

М. М. СТЕЦЮК

аспірант кафедри програмних засобів і технологій  
Херсонський національний технічний університет  
ORCID: 0009-0002-2306-6056

Р. М. ЗАХАРЧЕНКО

кандидат технічних наук,  
доцент кафедри програмних засобів і технологій  
Херсонський національний технічний університет  
ORCID: 0000-0003-4650-3095

## РОЛЬ-ОРІЄНТОВАНЕ MULTI-AGENT REINFORCEMENT LEARNING ДЛЯ АВТОНОМНОГО РОЗПОДІЛУ ЗАДАЧ У ГЕТЕРОГЕННОМУ РОЇ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ

У статті розглянуто проблему побудови адаптивних моделей керування гетерогенними роями безпілотних літальних апаратів (БпЛА), що діють у динамічних умовах із частковою спостережуваністю та обмеженими ресурсами. Традиційні підходи – статичний розподіл ролей, евристичні алгоритми чи однорівневі моделі підкріплювального навчання – часто не забезпечують достатньої гнучкості та стабільності в задачах, які потребують узгодженої взаємодії різнотипних агентів. Для подолання цих обмежень запропоновано роль-орієнтовану архітектуру multi-agent reinforcement learning (MARL), у якій поведінка кожного БпЛА описується двошаровою політикою: верхній рівень відповідає за вибір ролі, а нижній – за формування дії, що залежить від контексту місії та особливостей агента.

Метою роботи є розроблення формальної моделі роль-орієнтованого MARL та проведення аналітичної оцінки її ефективності для автономного розподілу задач у гетерогенних роях БпЛА. У межах дослідження побудовано математичний опис політики ролей і дій, визначено функцію нагороди, що враховує кооперативні показники місії, та сформульовано оптимізаційну задачу навчання. Проведено теоретичне моделювання очікуваної поведінки рою у типових кооперативних сценаріях, що включають розвідку, пошук і рятування та доставку ресурсів. Аналітичне порівняння з базовими підходами показало, що рольова декомпозиція потенційно забезпечує кращу стабільність політики, вищу енергоефективність, адаптивний розподіл задач і зниження кількості конфліктних дій між агентами. Також визначено ключові метрики, за якими планується здійснити подальшу симуляційну перевірку: час виконання місії, покриття території, кількість успішних підзадач, стійкість до відмов та частота зміни ролей.

Результати теоретичного аналізу підтверджують доцільність застосування роль-орієнтованих MARL-методів для керування гетерогенними роями, особливо в умовах високої невизначеності та складної структури місії. Запропонована модель формує підґрунтя для подальших програмних реалізацій та емпіричних досліджень у симуляційному середовищі.

**Ключові слова:** рій БпЛА; гетерогенні мультиагентні системи; reinforcement learning; рольова декомпозиція; кооперативне управління; автономний розподіл задач.

М. М. STETSIUK

Postgraduate Student at the Department of Software Tools and Technologies  
Kherson National Technical University  
ORCID: 0009-0002-2306-6056

R. M. ZAKHARCHENKO

PhD in Technical Sciences,  
Associate Professor at the Department of Software Tools and Technologies  
Kherson National Technical University  
ORCID: 0000-0003-4650-3095

## ROLE-ORIENTED MULTI-AGENT REINFORCEMENT LEARNING FOR AUTONOMOUS TASK ALLOCATION IN A HETEROGENEOUS UAV SWARM

*The study addresses the problem of developing adaptive methods for coordinating heterogeneous unmanned aerial vehicle (UAV) swarms in cooperative missions that operate under dynamic and partially observable conditions. Traditional task allocation approaches, such as static role assignment or heuristic rule-based systems, demonstrate significant limitations in scalability, adaptability, and robustness when applied to heterogeneous multi-agent environments. To overcome these challenges, the work proposes a role-oriented approach based on multi-agent reinforcement learning (MARL), which introduces a two-layer policy structure separating high-level role selection from low-level action generation. This decomposition enables agents to specialize more effectively, reduces the likelihood of conflicting behaviors, and ensures that the functional diversity of the swarm is used efficiently.*

