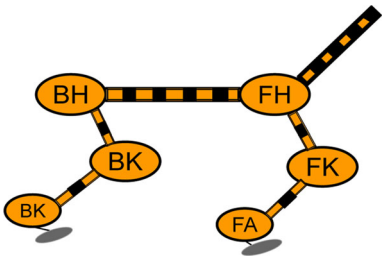

Fig. 1 The illustration of HalfCheetah as a collection of six different agents

To quantify the concept of "Overall Vulnerability", we employed the following mathematical formula:

$$V_c = \frac{100}{8} \sum_{i \in A} \left(\frac{B_{\max,c} - P_{\max,i,c}}{B_{\max,c}} + \frac{B_{\text{avg},c} - P_{\text{avg},i,c}}{B_{\text{avg},c}} \right)$$

where V_c represents the Overall Vulnerability for a given agent combination c. A denotes the set of considered attacks, specifically: {noise, reversal, scaling, shuffling}. The terms $B_{\max,c}$ and $B_{\text{avg},c}$ denote the maximum and average returns for the base case with agent combination c, respectively. Similarly, $P_{\max,i,c}$ and $P_{\text{avg},i,c}$ represent the maximum and average returns for the i-th poisoned case with agent combination c, respectively. For a detailed explanation of the 'maximum return' and 'average return' metrics, refer to Sect. 4.3.

4 Experimental Setup

Building on the approaches and statistical framework discussed earlier, this section provides an exhaustive exposition of our experimental design, including the specifics of IDDPG training. Our focus is to scrutinize the influence of observation poisoning on MARL systems, particularly in the HalfCheetah environment. Here, we detail the HalfCheetah setup, elaborate on the performance metrics, and outline the step-by-step procedure followed, thereby setting the stage for the forthcoming results and analysis.

4.1 The HalfCheetah Setup

The HalfCheetah environment [62, 63], part of the Mujoco simulation software suite [64], simulates a two-legged robot designed for rapid forward motion. While Mujoco's default configuration perceives the HalfCheetah as a singular entity, our adaptation [37], based on Multi-Agent Mujoco (MaMujoco), views it as a system comprised of six cooperative agents: the back ankle (BA), back knee (BK), back hip (BH), front ankle (FA), front knee (FK), and front hip (FH). For a more intuitive understanding, Fig. 1 visually depicts these agents and their interconnections. These agents work collectively toward the objective of forward progression. We selected this setup due to its relevance to practical robotics, complex agent dynamics, continuous action domains, and compatibility with the MARL framework. This configuration is also well-suited for the selective reincarnation process, which utilizes a dataset of experiences derived from Tabula Rasa.

4.2 IDDPG Training Details

Having outlined the theoretical foundations of IDDPG in 3.1, we now turn to details its practical implementation [10] in our multi-agent setup. Each agent employs an individual policy network for action decision-making given an environment observation. Concurrently, each agent is equipped with a critic network tasked with estimating the value of taking a specific action in a given state.

To ensure stability during the training process, target networks are utilized for both the policy and the critic. Separate optimizers are employed for the policy and critic networks to facilitate an effective and efficient training process. During the exploration phase, each agent is programmed to interact with the environment for a default of 10,000 timesteps. Agents designated for selective reincarnation undergo additional training on a specialized teacher dataset for 200,000 timesteps.

The hyperparameters for the experiments include a default batch size of 32 for training, a discount factor of 0.99, a lambda value of 0.6 for temporal-difference bootstrapping, and a noise standard deviation of 0.1 for exploration.

4.3 Performance Metrics

In our evaluation, two primary metrics were employed to assess the performance of the MARL system in the HalfCheetah environment: the 'maximum return' and the 'average return'. These metrics offer a systematic evaluation of both the system's peak performance capability and its overall consistency.

Maximum Return:
The 'maximum return', denoted as $R_{\max}$, encapsulates the highest return secured during the training process, averaged across all seeds. If $R(t, s)$ represents the return at timestep t for seed s, the maximum return is formulated as:

$$R_{\max} = \max_{t \in [1,T]} \left(\frac{1}{S} \sum_{s \in \mathcal{S}} R(t, s) \right)$$

where T is the total number of timesteps and $\mathcal{S}$ is the set of seeds.

Average Return:
The 'average return', denoted as R_{avg}, offers a measure of the consistent performance of the system, with the return averaged over all timesteps and seeds. It is defined as:

$$R_{\text{avg}} = \frac{1}{TS} \sum_{t \in [1,T]} \sum_{s \in \mathcal{S}} R(t, s)$$

We utilized these metrics to evaluate the HalfCheetah MARL system within a Decentralised Partially Observable Markov Decision Process (Dec-POMDP) [60] framework. The aim was to determine a joint policy that maximizes each agent's return in relation to other agents' policies. The effects of observation poisoning techniques on these metrics were then examined.

