УДК 004.896 EDN EWGAAM

Адаптивный пороговый метод обнаружения сговора агентов в рое БПЛА при распределённом обучении

В.И. Пронин*, С.А. Горяйнов, Н. Дибров, Д.А. Диканский, К.А. Неборачек

Северо-Кавказский федеральный университет, Ставрополь, Россия

*E-mail: vladinfotech23@gmail.com

Аннотация. Предложен адаптивный пороговый метод обнаружения сговора в рое БПЛА, пересчитывающий порог по темпу изменения награды, разбросу Q-оценок и числу соседей. Эксперименты с восемью дронами показали: полнота обнаружения сговора выросла на 25-30%, ложных тревог стало на треть меньше. Метод не требует переобучения, практически не нагружает бортовой вычислитель и не вмешивается в протоколы аутентификации.

Ключевые слова: рой БПЛА, сговор агентов, адаптивный порог, распределённое обучение, обнаружение аномалий, взаимная информация, DQN.

Adaptive threshold method for detecting agent collusion in a UAV swarm under distributed learning

V.I. Pronin, S.A. Goryainov, N. Dibrov, D.A. Dikansky, K.A. Neborachek

North-Caucasus Federal University, Stavropol, Russia

*E-mail: vladinfotech23@gmail.com

Abstract. An adaptive threshold method for detecting collusion is proposed, where the threshold is recalculated based on the reward change rate, Q-value estimates, and the number of neighbors within communication range. Experiments with eight drones showed that collusion detection recall increased by 25-30%, while the false alarm rate dropped by one third. The method requires no retraining, adds negligible computational load, and does not interfere with authentication protocols.

Keywords: UAV swarm, agent collusion, adaptive threshold, distributed learning, anomaly detection, mutual information, DQN.

1. Введение

Распределённое обучение в рое БПЛА основано на обмене опытом между агентами. Это ускоряет сходимость, но открывает возможность для скоординированных атак. Группа агентов может договориться и действовать в собственных интересах, внешне сохраняя нормальное поведение. Например, два дрона начинают завышать взаимные награды или синхронно избегать невыгодных действий. Другой вариант – подавление сигналов от честных участников. Такой сговор практически незаметен, однако постепенно искажает общую политику роя.

Стандартный подход к обнаружению подобных аномалий – вычисление метрики согласованности действий, например взаимной информации, и сигнализация тревоги при превышении фиксированного порога. В стационарной среде этот приём работает. Но рой БПЛА не статичен: меняется плотность, дроны переходят из транзитной фазы в рабочую, появляются новые препятствия. В таких условиях фиксированный порог становится бесполезен. Низкое значение вызывает частые ложные тревоги и мешает работе, высокое – пропускает сговор.

В данной работе предлагается отказаться от подбора одного порога и сделать его адаптивным. Логика проста. В нестабильной среде нормальные корреляции действий также возрастают, поэтому порог должен повышаться. При высоком уровне шума в оценках агентов порог тоже поднимается, иначе детектор начнёт принимать случайный шум за сговор. Наконец, чем больше соседей у дрона, тем выше вероятность случайной корреляции – порог должен расти, но не линейно, а логарифмически.

2. Обзор литературы

Задача обнаружения аномалий в распределённом обучении привлекает всё большее внимание исследователей. Wu и др. [1] показали, что группа агентов способна имитировать равновесие Нэша, скрытно искажая обучение через отравление данных. В обзорной работе Standen с соавторами [2] представлена комплексная классификация атак adversarial machine learning на многоагентное обучение с подкреплением, включая анализ уязвимостей MARL к скоординированным атакам execution-time типа.

Проблема скрытого сговора с использованием стеганографических протоколов исследуется в работе Motwani и др. [3], где авторы формализуют задачу выявления нелегальной координации между агентами и предлагают методы её детектирования. В обзоре Ghaemi [4], представленном на конференции NeurIPS 2025, анализируются три

фундаментальные стратегии сговора в LLM-агентах: негласная координация через поведенческое обучение, формирование неформальных картелей и скрытая стеганографическая коллаборация. Авторы выделяют приоритетные направления разработки робастных методов обнаружения, отличающих сговор от легитимной кооперации.

В контексте кибербезопасности распределённых систем Finistrella и др. [5] предлагают классификацию методов MARL для защиты сетевой инфраструктуры, подчёркивая важность адаптивных механизмов обнаружения атак.

