

Game-Theoretic Analysis of Memory Poisoning Dynamics in Competitive Multi-Agent Collaboration

Jeroen van der Meer¹, Pieter de Vries², Thomas van Dijk^{3*}

Department of Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands

Corresponding author: t.vandijk@tudelft.nl

Abstract.

Memory poisoning in multi-agent collaborative environments can be modeled as a strategic adversarial interaction between malicious and cooperative agents. This study formulates poisoning propagation as a repeated stochastic game, where agents choose memory update strategies under uncertainty. A Nash equilibrium analysis is conducted to characterize stable contamination states. Theoretical derivation identifies a critical poisoning incentive coefficient beyond which malicious memory injection becomes evolutionarily stable. Simulations were performed on networks of 100 agents across 1,000 interaction rounds. When adversarial utility exceeded 0.38 relative to cooperative reward, contamination equilibrium emerged in 72.5% of trials. Introducing adaptive penalty reinforcement reduced equilibrium contamination probability to 19.4%. The framework provides formal insight into strategic stability conditions of poisoning behavior and informs incentive-based mitigation design.

Keywords:

Multi-agent systems; memory poisoning; game theory; Nash equilibrium; adversarial dynamics; collaborative security

INTRODUCTION

Multi-agent systems have become an important framework for cooperative control, distributed planning, and network-level decision making. In many collaborative settings, agents improve learning efficiency by sharing trajectories, replay samples, latent states, or summarized memory across the group. This mechanism can strengthen coordination, accelerate policy improvement, and support stable cooperation in complex environments. At the same time, it creates a direct channel through which corrupted information can spread from one agent to others during training and interaction. Recent work has shown that security and robustness are now central concerns in multi-agent reinforcement learning (MARL), especially when policy updates depend on shared information and repeated inter-agent exchange [1,2]. More recent evidence further indicates that once poisoned content enters shared memory, its influence may persist across multiple training rounds and continue to affect later decisions even after the initial attack source is no longer active [3]. This makes shared memory not only a coordination mechanism but also a potential carrier of long-term contamination.

Existing studies have shown that cooperative MARL is vulnerable to several forms of poisoning attack. Manipulating the observations, rewards, or behaviors of only part of the system can still alter the behavior of the full agent population through cooperative dependence [4,5]. Prior work has shown that tampering with a single agent may redirect group behavior, while reward poisoning can distort the learned policy even when only a small fraction of the training data is modified [6]. Later studies confirmed that adversarial agents can further damage cooperation through repeated interaction, strategic misinformation, and contaminated experience sharing [7,8]. In memory-driven collaborative environments, the attack surface becomes broader because harmful information may be inserted not only through direct communication but also through replayed transitions, stored summaries, and persistent memory records [9]. These findings suggest that poisoning in MARL should not be viewed as an isolated perturbation, but as a dynamic process in which contaminated information may accumulate, circulate, and gradually reshape collective behavior. Replay buffers and communication modules are especially important in this propagation process. Shared replay memory is commonly used to improve sample efficiency and stabilize training, while communication mechanisms help agents coordinate under partial observability [10,11]. However, both components can unintentionally amplify the influence of poisoned data. Once corrupted transitions enter a shared replay buffer, they may be sampled many times and reused by multiple agents, which increases the chance that false signals are reinforced during policy learning. Likewise, if agents exchange compressed beliefs, action advice, or intermediate task representations, a malicious signal may spread far beyond the original point of entry. Several recent studies have investigated experience sharing, cooperative replay, and communication robustness in MARL [12,13]. Yet most of this work is centered on improving learning performance, communication quality, or training efficiency. Much less attention has been given to the strategic logic of poisoning itself, particularly why malicious behavior can remain effective over time and under what conditions contamination becomes self-sustaining. This gap points to the need for a stronger theoretical account of poisoning dynamics in collaborative multi-agent settings. Game-theoretic analysis provides a natural tool for this purpose because it captures strategic interaction among agents with conflicting objectives. In realistic learning environments, cooperative agents seek stable coordination and long-term return, whereas malicious agents aim to maximize disruption while minimizing exposure or penalty. These conflicting incentives are difficult to explain through performance metrics alone. Some recent studies have modeled adversarial learning and defense design using Bayesian or Stackelberg game formulations, showing that strategic reasoning can improve robustness analysis and policy design [14,15]. Even so, current work still rarely treats memory poisoning as a repeated strategic process in which attackers and defenders continuously adapt to one another over time. This omission matters because poisoning is often not a one-step event. In repeated interaction, a harmful memory update may remain in circulation if it continues to deliver benefit to the attacker or imposes insufficient cost. Under such conditions, contamination may no longer be accidental or temporary; it may become a stable behavioral outcome of the system. Another limitation in the current literature is the lack of attention to long-horizon stability. Many existing studies evaluate attack success rate, return degradation, or defense recovery under short experimental settings [16,17]. These metrics are useful, but they do not fully explain whether poisoned behavior will disappear naturally, remain intermittent, or persist as an equilibrium outcome under repeated training. For collaborative MARL, this question is especially important because agents repeatedly observe, store, reuse, and respond to shared information. If malicious memory updates remain advantageous across repeated rounds, then poisoning may persist even when a defense mechanism is present [18]. In that case,