4.4 Experimental Procedure

Inspired by [10], we developed our own variant procedure 4.4.1 from scratch and used the 'Good-Medium teacher dataset' of stored experiences based on their guidelines. We

utilize the multi-agent environment Mujoco with a configuration of "HalfCheetah". This environment involves multiple agents interacting with each other in a simulated physical space. The primary algorithm used for training the agents is IDDPG, which is a variant of the Deep Deterministic Policy Gradient algorithm tailored for multi-agent settings.

4.4.1 Procedure for Training, Dataset Poisoning, Agent Reincarnation, and Performance Evaluation

In this subsection, we present the procedure that encompasses the training, dataset poisoning, reincarnation, and performance evaluation of the agents. The procedure is methodically laid out in seven distinct steps. Subsequent to the enumeration, a more in-depth description of each step provides clarity and insight into the methodology and processes involved. The steps are:

1. Initial Training of Agents and Creation of Teacher Datasets.
2. Observation Poisoning.
3. Enumeration of Agent Combinations for Reincarnation.
4. Retraining of the System for Each Combination.
5. Performance Evaluation.
6. Grouping and Sorting of Reincarnating Agents Based on Metric Value.
7. Kendall's Tau Rank Calculation.

Description of Each Step in Procedure 4.4.1:

1. **Initial Training of Agents and Creation of Teacher Datasets:** The six agents are initially trained on the HalfCheetah environment using Independent Deep Deterministic Policy Gradient (IDDPG) over 1 million training steps. Training experiences are saved as teacher dataset [39].
2. **Observation Poisoning:** The teacher dataset is manipulated through triggers like Gaussian noise addition, observation reversal, shuffling, and scaling. This "observation poisoned" dataset is later given to the reincarnating agents.
3. **Enumeration of Agent Combinations for Reincarnation:** All $2^6 = 64$ subsets of agent combinations for reincarnation are enumerated. Each subset accesses its corresponding offline teacher dataset during retraining on the HalfCheetah environment.
4. **Retraining of the System for Each Combination:** Each agent combination undergoes training for 200k timesteps, post-teacher data removal, and an additional 50k timesteps on student data alone, with this process repeated over five seeds (0–4).
5. **Performance Evaluation:** Using 'maximum return' and 'average return' metrics as explained in the performance metrics subsection, we assessed the system's performance and speed to convergence, respectively.
6. **Grouping and Sorting of Reincarnating Agents Based on Metric Value:** Reincarnating agents are grouped by number involved, then within groups, sorted descendingly by either maximum or average return.
7. **Kendall's Tau Rank Calculation:** Calculate Kendall's tau to assess the degree of correspondence between the orders of agent combinations based on the performance metrics. It serves to measure the impact of poisoning on the ordering of reincarnating agents.

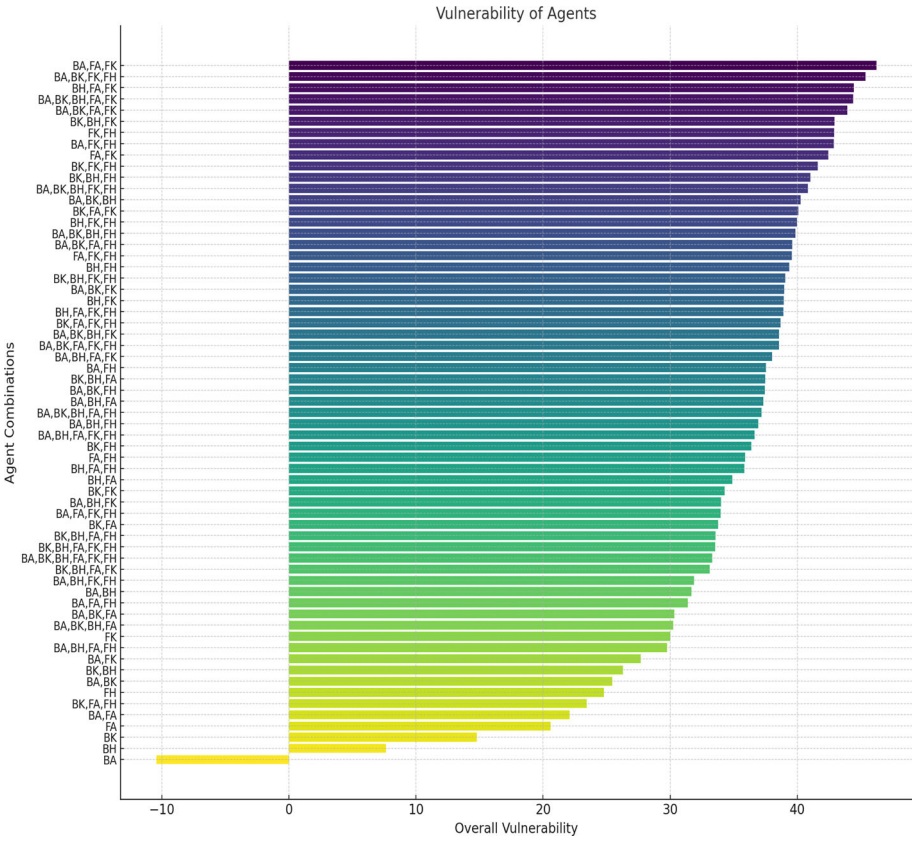

Fig. 2 Bar chart depicting the overall vulnerability percentages of different reincarnating agent combinations under the four observation poisoning attacks

