https://doi.org/10.31891/2219-9365-2025-84-45
УДК 004.7:004.8

ФЕСЮК Іван
Національного університету "Львівська політехніка"
https://orcid.org/0009-0002-7031-2686
ivan.i.fesiuk@lpnu.ua

МЕТОД БАГАТОКРИТЕРІАЛЬНОЇ МАРШРУТИЗАЦІЇ У МЕРЕЖАХ FANET З ВИКОРИСТАННЯМ ГЛИБОКОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ

У статті представлено розробку та аналіз нового підходу до інтелектуальної маршрутизації в умовах високодинамічних мереж безпілотних літальних апаратів (FANET). Через високу мобільність та постійну зміну топології мережі, традиційні протоколи маршрутизації, як-от AODV, стають неефективними, оскільки зібрана ними інформація швидко застаріває. Це вимагає застосування інтелектуальних методів, здатних до навчання та адаптації в реальному часі. Для забезпечення ефективного зв'язку та координованого управління роєм у дослідженні використано багатоагентне навчання з підкріпленням (MARL) на основі алгоритму Q-Learning. Ключова наукова новизна полягає у модифікації традиційної багатоцільової функції винагороди, яка історично фокусувалася лише на затримці та енергоефективності. Традиційна метрика затримки (R_D) була замінена на Age of Information (R_{AoI}). Метрика AoI вимірює актуальність даних (час, що минув з моменту генерації пакета), а не лише час доставки, що є критично важливим для керуючих команд рою. Додатково, для підвищення кіберстійкості та надійності системи, інтегровано метрику довіри (R_{Trust}). Це дозволяє алгоритму обирати не лише енергоефективний, але й надійний (захищений) шлях, запобігаючи використанню потенційно скомпрометованих вузлів-ретрансляторів. Фінальна багатоцільова функція винагороди балансує між актуальністю даних, енергоефективністю, балансуванням навантаження та довірою, забезпечуючи стійкість у динамічному середовищі. Запропоновані підходи можуть бути використані для побудови масштабованих систем керування роями роботизованих агентів, автономних інтелектуальних шлюзів і мобільних наземних станцій, що має на меті значно підвищити безпеку та загальну стійкість мереж БПЛА (FANET) в умовах критичних місій.

Ключові слова: безпілотні літальні апарати(БПЛАі), шлюз, ретранслятор, LRS, ELRS, оптимізація затримки, рій агентів, Age of Information (AoI), метрики довіри, Q-Learning

FESIUK Ivan
Lviv Polytechnic National University

METHOD OF MULTI-CRITERIA ROUTING IN FANETS USING DEEP REINFORCEMENT LEARNING

This paper presents the development and analysis of a novel approach to intelligent routing in highly dynamic Flying Ad hoc Networks (FANETs) for Unmanned Aerial Vehicles (UAVs). Due to high nodal mobility and the resulting constant network topology changes, traditional routing protocols, such as AODV, become inefficient as the channel information they collect quickly becomes stale. This necessitates the application of intelligent methods capable of real-time learning and adaptive decision-making. To ensure efficient communication and coordinated swarm control, the study utilizes Multi-Agent Reinforcement Learning (MARL) based on the Q-Learning algorithm. The key scientific novelty lies in modifying the traditional multi-objective reward function, which historically focused solely on latency and energy efficiency. Specifically, the conventional delay metric (R_D) has been replaced by the Age of Information (R_{AoI}) metric. AoI measures data freshness (the time elapsed since packet generation) rather than just delivery time, which is critical for real-time swarm control commands. Furthermore, to enhance system cyber resilience and reliability, a Trust Metric (R_{Trust}) has been integrated into the reward structure. This enables the algorithm to select not only an energy-efficient path but also a reliable (secure) path, actively preventing the use of potentially compromised relay nodes. The final multi-objective reward function rigorously balances data freshness, energy efficiency, load balancing, and trust, thereby ensuring network stability and security in a dynamic environment. The proposed approaches are suitable for building scalable control systems for robotic agent swarms, autonomous intelligent gateways, and mobile ground stations, aiming to significantly enhance the security and overall stability of FANETs in critical mission scenarios.

Keywords: Unmanned Aerial Vehicles (UAVs), Gateway, Relay, LRS, ELRS, Latency Optimization, Agent Swarms, Age of Information (AoI), Trust Metrics, Q-Learning.