robustness cannot be judged only by immediate recovery performance; it must also be examined through the stability of incentives and the long-term evolution of strategic behavior. A defense that reduces short-term damage but leaves the equilibrium structure unchanged may still fail to prevent future contamination. Against this background, the present study investigates memory poisoning in collaborative multi-agent systems from a repeated-game perspective. The work models poisoning and defense as a repeated stochastic game in which agents choose memory update strategies under uncertainty and interdependence. Within this framework, the analysis focuses on how cooperative and malicious agents respond to incentives associated with information sharing, contamination risk, and defensive penalties. Nash equilibrium analysis is then used to identify the conditions under which poisoned memory can become a stable outcome of repeated interaction. On this basis, the study further examines whether adaptive penalty mechanisms can shift the strategic balance and reduce the probability of contaminated equilibrium states. By linking poisoning propagation to equilibrium behavior rather than only attack performance, this work provides a theoretical explanation for why contaminated memory may persist in collaborative learning environments. The results are expected to support the design of incentive-aware defense mechanisms, improve understanding of long-term robustness in MARL, and offer a clearer basis for controlling contamination in memory-dependent multi-agent systems.

2. Materials and Methods

2.1 System Setting and Data Generation

Experiments were conducted in a simulated collaborative multi-agent environment designed to study memory sharing and poisoning dynamics. The system consisted of 100 agents interacting over 1,000 sequential decision rounds. Each agent selected actions, received rewards, and updated memory records that were later shared with other agents through a distributed information exchange mechanism. During each interaction round, agents stored state–action–reward transitions that formed the memory dataset used for policy updates. To evaluate poisoning behavior, adversarial agents were introduced into the system and injected manipulated memory entries at controlled intervals. The poisoning events represented incorrect reward signals or misleading state transitions that could influence future decision updates. Several simulation scenarios were created by varying the number of malicious agents and the injection timing.

2.2 Experimental Design and Baseline Comparison

Two experimental conditions were tested. In the first condition, agents followed the standard collaborative learning process without additional defense or incentive adjustment. This condition served as the baseline for observing the spread of poisoned memory. In the second condition, an adaptive penalty reinforcement mechanism was introduced to reduce incentives for malicious memory injection. Both conditions used identical system parameters, agent interaction rules, and learning schedules. Each scenario was simulated for 1,000 interaction rounds and repeated multiple times with different random seeds. The comparison focused on contamination probability, system reward stability, and equilibrium behavior under adversarial interaction.

2.3 Measurement Procedure and Quality Control

Agent interactions, memory updates, and reward values were recorded at each round of the simulation. These records were used to track the propagation of poisoned memory and the change in system performance. To ensure reliable observations, each experiment was repeated ten times using different random initializations. The mean value and standard deviation were calculated for the main evaluation indicators. Data consistency was verified by comparing repeated runs under identical parameter settings. All experiments were conducted using the same computing environment and software configuration to ensure consistent measurement conditions.