5 Results, Analysis and Discussion

We carried out exhaustive experiments to study the impact of observation poisoning on the HalfCheetah MARL system, specifically focusing on reincarnation decisions. Kendall's tau correlation coefficient, a non-parametric statistic, was used to quantify the order correlation between the different poisoning techniques and reincarnation decisions based on maximum and average returns.

Tables 1 and 2 summarize the experiment outcomes, each showcasing the performance metrics for different reincarnating agent combinations under diverse observation poisoning attacks. Each configuration represents a subset of the six agents. Table 1 focuses on maximum returns, while Table 2 displays average returns. The reported values are average values over five seeds and the standard deviation to show result variability. To avoid excessive listing, we present only the best and worst-performing configurations for each number of reincarnating agents, illustrating the performance range for each attack scenario.

The bar chart in Fig. 2 visualizes the overall vulnerability of different reincarnating agent combinations under the four poisoning attacks. Notably, the agent combination 'BA' exhibits the lowest vulnerability, while 'BA, FA, FK' shows the highest. This analysis lends empir-

ical support to the notion that some agent combinations are inherently more robust against observation poisoning attacks than others.

The findings of this study from Table 1 spotlight the considerable variation in the performance of different agent combinations when subjected to diverse types of poisoning attacks. For instance, the BA, BK, BH, FK, FH configuration, despite its superior performance in the base case, i.e., without poisoning (5441.5 ± 289.4), fails to keep its place in a noisy environment, being outperformed by the BA, BK, FA, FK, FH configuration (4923.9 ± 156.8). In the reversal attack, even the BA FA configuration, with a lower maximum return (4687.7 ± 588.0), outperforms it. Interestingly, the fully reincarnated configuration stands robust against the scaling attack with a return of 4352.9 ± 164.8. The result suggests that the agents' adaptability depends heavily on the nature of the attack, emphasizing the need for versatile combinations.

An analysis of the average returns, as detailed in Table 2, suggested that some agent combinations maintained more consistent performances while others exhibited high variability. It is apparent from the associated standard deviations. Table 2 reveals that the BA, FA pairing yields the highest average returns under the 'Reversal Attack,' albeit with the highest variance at a standard deviation of 566.4. Meanwhile, the FK, FH pairing under the 'Base Case' offers high average returns with less volatility, indicated by a standard deviation 297.1. It suggests the 'Base Case' FK, FH pairing's superior stability, despite BA, FA's high returns under the 'Reversal Attack', highlighting the trade-off between performance and consistency across configurations.

The fully reincarnated configuration, representing the scenario of all agents reincarnated simultaneously, provided an intriguing benchmark. It performed well under the base case and noise addition attacks but struggled against reversal and random shuffling attacks. It suggests that these latter types of attacks particularly disrupt inter-agent cooperation and coordination.

This data provides comprehensive insights into the resilience of each agent to observation poisoning and their ability to recover through reincarnation; by examining both the peak and average performance of different agent configurations under attack scenarios, we highlight the potential vulnerability of reincarnation decisions in HalfCheetah MARL systems to observation poisoning.

The Overall Vulnerability chart serves as a supplementary guide to the tables, providing a quick, at-a-glance view of which reincarnating agent combinations are most and least vulnerable to poisoning attacks. This added layer of analysis aids in the strategic decision-making process for configuring agents in different attack scenarios.

Table 3 offers a comparative analysis of various poisoning techniques using Kendall's Tau correlation coefficient, measuring their impact on performance rankings across diverse agent reincarnations. An initial ranking positions 'FH' first, followed by 'FK', 'FA', 'BK', 'BH', and 'BA'. Post the shuffling attack, the order significantly alters, with 'BH' leading, then 'BA', 'BK', 'FA', 'FH', and 'FK', a shift evidenced by the Kendall's Tau value of -0.733, signaling substantial ranking reshuffle.

These trends are also evident in Table 3, where techniques like noise addition and reversal significantly impact rankings, especially with a single reincarnating agent. However, as the number of reincarnating agents increases, the impact lessens. Nevertheless, the scaling technique consistently influences rankings regardless of the number of reincarnating agents. This drastic ranking effect underscores poisoning's disruptive potential in reincarnating agent selection, reinforcing the need for robust strategies to counter such disturbances.

Table 3 crucially informs reincarnating agent selection decisions by illuminating poisoning techniques' impacts on rankings, helping identify susceptible and sturdy agents. The table thus becomes a vital strategic resource, providing extensive data on the varied poisoning techniques' efficacy and impact across diverse agent combinations.