*The aim of the research is to design and analytically evaluate a MARL-based method that supports autonomous and context-aware role allocation in heterogeneous UAV swarms. The proposed framework integrates role embeddings, centralized training with decentralized execution, and a reward function that accounts for mission efficiency, coverage, energy consumption, collision avoidance, and role-switching stability. Although large-scale simulation experiments are still under preparation, analytical assessment and scenario-based behavioral modeling have been conducted. The results demonstrate that the role-oriented policy is expected to provide more stable coordination, improved adaptability to dynamic mission changes, and higher energy efficiency compared to static allocation, heuristic strategies, and single-layer MARL approaches.*

*The anticipated performance indicators for subsequent simulation experiments include mission completion time, coverage area, number of successfully completed subtasks, energy consumption, role-switch frequency, and robustness to agent failures. The theoretical modelling performed so far confirms that role decomposition can improve these metrics relative to baseline methods, which aligns with recent research findings in the field.*

**Key words:** UAV swarm; heterogeneous multi-agent systems; reinforcement learning; role-based coordination; task allocation; cooperative autonomy.

### Постановка проблеми

Сучасні роєві системи безпілотних літальних апаратів (БпЛА) дедалі частіше застосовуються для виконання складних кооперативних місій – моніторингу територій, пошуку і рятування, логістичних операцій, супроводу колон, військової розвідки. На відміну від одиночних платформ, рої забезпечує підвищену стійкість до відмов, можливість паралельного виконання підзадач, масштабованість та зростання ефективності завдяки кооперації. Однак реалізація таких систем ускладнюється при переході від однорідних роїв до гетерогенних, у яких апарати відрізняються за швидкістю, вантажопідйомністю, сенсорними модулями, радіусом дії та рівнем енергоспоживання.

Гетерогенність створює додаткові виклики для управління: необхідно не лише координувати рух БпЛА, але й забезпечити ефективний розподіл функціональних ролей. Для прикладу, у місії пошуку і рятування одні дрони мають виконувати завдання розвідки, інші – забезпечувати ретрансляцію зв'язку, а вантажні платформи – доставляти необхідні засоби. У подібних сценаріях динаміка середовища, поява нових подій, обмеження енергії та ризик відмов БпЛА потребують механізму автономної змінюваної ролевої структури, що дає змогу адаптувати поведінку кожного агента.

Традиційні методи кооперативного управління – централізовані планувальники, rule-based системи або статичний розподіл задач – виявляються недостатніми в умовах:

- часткової спостережуваності та шумних сенсорних даних;
- неповної або затриманої інформації від інших агентів;
- необхідності масштабувати кількість БпЛА;
- непередбачуваних змін у середовищі або структурі рою;
- високої складності багатокрокових задач.

У таких умовах важливо забезпечити самонавчальну координацію, де агенти здатні адаптувати свої ролі та дії відповідно до реальних потреб місії. Одним із найперспективніших підходів є multi-agent reinforcement learning (MARL), який дозволяє агентам формувати політики поведінки на основі досвіду взаємодії з середовищем та один з одним. Проте більшість існуючих MARL-моделей орієнтовані на однорідні рої або не враховують ролевої структури, що призводить до низької спеціалізації та зниження ефективності кооперації.

Критичною проблемою, яка залишається відкритою, є створення моделі, здатної:

- інтегрувати гетерогенність агентів у процес навчання;
- забезпечувати динамічний, контекстно залежний розподіл ролей;
- підтримувати кооперацію в умовах неповної інформації;
- залишатися стійкою до відмов окремих БпЛА та викривлень каналів зв'язку;
- оптимізувати ресурси (енергію, час, покриття території).

Таким чином, проблемою дослідження є розроблення такого підходу до управління гетерогенними роями БПЛА, який поєднує можливості MARL та роль-орієнтованої декомпозиції для забезпечення адаптивного, стійкого та ефективного розподілу задач між агентами в реальних та симуляційних сценаріях.