Проведённый анализ выявляет три ограничения существующих методов. Во-первых, порог обнаружения в большинстве подходов фиксирован и не адаптируется к изменяющимся условиям среды. Во-вторых, не учитывается структура сговора, когда агенты могут коррелировать действия, не изменяя индивидуальные награды. В-третьих, отсутствует обратная связь от истории ложных срабатываний для коррекции порога. Устранение этих ограничений – цель настоящей статьи.

3. Постановка задачи

Рассматривается рой из N БПЛА. Каждый агент i обучает собственную политику π_i и периодически обменивается данными с соседями. Среда нестационарна: скрытый параметр θ_t меняется во времени – скачкообразно или плавно.

Сговором считается подгруппа C размера не менее двух агентов, согласованно отклоняющихся от нормального обучения. На практике это проявляется как завышение взаимных наград, аномально высокая корреляция действий без видимой внешней причины или синхронное занижение Q-значений для действий, выгодных другим агентам.

Детектор на каждом шаге t вычисляет статистику $\xi(t)$, отражающую степень скоординированности действий в рое. При превышении порога $\tau(t)$ фиксируется подозрение на сговор. Требуется построить правило вычисления $\tau(t)$, которое автоматически подстраивается под текущие условия, не требует переобучения всей системы и не вносит заметных задержек в контур управления.

4. Предлагаемый метод

Базовая статистика $\xi(t)$ вычисляется как средняя взаимная информация действий по всем парам агентов предполагаемой группы на скользящем окне длины W. В экспериментах $W = 200$ шагов. Сама по себе эта метрика известна; проблема – в выборе порога для сравнения.

Вместо одного числа предлагается порог, зависящий от трёх величин:

- Нестационарность среды $\Delta(t)$ – скользящее среднее абсолютного изменения награды агентов за 50 последних эпизодов. При резких изменениях среды нормальная корреляция действий временно возрастает, и порог должен быть выше, чтобы избежать ложных тревог.
- Шум оценок $\sigma_Q(t)$ – дисперсия Q-значений по рою (или по отдельному агенту). Когда оценки сильно варьируются, метрика согласованности становится ненадёжной, и порог поднимается.
- Плотность роя $D(t)$ – число соседей в радиусе связи. Влияние плотности взято логарифмическим, так как каждый дополнительный агент даёт убывающий прирост вероятности случайной корреляции.

Простейший вариант порога – аналитическая формула:

$$\tau(t) = \tau_0 \cdot \left(1 + \alpha \cdot \Delta(t) - \beta \cdot \sigma_Q(t)\right) \cdot \left(1 + \gamma \cdot ln\left(1 + D(t)\right)\right),$$

где коэффициенты α, β, γ однократно подбираются по имитационным прогонам и не требуют обучения на борту.

Второй вариант – четырёхнейронный многослойный перцептрон (MLP), на вход которого подаются Δ, σ_Q, D и $\xi(t-1)$, а на выходе формируется τ. Он даёт несколько более высокую точность, но требует предварительного обучения на синтетических данных. Для практического развёртывания на рое проще применять аналитическую формулу, так как она прозрачна и не нуждается в дополнительных обоснованиях.

В метод введена обратная связь. При длительном отсутствии подтверждённых сговоров (например, оператор классифицирует тревоги как ложные) порог постепенно снижается. Если сговор подтверждается, порог кратковременно повышают, чтобы подавить лавину повторных срабатываний по инерции.

Метод функционирует параллельно с основным обучением, не затрагивая логику аутентификации и согласования ключей. Каждые 100 шагов детектор вычисляет $\xi(t)$, формирует $\tau(t)$ и при превышении генерирует событие. Вычислительные затраты минимальны: на Cortex M4 подсчёт взаимной информации по окну в 200 шагов занимает единицы миллисекунд, расчёт адаптивного порога – микросекунды.

5. Экспериментальная среда

Эксперименты проводились в симуляторе навигации на сетке 10×10. Рой состоял из 8 БПЛА; задача – достижение целевых точек при наличии динамических препятствий. Каждый

агент обучался с помощью DQN с буфером воспроизведения объёмом 2000 эпизодов и ε-жадной стратегией (ε затухало с 1,0 до 0,01 за 1000 эпизодов). Обмен эпизодами выполнялся каждые 40 шагов.