Table 1 Maximum return values for best and worst runs of reincarnated agents with & without observation poisoning

Base case		Noise addition attack		Reversal attack	
Configuration	Maximum returns	Configuration	Maximum returns	Configuration	Maximum returns
Tabula Rasa	2484.7 ± 357.4	N/A	N/A	N/A	N/A
FH	4092.3 ± 589.7	FH	3418.9 ± 373.7	BA	4347.9 ± 1142.4
BA	2826.6 ± 514.5	BA	2484.1 ± 355.7	FK	3409.3 ± 625.4
FA, FK	5111.2 ± 190.9	FK, FH	3624.3 ± 282.5	BA, FA	**4687.7 ± 588.0**
BA, FA	4285.1 ± 729.7	FA, FK	2862.1 ± 221.4	FK, FH	3299.5 ± 1082.0
BH, FK, FH	5052.5 ± 538.9	BA, BK, FH	4050.8 ± 271.1	BA, BK, FA	3545.4 ± 876.2
BK, FA, FH	3859.0 ± 810.9	BK, BH, FK	2601.6 ± 396.6	BA, FA, FK	1727.8 ± 406.2
BA, BK, FK, FH	5077.4 ± 284.4	BA, BH, FK, FH	4390.1 ± 150.3	BA, BH, FA, FH	2764.8 ± 431.1
BA, BK, BH, FA	4212.4 ± 340.7	BA, BK, FA, FK	3112.2 ± 517.5	BK, BH, FK, FH	1489.2 ± 296.1
BA, BK, BH, FK, FH	**5441.5 ± 289.4**	BA, BK, FA, FK, FH	**4923.9 ± 156.8**	BA, BH, FA, FK, FH	2146.9 ± 260.4
BK, BH, FA, FK, FH	4962.0 ± 643.2	BA, BK, BH, FA, FK	4141.4 ± 505.3	BA, BK, BH, FK, FH	1474.0 ± 391.4
Fully Reincarnated	5148.7 ± 275.9	Fully Reincarnated	4543.8 ± 351.2	Fully Reincarnated	2127.4 ± 164.3

Random shuffling attack		Scaling attack	
Configuration	Maximum returns	Configuration	Maximum returns
N/A	N/A	N/A	N/A
BH	3341.4 ± 426.5	FH	2418.4 ± 143.1
FK	2717.3 ± 275.7	BA	1898.9 ± 584.2
FA, FK	**3590.5 ± 232.6**	BK, BH	3188.2 ± 409.1
FK, FH	2835.6 ± 278.6	BK, FK	2274.4 ± 409.2
BA, BK, FH	3350.8 ± 115.6	FA, FK, FH	4072.3 ± 279.8
FA, FK, FH	2363.5 ± 502.1	BK, BH, FK	2899.3 ± 75.2

Table 1 continued

Random shuffling attack		Scaling attack	
Configuration	Maximum returns	Configuration	Maximum returns
BA, FA, FK, FH	3292.8 ± 297.4	BA, FA, FK, FH	4234.6 ± 194.7
BA, BK, FK, FH	2546.2 ± 336.2	BA, BK, BH, FK	3769.6 ± 255.1
BA, BH, FA, FK, FH	3479.1 ± 170.5	BA, BK, BH, FK, FH	4241.0 ± 183.0
BA, BK, BH, FA, FK	3027.0 ± 276.3	BA, BK, BH, FA, FK	3958.2 ± 297.6
Fully Reincarnated	3381.7 ± 264.2	Fully Reincarnated	$\mathbf{4352.9 \pm 164.8}$

Bold values indicate the highest performing results

Table 2 Average return values for best and worst runs of reincarnated agents with & without observation poisoning

Base case		Noise addition attack		Reversal attack	
Configuration	Average returns	Configuration	Average returns	Configuration	Average returns
Tabula Rasa	1098.8 ± 262.8	N/A	N/A	N/A	N/A
FH	2586.0 ± 613.2	FH	2001.2 ± 383.6	BH	2956.7 ± 550.0
BA	1526.9 ± 516.2	BA	1273.0 ± 311.5	FK	2127.3 ± 359.0
FK, FH	3472.3 ± 297.1	BK, FH	2435.1 ± 441.6	BA, FA	**3167.4 ± 566.4**
BA, FA	2704.2 ± 760.7	FA, FK	1540.2 ± 132.4	BK, FH	2014.0 ± 408.2
BA, BK, FH	3719.0 ± 135.3	BA, BK, FH	2775.8 ± 229.4	BA, BK, FK	2223.1 ± 882.5
BK, FA, FH	2240.2 ± 654.8	FA, FK, FH	1544.2 ± 557.6	BA, FA, FK	758.4 ± 482.7
BK, BH, FK, FH	3670.7 ± 239.5	BK, BH, FK, FH	2943.0 ± 261.3	BA, BH, FA, FH	1208.1 ± 269.6
BA, BK, BH, FA	2820.2 ± 407.2	BK, BH, FA, FK	1676.0 ± 185.6	BK, FA, FK, FH	577.5 ± 193.5
BA, BK, BH, FK, FH	**4144.0 ± 426.1**	BA, BK, BH, FK, FH	**3333.6 ± 149.6**	BA, BH, FA, FK, FH	1148.8 ± 289.9
BK, BH, FA, FK, FH	3553.4 ± 498.9	BA, BK, BH, FA, FK	2353.9 ± 502.5	BA, BK, BH, FA, FH	580.1 ± 121.8
Fully Reincarnated	3873.7 ± 332.6	Fully Reincarnated	2954.8 ± 235.6	Fully Reincarnated	1224.4 ± 193.0