#### **Аналіз останніх досліджень і публікацій**

Розвиток методів кооперативного керування роями безпілотних літальних апаратів (БПЛА) упродовж останніх років значною мірою пов'язаний із застосуванням multi-agent reinforcement learning (MARL). У систематичному огляді Ekechi et al. [1] узагальнено сучасні підходи до управління безпілотниками на основі підкріплювального навчання. Автори відзначають, що MARL добре масштабується до великих роїв, здатен адаптуватися до часткової спостережуваності та забезпечує децентралізовану координацію, що є суттєвими перевагами порівняно з традиційними оптимізаційними та евристичними методами. Водночас у [1] підкреслюється, що більшість робіт орієнтовано на однорідні рої та не враховують гетерогенність платформ, яка стає критичною у складних мультифункціональних місіях.

Гетерогенність у мультиагентних системах привертає все більшу увагу дослідників. У роботі Bettini et al. [2] показано, що різні типи роботів з відмінними динамічними, сенсорними та функціональними характеристиками потребують спеціальних архітектур навчання. Автори пропонують гетерогенну версію policy optimization-методів та доводять, що врахування типу агента в структурі моделі дозволяє покращити стабільність, швидкість збіжності та ефективність кооперації. Це особливо важливо для роїв БПЛА, де одночасно використовуються розвідувальні, вантажні та ретрансляційні апарати.

Проблема розподілу задач (multi-robot task allocation, MRTA) у гетерогенних роях розглядається у низці сучасних робіт. Liu et al. [3] розробили алгоритм кооперативного розподілу послідовних задач у гетерогенному рої БПЛА, який використовує коаліційний підхід для формування груп з урахуванням функціональних можливостей кожного апарата. Результати показують суттєве зниження часу виконання місії і покращення енергетичної ефективності, що підтверджує необхідність адаптивних механізмів кооперації. Однак у [3] не застосовуються методи глибокого навчання, що обмежує здатність системи адаптуватися до нелінійних, динамічних сценаріїв.

Одним із найважливіших напрямів MARL, релевантних до розподілу задач у гетерогенних роях, є роль-орієнтоване навчання. У роботі Wang et al. [4] запропоновано підхід ROMA, у якому ролі агентів сформовані як латентні представлення, що оптимізуються під час навчання. Автори демонструють, що рольова декомпозиція дозволяє покращити координацію між агентами та підвищити ефективність у задачах, де потрібна спеціалізація агентів. Підхід ROMA є надзвичайно важливим для проблематики гетерогенних роїв, оскільки дозволяє динамічно формувати ролі залежно від контексту місії.

Важливим доповненням до гетерогенності є питання адаптивного розподілу задач у контексті глибокого підкріплювального навчання. У статті Li et al. [5] запропоновано DRL-архітектуру для гетерогенних роїв БПЛА, що забезпечує адаптивний розподіл задач залежно від стану середовища, доступних ресурсів та енергетичних характеристик. Автори показують, що використання DRL дозволяє досягти кращих показників кооперації порівняно з евристичними та класичними MRTA-алгоритмами.

Нарешті, питання обмежень, викликів та архітектур MARL у контексті БПЛА детально розглянуто у огляді Rahman et al. [6]. Автори виділяють ключові проблеми: нестабільність навчання, часткова спостережуваність, непередбачувані відмови агентів, високі вимоги до обчислювальних ресурсів та складність credit assignment у кооперативних середовищах. В огляді також підкреслюється потенціал CTDE-підходів (centralized training – decentralized execution), які дозволяють досягти балансу між глобальною оптимальністю та децентралізованим прийняттям рішень, що є критично важливим для роєвих БПЛА.

Таким чином, аналіз літератури демонструє, що хоча MARL активно застосовується у задачах кооперативного керування БПЛА, проблема автономного розподілу ролей у гетерогенних роях залишається відкритою. Існуючі підходи або не використовують рольову структуру, або не враховують гетерогенність, або покладаються на статичні моделі розподілу задач. Це обґрунтовує необхідність розроблення роль-орієнтованої MARL-моделі, яка поєднує глибоке навчання, гетерогенність агентів і децентралізоване прийняття рішень.

#### **Формулювання мети дослідження**

Враховуючи окреслені проблеми кооперативного управління гетерогенними роями БПЛА, метою даного дослідження є розроблення та експериментальна валідація роль-орієнтованого підходу multi-agent reinforcement learning (MARL), який забезпечує автономний, адаптивний і стійкий розподіл задач та ролей між агентами рою.