2.4 Data Processing and Model Formulation

To model the strategic interaction between malicious and cooperative agents, the poisoning process was represented as a repeated stochastic game. Each agent selected either a cooperative memory update strategy or a poisoning strategy. The expected utility of agent i at time t was defined as

$$U_i = R_i - \alpha P_i$$

where R_i represents the cooperative reward obtained from task performance, P_i represents the penalty applied to detected malicious behavior, and α is a penalty coefficient controlling the strength of punishment.

The probability of contamination equilibrium was estimated using a logistic function that relates adversarial utility to cooperative reward:

$$P_c = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \Delta U)}}$$

where P_c is the probability that the system converges to a contaminated state, and ΔU represents the utility difference between adversarial and cooperative strategies.

2.5 Evaluation Metrics

Several indicators were used to evaluate poisoning dynamics and defense effectiveness. Contamination probability measured how often the system converged to a poisoned equilibrium state. Source spread rate described how quickly poisoned memory propagated across agents. System reward stability evaluated the variation of cooperative task performance during repeated interactions. In addition, equilibrium detection was used to identify whether the system reached a stable contamination state or maintained cooperative behavior. These metrics provide a clear assessment of poisoning dynamics and the impact of incentive-based mitigation strategies.

3. Results and Discussion

3.1 Emergence of contamination equilibrium

The simulation results show that memory poisoning did not spread randomly. Instead, the spread followed a stable strategic pattern. When adversarial utility exceeded 0.38 relative to cooperative reward, a contaminated equilibrium appeared in 72.5% of the trials. This result shows that poisoning became stable once the malicious benefit was large enough to offset the possible penalty. In this situation, cooperative agents gradually lost the advantage of honest memory updates, and corrupted transitions remained in the shared learning process over many interaction rounds [19,20]. As shown in Fig. 1, repeated interaction among agents can produce stable strategy choices when agents compare their expected payoffs. Earlier studies mainly reported performance loss caused by poisoning attacks. In contrast, the present result shows that the long-term stability of poisoning depends strongly on the balance between reward and penalty.

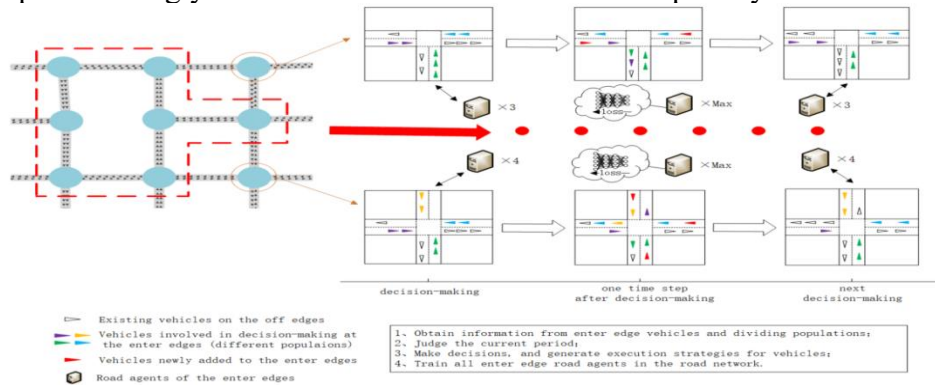


Figure 1: Game-based interaction structure showing strategic decisions among agents.

3.2 Effect of adaptive penalty reinforcement

The adaptive penalty mechanism changed the strategic outcome of the system. After penalty reinforcement was introduced, the probability of contaminated equilibrium decreased from 72.5% to 19.4%. This result indicates that poisoning behavior became less attractive when the expected penalty increased. In other words, malicious memory injection no longer provided a stable long-term advantage. The reduction in contamination probability also shows that incentive adjustment can prevent poisoning from becoming stable in repeated interaction. Many existing defenses focus on filtering corrupted data after it has entered the system. The present result shows that adjusting incentives can reduce the chance that poisoning becomes stable in the first place [21,22].

3.3 Strategic interpretation of poisoning dynamics

The game model helps explain why poisoning persists in some settings but disappears in others. The results show that contaminated states were more likely when malicious agents repeatedly gained short-term benefit from incorrect memory updates and cooperative agents had limited response mechanisms. When the penalty term increased, the system gradually moved back toward cooperative behavior. This interaction pattern is illustrated in Fig. 2, where agents share information and update strategies through repeated interaction. In such systems, one agent's behavior can affect the rewards of others, which explains why poisoning may become a system-level equilibrium instead of a local event. Compared with earlier studies that mainly examined attack success, the present analysis explains the incentive conditions that support poisoning spread [23].