Random shuffling attack		Scaling attack	
Configuration	Average returns	Configuration	Average returns
N/A	N/A	N/A	N/A
BH	1961.5 ± 267.4	FA	1322.4 ± 441.5
FK	1548.2 ± 114.8	BA	1022.2 ± 337.6
BA, BK	1835.8 ± 202.5	BK, BH	2179.3 ± 555.0
BH, FK	1356.3 ± 412.0	FK, FH	1245.5 ± 127.2
BA, BK, FH	2254.7 ± 323.0	FA, FK, FH	2618.4 ± 208.4
FA, FK, FH	1219.9 ± 140.9	BA, BK, BH	1734.4 ± 220.2

Table 2 continued

Random shuffling attack		Scaling attack	
Configuration	Average returns	Configuration	Average returns
BH, FA, FK, FH	2029.4 ± 473.8	BK, BH, FA, FK	3141.4 ± 166.8
BA, BH, FA, FK	1472.7 ± 107.2	BK, FA, FK, FH	2476.5 ± 348.1
BA, BH, FA, FK, FH	2330.3 ± 163.3	BA, BK, BH, FK, FH	3074.7 ± 68.9
BA, BK, FA, FK, FH	2019.0 ± 368.0	BA, BH, FA, FK, FH	2774.2 ± 205.7
Fully Reincarnated	**2417.3 ± 351.1**	Fully Reincarnated	**3227.7 ± 168.2**

Bold values indicate the highest performing results

Table 3 Comparison of poisoning techniques based on evaluation metrics for different numbers of reincarnating agents

Metric	Poisoning Technique	No. of reincarnating agents					Overall Sequence
		1	2	3	4	5	
Kendall τ for max returns	Noise addition	−0.6	−0.028	0.368	0.009	0.066	−0.450
	Reversal	−0.6	−0.466	0.115	0.123	−0.6	−0.290
	Random shuffling	−0.2	0.009	0.294	−0.180	−0.2	0.152
	Scaling	0.6	−0.161	0.168	0.371	0.333	0.438
Kendall τ for avg returns	Noise addition	0.733	0.180	0.526	0.200	0.333	0.504
	Reversal	−0.733	−0.2	0.326	−0.161	−0.066	−0.026
	Random shuffling	−0.733	0.047	0.336	0.066	−0.466	0.265
	Scaling	−0.066	−0.352	−0.073	−0.028	0.2	0.381

In summary, our findings reveal that selective incarnation remains vulnerable, even in the presence of subtle changes in observations affecting all agents uniformly. Notably, certain agent combinations exhibit high susceptibility to poisoning attacks, while others display remarkable resilience. This empirical data provides valuable insights essential for informed decision-making regarding reincarnation strategies.

6 Conclusion and Future Work

This study has uncovered the critical role of agent combinations and reincarnation scenarios in determining the resilience of HalfCheetah MARL systems to different observation poisoning attacks. This highlights the importance of incorporating adaptive strategies and maintaining a balance between performance and consistency in system design.

Our research specifically focuses on the applying basic, environment-independent triggers in the context of selective reincarnation in MARL. We have demonstrated that selective reincarnation, when faced with observation poisoning, exhibits varying levels of vulnerability based on agent combinations, underscoring the necessity for targeted defenses. Through rigorous experimentation, we have quantitatively measured the impact of poisoning using Kendall's Tau metric, thereby providing a statistical foundation for assessing the robustness of MARL systems against adversarial threats. Our findings revealed that certain agent combinations are inherently more resilient, offering pathways to enhance system security and stability.

Our future research will focus on broadening the study of observation poisoning attacks in the selective reincarnation of MARL by testing various poisoning methods and advanced triggers in a variety of environments, including multi-agent Humanoid and HumanoidStandup. We aim to understand how selective reincarnation performs in cooperative, competitive, and mixed environments, assess the resilience of multi-agent systems to advanced poisoning attacks, identify and understand the effects of poisoning, and gain a comprehensive understanding of the strengths and limitations of current methods. This effort will inspire the creation of more adaptable and resilient multi-agent system strategies. Through this thorough analysis, we will highlight areas for improvement and contribute to the development of more effective defense strategies in the field.