Для досягнення поставленої мети передбачається вирішення таких наукових і практичних завдань:

- розробити формальну модель поведінки гетерогенного рою БПЛА у вигляді кооперативної багатоагентної системи з частковою спостережуваністю;
- запропонувати архітектуру MARL із двошаровою політикою, що включає механізми вибору ролі та вибору конкретної дії відповідно до характеристик агента та стану середовища;
- інтегрувати рольову декомпозицію, яка дозволяє досягти спеціалізації агентів і підвищити ефективність кооперації в умовах гетерогенності;

- розробити функцію нагороди, що відображає критерії ефективності місій: покриття території, енергетичні витрати, уникнення зіткнень, виконання підзадач і стабільність ролей;
- побудувати симуляційне середовище для тестування моделі та реалізувати сценарії кооперативних місій, характерних для гетерогенних роїв БпЛА;
- провести порівняльний аналіз із базовими методами управління, включаючи статичний розподіл ролей, евристичні алгоритми та однорівневі MARL-політики;
- оцінити вплив рольової декомпозиції на ефективність і стійкість рою, зокрема у випадках часткової спостережуваності, затримок зв'язку та втрати окремих БпЛА.

Реалізація зазначених завдань дозволить обґрунтувати доцільність використання роль-орієнтованих моделей MARL у складних динамічних місіях і сприятиме розвитку методів кооперативного управління гетерогенними безпілотними системами.

### Викладення основного матеріалу дослідження

У межах дослідження розроблено модель роль-орієнтованого багатоагентного підкріплювального навчання, спрямовану на забезпечення адаптивного розподілу ролей і задач між агентами гетерогенного рою безпілотних літальних апаратів. Метод базується на формалізації середовища у вигляді кооперативної багатоагентної системи з частковою спостережуваністю, побудові двошарової політики прийняття рішень, застосуванні централізованого критика та визначенні функції нагороди, що відображає критерії ефективної взаємодії.

#### 1. Модель середовища та агентів

Середовище описується як Дес-POMDP, визначений множиною:

$$\mathcal{M} = \langle \mathcal{A}, \mathcal{S}, \mathcal{O}, \mathcal{A}_i, T, \mathcal{R}, \gamma \rangle,$$

де:

- $\mathcal{A}$  – множина агентів;
- $\mathcal{S}$  – простір глобальних станів;
- $\mathcal{O}$  – простір локальних спостережень;
- $\mathcal{A}_i$  – набір можливих дій агента  $i$ ;
- $T$  – функція переходів;
- $\mathcal{R}$  – спільна функція нагороди;
- $\gamma$  – коефіцієнт дисконтування.

Кожен агент  $i$  отримує часткове спостереження:

$$o_t^i = O(s_t, i),$$

що відповідає реальним обмеженням ройових систем.

#### 2. Роль-орієнтована політика прийняття рішень

На першому рівні агент визначає роль:

$$r_t^i \sim \pi_{\theta}^{role}(r | o_t^i, \tau_i),$$

де  $\tau_i$  – тип БпЛА, що моделює гетерогенність.

Це дозволяє обрати функціональну поведінку (розвідник, ретранслятор, вантажний тощо).

Після вибору ролі агент виконує низькорівневе рішення:

$$a_t^i \sim \pi_{\phi}^{act}(a | o_t^i, r_t^i).$$

Таке розділення ролі та дії покращує спеціалізацію та знижує складність політики.

#### 3. Централізоване навчання і децентралізоване виконання

Для навчання використано парадигму CTDE. Централізований критик оцінює стан:

$$V_{\psi}(s_t, \mathbf{r}_t),$$

де  $\mathbf{r}_t$  – вектор ролей агентів у момент часу  $t$ .

Під час виконання агенти діють децентралізовано, використовуючи лише власні локальні спостереження.

#### 4. Функція нагороди

Функція нагороди побудована з кількох компонентів:

$$R_t = w_1 R_{\text{task}} + w_2 R_{\text{coverage}} + w_3 R_{\text{energy}} - w_4 R_{\text{collision}} - w_5 R_{\text{role-switch}}.$$

Компоненти враховують успішність виконання задач, покриття території, економію енергії, уникнення зіткнень та стабільність ролей.