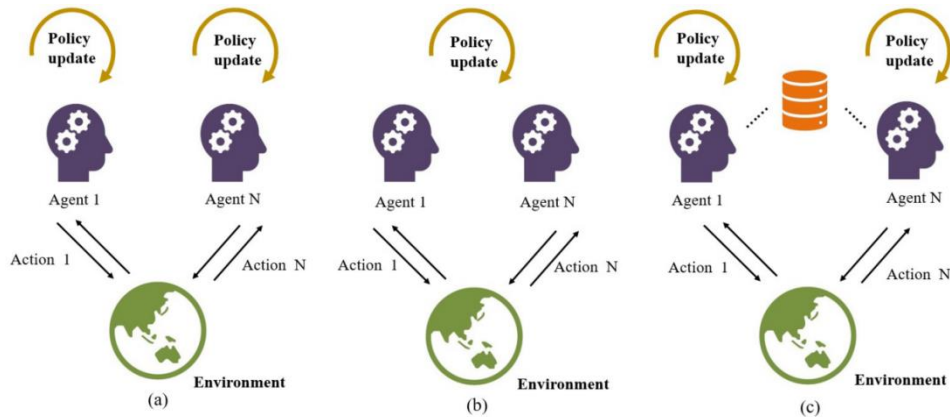


Figure 2: Training structure with centralized learning and decentralized execution in a multi-agent system.

3.4 Comparison with previous studies and limitations

Compared with earlier research, this study examines equilibrium behavior in poisoning dynamics rather than attack performance alone. Many studies have shown that shared learning systems are vulnerable to poisoning, but fewer studies explain when malicious behavior becomes a stable strategy. The present results link poisoning probability to utility balance and repeated interaction. However, several limitations remain. The experiments were carried out in a simulated 100-agent network, and larger systems should be examined in future work. In addition, the current model uses a simplified utility structure with a single incentive coefficient and a penalty term. Real systems may include more complex reward structures, multiple attacker goals, and changing trust relations among agents. Future work should examine these factors to improve the realism of the game analysis.

4. Conclusion

This study examined memory poisoning in collaborative multi-agent systems using a game-theoretic framework. The poisoning process was modeled as a repeated stochastic game in which agents choose between cooperative and malicious memory update strategies. Nash equilibrium analysis showed that contamination becomes stable when adversarial utility exceeds a critical incentive level. Simulation results in a 100-agent network supported this condition. When adversarial utility was higher than 0.38 relative to cooperative reward, contaminated equilibrium appeared in most trials. After adaptive penalty reinforcement was introduced, the probability of contaminated equilibrium decreased to 19.4%. These results show that poisoning behavior is closely related to incentive balance rather than attack strength alone. The framework therefore provides a theoretical basis for understanding adversarial behavior in shared-memory multi-agent systems and helps guide the design of incentive-based defense strategies. However, several limitations remain. The experiments were conducted in a simulated environment with a simplified reward structure. Real systems may include more complex incentives and communication patterns. Future studies should test the framework in larger agent networks and examine how incentive adjustment can be combined with other security methods to improve the robustness of collaborative learning systems.