Acknowledgements We acknowledge the Machine Learning and Optimization lab at IIITA for computational support and the Ministry of Education, Government of India, for financial support. Thanks to Avadhoot Bangal for initial discussions.

Author Contributions Putla Harsha (PH): Conceived the core concept, significantly contributed to the experimental design, and was instrumental in data collection and analysis. PH also drafted the manuscript and revised it critically for important intellectual content. Patibandla Chanakya (PC): Provided valuable inputs in the experimental design and critically reviewed and revised the manuscript, ensuring it met the requisite standards for final acceptance. PC has given consent to submit the work for publication. Krishna Pratap Singh (KPS): Collaborated with PH in refining the data analysis. KPS contributed valuable insights throughout the research process, approved the final version for publication, and agrees to be accountable for all aspects of the work in ensuring accuracy and integrity. P Nagabhushan (PN): Guided the revision process, playing a pivotal role in maintaining adherence to standards. PN critically evaluated and approved the final version for quality and accuracy, underscoring the work's intellectual content. He also agrees to be accountable for all aspects of the work to ensure proper investigation and resolution of any questions related to the accuracy or integrity of the work. All authors have read and agreed to the published version of the manuscript. Those who contributed to the work but did not meet all criteria for authorship have been listed in the Acknowledgements section.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Abdallah S, Lesser V (2008) A multiagent reinforcement learning algorithm with non-linear dynamics. J Artif Intell Res 33:521–549
2. Witt CSD, Peng B, Kamienny P-A, Torr PHS, Böhmer W, Whiteson S (2020) Deep multi-agent reinforcement learning for decentralized continuous cooperative control. arXiv:2003.06709
3. Kim DK, Liu M, Riemer MD, Sun C, Abdulhai M, Habibi G, Lopez-Cot S, Tesauro G, How J (2021) A policy gradient algorithm for learning to learn in multiagent reinforcement learning. In: International Conference on Machine Learning, pp. 5541–5550 . PMLR
4. Hu J, Wellman MP (1998) Multiagent reinforcement learning: theoretical framework and an algorithm. ICML 98:242–250
5. Bazzan AL (2009) Opportunities for multiagent systems and multiagent reinforcement learning in traffic control. Auton Agents Multi-Agent Syst 18:342–375
6. Castanyer RC (2023) Centralized control for multi-agent rl in a complex real-time-strategy game. arXiv preprint arXiv:2304.13004
7. Xiong C, Ma Q, Guo J, Lewis FL (2023) Data-based optimal synchronization of heterogeneous multiagent systems in graphical games via reinforcement learning. IEEE Trans Neural Netw Learn Syst
8. Zhang G, Li Y, Xu X, Dai H (2019) Efficient training techniques for multi-agent reinforcement learning in combat tasks. IEEE Access 7:109301–109310
9. Canese L, Cardarilli GC, Di Nunzio L, Fazzolari R, Giardino D, Re M, Spanò S (2021) Multi-agent reinforcement learning: A review of challenges and applications. Appl Sci 11(11):4948
10. Formanek C, Tilbury CR, Shock J, Tessera K-a, Pretorius A (2023) Reduce, reuse, recycle: Selective reincarnation in multi-agent reinforcement learning. In: Workshop on Reincarnating Reinforcement Learning at ICLR 2023
11. Agarwal R, Schwarzer M, Castro PS, Courville AC, Bellemare M (2022) Reincarnating reinforcement learning: Reusing prior computation to accelerate progress. Adv Neural Inf Process Syst 35:28955–28971
12. Shenfeld I, Hong Z-W, Tamar A, Agrawal P (2023) TGRL: Teacher guided reinforcement learning algorithm for POMDPs. In: Workshop on Reincarnating Reinforcement Learning at ICLR 2023 . https://openreview.net/forum?id=kTqjkIvjj7
13. Xu K, Bai C, Qiu S, He H, Zhao B, Wang Z, Li W, Li X (2023) On the value of myopic behavior in policy reuse. arXiv preprint arXiv:2305.17623
14. Rahman MM, Xue Y (2023) Accelerating policy gradient by estimating value function from prior computation in deep reinforcement learning. arXiv preprint arXiv:2302.01399
15. Xiong Z, Eappen J, Zhu H, Jagannathan S (2023) Defending observation attacks in deep reinforcement learning via detection and denoising. In: Amini M-R, Canu S, Fischer A, Guns T, Kralj Novak P, Tsoumakas G (eds) Machine Learning and Knowledge Discovery in Databases. Springer, Cham, pp 235–250
16. Zhang H, Chen H, Xiao C, Li B, Liu M, Boning D, Hsieh C-J (2020) Robust deep reinforcement learning against adversarial perturbations on state observations. Adv Neural Inf Process Syst 33:21024–21037
17. Li M, Sun Y, Lu H, Maharjan S, Tian Z (2019) Deep reinforcement learning for partially observable data poisoning attack in crowdsensing systems. IEEE Internet Things J 7(7):6266–6278
18. Gu S, Yang L, Du Y, Chen G, Walter F, Wang J, Yang Y, Knoll A (2022) A review of safe reinforcement learning: Methods, theory and applications. arXiv preprint arXiv:2205.10330