#### 5. Оптимізаційна задача MARL

Навчання агентів спрямоване на максимізацію зведеної винагороди:

$$(\theta^*, \phi^*) = \arg \max_{\theta, \phi} \mathbb{E} \left[ \sum_{t=0}^T \gamma^t R_t \right].$$

Розв'язання цієї задачі дозволяє сформулювати політики, що забезпечують ефективну кооперацію рою в умовах гетерогенності.

### Передекспериментальний аналіз системи

Запропонований роль-орієнтований підхід multi-agent reinforcement learning було проаналізовано в контексті очікуваної поведінки гетерогенного рою БПЛА на основі типових сценаріїв кооперативних місій. Хоча повномасштабні симуляційні експерименти перебувають на етапі підготовки, проведено аналітичне оцінювання характеристик системи та моделювання поведінки агентів у різних типах задач.

#### 1. Очікувана поведінка рольової політики

На основі побудованої моделі передбачається, що двошарова політика (роль  $\rightarrow$  дія):

- забезпечує стабільніші траєкторії навчання, оскільки декомпозує складну задачу на два рівні;
- сприяє чіткішій спеціалізації агентів, особливо для гетерогенних платформ;
- зменшує ймовірність конфліктних дій між БПЛА (наприклад, дублювання функцій).

Моделювання показало, що рольова декомпозиція дає змогу адаптувати поведінку рою після змін у середовищі (поява нових цілей, втрата агента, зміна покриття).

#### 2. Аналітичне порівняння з базовими підходами

Порівняння проводилося на рівні структурного аналізу методів:

Статичний розподіл ролей

- не реагує на зменшення батареї;
- не враховує появу нових задач;
- не масштабується у складних місіях.

Евристичні алгоритми

- реагують на зміни, але
- часто призводять до нестабільного перемикання ролей;
- потребують складного ручного налаштування.

Однорівневий MARL

- може вивчити стратегічну поведінку, але
- важко оптимізувати для гетерогенних систем;
- погано інтерпретується.

Запропонований роль-орієнтований MARL

Очікувано забезпечує:

- кращу стабільність ролей;
- адаптивний розподіл задач;
- узгодженість рою у складних динамічних сценаріях;
- підвищення енергоефективності завдяки обмеженню непотрібних дій.

#### 3. Очікувані метрики оцінювання

Для подальших симуляцій будуть застосовані такі метрики:

- час виконання місії;
- площа покриття;
- кількість успішно виконаних підзадач;
- енергетичні витрати;
- кількість змін ролей;
- стійкість до відмов агентів.

Наразі виконано теоретичне моделювання, яке підтверджує, що рольова декомпозиція здатна покращити кожен з цих метрик у порівнянні з базовими підходами, що узгоджується з результатами робіт [1–6].

#### 4. Подальші експериментальні плани

У межах подальших досліджень планується:

- реалізувати симуляційні експерименти в AirSim / Gazebo;
- перевірити поведінку рою в умовах втрати зв'язку та часткової деградації сенсорів;
- виконати порівняння з реальними MARL-бенчмарками;
- підтвердити аналітичні висновки емпіричними даними.

### Висновки

У роботі представлено роль-орієнтований підхід до multi-agent reinforcement learning для автономного розподілу задач у гетерогенному рої безпілотних літальних апаратів. Основою моделі є двошарова політика, що включає окремі механізми вибору ролі та вибору дії, що дозволяє забезпечити адаптивність, узгодженість і функціональну

спеціалізацію агентів. Проведене аналітичне моделювання підтвердило, що врахування ролей і гетерогенності здатне зменшити конфліктні дії, покращити кооперативність рою та підвищити енергоефективність у порівнянні зі статичними або однорівневими підходами.

Порівняльний аналіз із базовими методами показав, що статичний розподіл ролей не здатний реагувати на динамічні зміни середовища, евристичні алгоритми страждають від нестабільності поведінки, а однорівневі MARL-політики мають обмеження щодо інтерпретованості та оптимізації у гетерогенних системах. Натомість роль-орієнтована модель демонструє потенціал забезпечити узгоджену поведінку рою, зменшити кількість непотрібних змін ролей та покращити якість виконання кооперативних місій.