Рассмотрены три сценария сговора:

- Сценарий А (простой). Два соседних агента завышают Q-значения друг друга в два раза.
- Сценарий Б (скрытая координация). Три агента синхронно избегают зон, потенциально полезных для обучения. По отдельности их поведение не выделяется, скоординированность проявляется только при совместном рассмотрении.
- Сценарий В (нестационарный). Среда циклически меняет конфигурацию препятствий каждые 500 эпизодов, всего четыре цикла. Сговор сохраняется на всём протяжении.

В качестве базового метода использовался тот же детектор с фиксированным порогом 0,15. Порог был вручную подобран под сценарий А и не настраивался отдельно для других сценариев, чтобы не создавать искусственного преимущества для фиксированного подхода.

Результаты приведены в таблице 1 (полнота / точность).

Таблица 1. Результаты экспериментов различных сценариев.

Сценарий	Фиксированный порог	Адаптивный (формула)	Адаптивный (MLP)
А	72% / 68%	92% / 91%	95% / 93%
Б	58% / 51%	86% / 88%	90% / 92%
В	44% / 39%	83% / 85%	88% / 89%

В среднем по всем сценариям прирост полноты составил 25-30%, доля ложных срабатываний уменьшилась примерно на треть. Сравнение аналитического варианта с MLP показывает, что последний выигрывает 2-5 процентных пунктов, однако ценой необходимости предварительного обучения и некоторой потери интерпретируемости.

6. Анализ результатов

Предложенный метод продемонстрировал устойчивость к трём типовым искажениям, характерным для реальной эксплуатации роя. Его главное достоинство – простота реализации и отсутствие необходимости переобучать основные алгоритмы управления.

Вместе с тем следует отметить ограничения. Во-первых, метод предполагает наличие единственной группы сговора. При нескольких независимых коалициях статистика $\xi(t)$ усредняется, и слабый сговор может быть пропущен. Во-вторых, коэффициенты α, β, γ

подбирались эмпирически для симулированной среды; при переносе на реальные полётные данные может потребоваться их перекалибровка. Наконец, логарифмическая зависимость от плотности, хотя и физически обоснована, является аппроксимацией и в экстремально плотных или разреженных роях может нуждаться в уточнении.

7. Заключение

Разработан адаптивный пороговый метод обнаружения сговора в рое БПЛА при распределённом обучении. Порог подстраивается под нестационарность среды, уровень шума в Q-оценках и плотность роя. Метод не затрагивает основной протокол аутентификации и требует минимальных вычислительных ресурсов. В симуляции навигационной задачи выигрыш по полноте и точности составил 25-30% относительно фиксированного порога, доля ложных тревог снизилась примерно на треть.

Дальнейшие направления работы включают: обнаружение нескольких независимых сговоров одновременно; адаптацию коэффициентов по реальным полётным данным.

Список литературы

1. Data Poisoning to Fake a Nash Equilibria for Markov Games / Y. Wu, J. Mcmahan, X. Zhu, Q. Xie // Proceedings of the AAAI Conference on Artificial Intelligence. – 2024. – Vol. 38, No. 14. – P. 15979-15987. – DOI 10.1609/aaai.v38i14.29529. – EDN WKNLVO.

2. Standen, M. Adversarial Machine Learning Attacks and Defences in Multi-Agent Reinforcement Learning / M. Standen, J. Kim, C. Szabo // ACM Computing Surveys. – 2025. – Vol. 57, No. 5. – 124. – P. 1-35. – DOI: 10.1145/3708320.

3. Motwani, S.R. Secret Collusion among AI Agents: Multi-Agent Deception via Steganography / S.R. Motwani, et al. // 38th Conference on Neural Information Processing Systems. – 10-15 December 2024, Vancouver, Canada. – Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024. – P. 73439-73486. – DOI: 10.52202/079017-2336.

4. Ghaemi, S. A Survey of Collusion Risk in LLM-Powered Multi-Agent Systems / S. Ghaemi // NeurIPS 2025 Workshop on Reliable and Responsible AI. – Mexico City, 2025. – P. 1-13.

5. Finistrella, S. Multi-Agent Reinforcement Learning for Cybersecurity: Classification and survey / S. Finistrella, S. Mariani, F. Zambonelli // Intelligent Systems with Applications. – 2025. – Vol. 26. – P. 200495. – DOI 10.1016/j.iswa.2025.200495. – EDN ZVMIJA.