19. Schmidt LM, Kontes G, Plinge A, Mutschler C (2021) Can you trust your autonomous car? interpretable and verifiably safe reinforcement learning. In: 2021 IEEE Intelligent Vehicles Symposium (IV), pp. 171–178 . IEEE
20. Amani S, Thrampoulidis C, Yang L (2021) Safe reinforcement learning with linear function approximation. In: International Conference on Machine Learning, pp. 243–253 . PMLR
21. Thomas G, Luo Y, Ma T (2021) Safe reinforcement learning by imagining the near future. Adv Neural Inf Process Syst 34:13859–13869
22. Pfrommer S, Gautam T, Zhou A, Sojoudi S (2022) Safe reinforcement learning with chance-constrained model predictive control. In: Learning for Dynamics and Control Conference, pp. 291–303 . PMLR
23. Bastani O, Li S, Xu A (2021) Safe reinforcement learning via statistical model predictive shielding. In: Robotics: Science and Systems, pp. 1–13
24. Liu G, Lai L (2021) Provably efficient black-box action poisoning attacks against reinforcement learning. Adv Neural Inf Process Syst 34:12400–12410
25. Ma Y, Zhang X, Sun W, Zhu J (2019) Policy poisoning in batch reinforcement learning and control. Adv Neural Inf Process Syst 32
26. Wu Y, McMahan J, Zhu X, Xie Q (2023) Reward poisoning attacks on offline multi-agent reinforcement learning. Proc AAAI Conf on Artif Intell 37(9):10426–10434. https://doi.org/10.1609/aaai.v37i9.26240
27. Liu G, LAI L (2023) Efficient adversarial attacks on online multi-agent reinforcement learning. In: Oh A, Neumann T, Globerson A, Saenko K, Hardt M, Levine S (eds.) Advances in Neural Information Processing Systems, vol. 36, pp. 24401–24433 . https://proceedings.neurips.cc/paper_files/paper/2023/file/4cddc8fc57039f8fe44e23aba1e4df40-Paper-Conference.pdf
28. Li S, Guo J, Xiu J, Feng P, Yu X, Liu A, Wu W, Liu X (2023) Attacking Cooperative Multi-Agent Reinforcement Learning by Adversarial Minority Influence
29. Lu Z, Liu G, Lai L, Xu W (2024) Camouflage Adversarial Attacks on Multiple Agent Systems
30. Figura M, Kosaraju KC, Gupta V (2021) Adversarial attacks in consensus-based multi-agent reinforcement learning. In: 2021 American Control Conference (ACC), pp. 3050–3055 . IEEE
31. Rakhsha A, Radanovic G, Devidze R, Zhu X, Singla A (2020) Policy teaching via environment poisoning: training-time adversarial attacks against reinforcement learning. In: International Conference on Machine Learning, pp. 7974–7984 . PMLR
32. Guo J, Chen Y, Hao Y, Yin Z, Yu Y, Li S (2022) Towards comprehensive testing on the robustness of cooperative multi-agent reinforcement learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 115–122
33. Xu H, Wang R, Raizman L, Rabinovich Z (2021) Transferable environment poisoning: training-time attack on reinforcement learning. In: Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems, pp. 1398–1406
34. Chen Y, Zheng Z, Gong X (2022) Marnet: Backdoor attacks against cooperative multi-agent reinforcement learning. IEEE Trans Dependable Secur Comput
35. Xie Z, Xiang Y, Li Y, Zhao S, Tong E, Niu W, Liu J, Wang J (2021) Security analysis of poisoning attacks against multi-agent reinforcement learning. In: International Conference on Algorithms and Architectures for Parallel Processing, pp. 660–675 . Springer
36. Zheng H, Li X, Chen J, Dong J, Zhang Y, Lin C (2023) One4all: Manipulate one agent to poison the cooperative multi-agent reinforcement learning. Comput & Secur 124:103005
37. Peng B, Rashid T, Witt C, Kamienny P-A, Torr P, Böhmer W, Whiteson S (2021) Facmac: Factored multi-agent centralised policy gradients. Adv Neural Inf Process Syst 34:12208–12221
38. Wu J, Li H (2020) Deep ensemble reinforcement learning with multiple deep deterministic policy gradient algorithm. Math Probl Eng 2020:1–12
39. Formanek C, Jeewa A, Shock J, Pretorius A (2023) Off-the-Grid MARL: datasets with Baselines for Offline Multi-Agent Reinforcement Learning
40. Kendall MG (1938) A new measure of rank correlation. Biometrika 30(1/2):81–93
41. Kono H, Kamimura A, Tomita K, Murata Y, Suzuki T (2014) Transfer learning method using ontology for heterogeneous multi-agent reinforcement learning. Int J Adv Comput Sci Appl 5(10)
42. Gao Z, Xu K, Ding B, Wang H (2021) Knowru: Knowledge reuse via knowledge distillation in multi-agent reinforcement learning. Entropy 23(8):1043
43. Nekoei H, Badrinaaraayanan A, Courville A, Chandar S (2021) Continuous coordination as a realistic scenario for lifelong learning. In: Meila M, Zhang T (eds.) Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 8016–8024 . https://proceedings.mlr.press/v139/nekoei21a.html
44. Hussenot L, Geist M, Pietquin O (2019) Targeted attacks on deep reinforcement learning agents through adversarial observations. arXiv preprint arXiv:1905.12282