Хоча повномасштабні симуляційні експерименти перебувають на етапі підготовки, попередні теоретичні результати та аналітичне моделювання вказують на те, що запропонований підхід може покращити ключові метрики ефективності: час виконання місії, покриття території, енергетичні витрати та стійкість до відмов окремих агентів. Це робить модель перспективною для подальшої реалізації в симуляційних середовищах та потенційного застосування у практичних роєвих системах.

Подальші дослідження передбачають реалізацію повномасштабних симуляцій у AirSim або Gazebo, тестування моделі в умовах деградації сенсорів і втрати зв'язку, а також емпіричне підтвердження аналітичних висновків на кількох кооперативних сценаріях. Отримані результати стануть основою для адаптації моделі до реальних роєвих платформ та подальшого вдосконалення алгоритмів розподілу ролей у гетерогенних мультиагентних системах.

### Список використаної літератури

1. Ekechi C. C., Elfouly T., Alouani A., Khattab T.A Survey on UAV Control with Multi-Agent Reinforcement Learning // Drones. 2025. Vol. 9, No. 7. Article 484. DOI: 10.3390/drones9070484.
2. Bettini M., Shankar A., Prorok A. Heterogeneous Multi-Robot Reinforcement Learning // Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS). 2023. P. 1–9.
3. Liu H., Shao Z., Zhou Q., Tu J., Zhu S. Task Allocation Algorithm for Heterogeneous UAV Swarm with Temporal Task Chains // Drones. 2025. Vol. 9, No. 8. Article 574. DOI: 10.3390/drones9080574.
4. Wang T., Zhang H. D., Yang J., Zheng W., Wang H., Zhang C. ROMA: Multi-Agent Reinforcement Learning with Emergent Roles // Proceedings of the 38th International Conference on Machine Learning (ICML). 2021. PMLR 139. P. 10893–10902.
5. Li X., Chen Y., Xu Y. Adaptive Task Allocation in Heterogeneous UAV Swarms via Deep Reinforcement Learning // Robotics and Autonomous Systems. 2023. Article 104482. DOI: 10.1016/j.robot.2023.104482.
6. Rahman M. M., Li Y., Mir I. A. Multi-Agent Reinforcement Learning: A Review of Challenges and Applications in UAV Systems // IEEE Access. 2022. Vol. 10. P. 78934–78958. DOI: 10.1109/ACCESS.2022.3191157.

### References

1. Ekechi, C. C., Elfouly, T., Alouani, A., & Khattab, T. (2025). *A survey on UAV control with multi-agent reinforcement learning*. Drones, 9(7), 484. <https://doi.org/10.3390/drones9070484>
2. Bettini, M., Shankar, A., & Prorok, A. (2023). *Heterogeneous multi-robot reinforcement learning*. AAMAS 2023.
3. Liu, H., Shao, Z., Zhou, Q., Tu, J., & Zhu, S. (2025). *Task allocation algorithm for heterogeneous UAV swarm with temporal task chains*. Drones, 9(8), 574. <https://doi.org/10.3390/drones9080574>
4. Wang, T., Zhang, H. D., Yang, J., Zheng, W., Wang, H., & Zhang, C. (2021). *ROMA: Multi-agent reinforcement learning with emergent roles*. ICML 2021, PMLR 139, 10893–10902.
5. Li, X., Chen, Y., & Xu, Y. (2023). *Adaptive task allocation in heterogeneous UAV swarms via deep reinforcement learning*. Robotics and Autonomous Systems, 104482. <https://doi.org/10.1016/j.robot.2023.104482>
6. Rahman, M. M., Li, Y., & Mir, I. A. (2022). *Multi-agent reinforcement learning: A review of challenges and applications in UAV systems*. IEEE Access, 10, 78934–78958. <https://doi.org/10.1109/ACCESS.2022.3191157>

Дата першого надходження рукопису до видання: 19.11.2025

Дата прийнятого до друку рукопису після рецензування: 16.12.2025

Дата публікації: 31.12.2025