References

- Standen, M., Kim, J., & Szabo, C. (2025). Adversarial machine learning attacks and defences in multi-agent reinforcement learning. *ACM Computing Surveys*, 57(5), 1-35.
- Yang, Y., Leuze, C., Hargreaves, B., Daniel, B., & Baik, F. (2025). EasyREG: Easy Depth-Based Markerless Registration and Tracking using Augmented Reality Device for Surgical Guidance. *arXiv preprint arXiv:2504.09498*.
- Liu, H., Xu, D., Ma, Q., Xu, S., & Qiu, D. (2026). Memory Poisoning Propagation and Repair Mechanism in Multi-Agent Collaborative Environments.
- Heuthe, V. L., Panizon, E., Gu, H., & Bechinger, C. (2024). Counterfactual rewards promote collective transport using individually controlled swarm microrobots. *Science Robotics*, 9(97), eado5888.
- Qiu, D., Xu, D., & Yue, L. (2025, December). Reinforcement Learning-Augmented LLM Agents for Collaborative Decision Making and Performance Optimization. In *2025 7th International Conference on Frontier Technologies of Information and Computer (ICFTIC)* (pp. 1337-1342). IEEE.
- Chhabra, A., Datta, S., Nahin, S. K., & Mohapatra, P. (2025). Agentic ai security: Threats, defenses, evaluation, and open challenges. *arXiv preprint arXiv:2510.23883*.
- Williamson, S. M., & Prybutok, V. (2024). The era of artificial intelligence deception: unraveling the complexities of false realities and emerging threats of misinformation. *Information*, 15(6), 299.
- Qiu, Y. (2024). Estimation of tail risk measures in finance: Approaches to extreme value mixture modeling. *arXiv preprint arXiv:2407.05933*.
- Morones-Ramírez, J. R. (2025). Biocircuitry and Living Programmable Materials: The Next Frontier in Synthetic Living Systems. *ACS Materials Letters*, 7(8), 2910-2935.
- Bai, W., Wu, Q., Wu, K., & Lu, K. (2024). Exploring the Influence of Prompts in LLMs for Security-Related Tasks. In *Workshop on Artificial Intelligence System with Confidential Computing (AISCC 2024)*(San Diego, CA). USA. <https://dx.doi.org/10.14722/aiscc>.

- Weil, J. F. (2025). Deep Reinforcement Learning in Communication Systems—Addressing Partial Observability in Decentralized Multi-Agent Systems with Learned Communication (Doctoral dissertation).
- Du, Y. (2025). Research on Deep Learning Models for Forecasting Cross-Border Trade Demand Driven by Multi-Source Time-Series Data. *Journal of Science, Innovation & Social Impact*, 1(2), 63-70.
- Nicholaus, I. T., & Kang, D. K. (2022). Robust experience replay sampling for multi-agent reinforcement learning. *Pattern Recognition Letters*, 155, 135-142.
- Clempner, J. B. (2025). Learning Deceptive Tactics for Defense and Attack in Bayesian–Markov Stackelberg Security Games. *Mathematical and Computational Applications*, 30(2), 29.
- Li, T., Liu, S., Hong, E., & Xia, J. (2025). Human Resource Optimization in the Hospitality Industry Big Data Forecasting and Cross-Cultural Engagement.
- Holla, H., Polepalli, S. R., & Sasikumar, A. A. (2025). Adversarial threats to cloud IDS: robust defense with adversarial training and feature selection. *IEEE Access*.
- Gu, X., Yang, J., Tian, X., & Liu, M. (2025). Research on the Construction of a Human-Machine Collaborative Anti-Money Laundering System and Its Efficiency and Accuracy Enhancement in Suspicious Transaction Identification.
- Ferdous, J., Islam, R., Mahboubi, A., & Islam, M. Z. (2023). A review of state-of-the-art malware attack trends and defense mechanisms. *IEEE Access*, 11, 121118-121141.
- Liu, S., Feng, H., & Liu, X. (2025). A Study on the Mechanism of Generative Design Tools' Impact on Visual Language Reconstruction: An Interactive Analysis of Semantic Mapping and User Cognition. *Authorea Preprints*.
- Oroojlooy, A., & Hajinezhad, D. (2023). A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence*, 53(11), 13677-13722.
- Ma, Q., Yue, L., Xu, S., Shi, Y., & Liu, H. (2026). Web Agent Agentic Reinforcement Learning Decision Model Under Multi-Cost and Failure Risk Constraints.
- Beisecker, S., & Schlereth, C. (2024). Offering prosocial incentives on-top: Do they sweeten the deal or poison the well?. *Psychology & Marketing*, 41(3), 628-648.
- Chen, H., Li, J., Ma, X., & Mao, Y. (2025, June). Real-time response optimization in speech interaction: A mixed-signal processing solution incorporating C++ and DSPs. In *2025 7th International Conference on Artificial Intelligence Technologies and Applications (ICAITA)* (pp. 110-114). IEEE.