45. Qiaoben Y, Ying C, Zhou X, Su H, Zhu J, Zhang B (2021) Understanding adversarial attacks on observations in deep reinforcement learning. Science China Information Sciences
46. Ashcraft C, Karra K (2021) Poisoning deep reinforcement learning agents with in-distribution triggers. In: ICLR 2021 Workshop on Security and Safety in Machine Learning Systems . https://aisecure-workshop.github.io/aml-iclr2021/papers/11.pdf
47. Kiourti P, Wardega K, Jha S, Li W (2020) Trojdrl: evaluation of backdoor attacks on deep reinforcement learning. In: 2020 57th ACM/IEEE Design Automation Conference (DAC), pp. 1–6 . IEEE
48. Rauber J, Brendel W, Bethge M (2017) Foolbox: A python toolbox to benchmark the robustness of machine learning models. In: Reliable Machine Learning in the Wild Workshop, 34th International Conference on Machine Learning . arXiv:1707.04131
49. Adeyemo A, Khalid F, Odetola T, Hasan SR (2021) Security analysis of capsule network inference using horizontal collaboration. In: 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), pp. 1074–1077 . IEEE
50. Voss JR, Rademacher L, Belkin M (2013) Fast algorithms for gaussian noise invariant independent component analysis. In: Burges CJ, Bottou L, Welling M, Ghahramani Z, Weinberger KQ (eds.) Advances in Neural Information Processing Systems, vol. 26 . https://proceedings.neurips.cc/paper_files/paper/2013/file/4d2e7bd33c475784381a64e43e50922f-Paper.pdf
51. Zhang X, Zhang W, Gong Y, Yang L, Zhang J, Chen Z, He S (2023) Robustness evaluation of multi-agent reinforcement learning algorithms using gnas
52. Tekgul BG, Wang S, Marchal S, Asokan N (2022) Real-time adversarial perturbations against deep reinforcement learning policies: attacks and defenses. In: European Symposium on Research in Computer Security, pp. 384–404. Springer
53. Korkmaz E (2021) Non-robust feature mapping in deep reinforcement learning. In: ICML 2021 Workshop on Adversarial Machine Learning
54. Standen M, Kim J, Szabo C (2023) Sok: Adversarial machine learning attacks and defences in multi-agent reinforcement learning. arXiv preprint arXiv:2301.04299
55. Korkmaz E (2021) Investigating vulnerabilities of deep neural policies. In: Uncertainty in Artificial Intelligence, pp. 1661–1670 . PMLR
56. Korkmaz E (2021) Adversarial training blocks generalization in neural policies. In: NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications
57. Quiring E, Rieck K (2020) Backdooring and poisoning neural networks with image-scaling attacks. In: 2020 IEEE Security and Privacy Workshops (SPW), pp. 41–47 . IEEE
58. Hu C, Shi W (2022) Impact of scaled image on robustness of deep neural networks. arXiv preprint arXiv:2209.02132
59. Wang Z, Zhang S, Li Y, Pan Q (2023) Dba: downsampling-based adversarial attack in medical image analysis. In: Third International Conference on Computer Vision and Pattern Analysis (ICCPA 2023), vol. 12754, pp. 220–227 . SPIE
60. Bernstein DS, Givan R, Immerman N, Zilberstein S (2002) The complexity of decentralized control of markov decision processes. Math Oper Res 27(4):819–840
61. Lowe R, Wu YI, Tamar A, Harb J, Pieter Abbeel O, Mordatch I (2017) Multi-agent actor-critic for mixed cooperative-competitive environments. Adv. Neural Inf Process Syst 30
62. Wawrzynski P (2007) Learning to control a 6-degree-of-freedom walking robot. In: EUROCON 2007-The International Conference On "Computer as a Tool", pp. 698–705 . IEEE
63. Todorov E, Erez T, Tassa Y (2012) Mujoco: A physics engine for model-based control. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5026–5033 . IEEE
64. Brockman G, Cheung V, Pettersson L, Schneider J, Schulman J, Tang J, Zaremba W (2016) OpenAI Gym