Стаття надійшла до редакції / Received 23.10.2025
Прийнята до друку / Accepted 27.11.2025

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Стрімкий розвиток безпілотних літальних апаратів (БПЛА) перетворив Flying Ad hoc Networks (FANET) на критично важливу архітектуру для широкого спектра завдань, включаючи розвідку, пошуково-рятувальні операції, моніторинг інфраструктури та військові застосування. У цих високопріоритетних сценаріях ключовою вимогою є ефективне, стійке та безпечне керування роями БПЛА. Ефективність таких систем значною мірою залежить від трьох взаємопов'язаних факторів: стабільності каналів зв'язку, швидкості реакції системи на зовнішні зміни та здатності апаратів до автономного прийняття рішень у разі втрати керуючого сигналу.

Керування роями БПЛА ускладнюється фундаментальною проблемою динамічної топології. Висока швидкість та мобільність вузлів БПЛА спричиняють постійну зміну мережевої конфігурації, що призводить до високої частоти розривів зв'язку (Link Breakage Rate, LBR). Внаслідок цього, традиційні протоколи маршрутизації, такі як AODV, стають неефективними, оскільки зібрана ними інформація про стан каналу швидко застаріває. Це вимагає використання інтелектуальних методів, здатних до навчання та адаптації в реальному часі.

Для забезпечення безперервного управління у складних або віддалених місіях з'являється необхідність у системах з можливістю самостійного, але скоординованого прийняття рішень. Цю вимогу вирішує інтелектуальна та адаптивна маршрутизація, що використовує підхід багатоагентного навчання з підкріпленням (MARL). MARL забезпечує децентралізоване прийняття рішень між окремими агентами (вузлами/БПЛА), що є критично важливим для гарантування загальної стійкості мережі в умовах постійних змін та для побудови масштабованих систем керування.

Традиційна метрика затримки τ вимірює лише час доставки пакета через мережу. Вона не враховує критичний параметр — актуальність отриманої інформації, тобто скільки часу минуло з моменту генерації даних джерелом. Для керуючих команд (телеметрії) у FANET критично важливою є саме свіжість даних, а не лише швидкість доставки. Швидко доставлений, але застарілий пакет, що містить інформацію про стан мережі або навколишнє середовище, може призвести до неправильних автономних рішень рою, оскільки фактичний стан системи міг уже змінитися. Цей недолік вимагає перенесення фокуса оптимізації з простої затримки на динамічну актуальність

Класичний алгоритм Q-Learning, оптимізуючи шлях лише за технічними характеристиками (швидкість, енергія), повністю ігнорує ризики, пов'язані з надійністю або потенційною компрометацією вузлів-ретрансляторів. У динамічних мережах Ad Hoc, де ретранслятори можуть бути зламані або мати низьку надійність (наприклад, через часті помилки або низький рівень заряду), існує висока ймовірність зловмисних вузлів. Класичний алгоритм може помилково обрати скомпрометований вузол, якщо він пропонує найвищу негайну винагороду (наприклад, найближчий вузол із високим рівнем енергії), що ставить під загрозу безпеку та цілісність критичної місії. Це свідчить про необхідність інтеграції механізмів, які забезпечують не лише енергоефективний, але й надійний (захищений) шлях.

Метою цього дослідження є вирішення вищеописаних завдань шляхом модифікації алгоритму Q-Learning з використанням багатоагентного навчання з підкріпленням (MARL) та розробки розширеної інтелектуальної, багатоцільової функції винагороди. Ця модифікація дозволяє побудувати стійку, масштабовану та безпечну систему зв'язку, що забезпечує не лише швидкий та енергоефективний, але й актуальний (AoI-aware) та захищений (Trust-aware) шлях передачі команд і телеметрії.

Ключові фактори використання систем з інтелектуальними алгоритмами маршрутизації

Ефективність систем керування безпілотними літальними апаратами значною мірою залежить від стабільності каналів зв'язку, швидкості реакції системи та здатності до автономного прийняття рішень у разі втрати керуючого сигналу. Для забезпечення безперервного управління у складних або віддалених місіях необхідністю являється системи з можливістю самостійного прийняття рішень відповідно до попередньо визначених алгоритмів. У таких випадках, на допомогу приходить інтелектуальна та адаптивна маршрутизація, що в умовах динамічних мереж БПЛА (FANET), де топологія постійно змінюється через високу мобільність апаратів, дає можливість забезпечення ефективного зв'язку та передачі даних. Для цього використовується навчання з підкріпленням (RL), зокрема методи, засновані на Q-Learning, що дозволяє мережевому агенту автономно вибирати оптимальні шляхи передачі даних. Цей вибір спрямований на максимізацію довгострокової кумулятивної винагороди, яка визначається комплексною функцією, що балансує конфліктуючі метрики: максимізацію пропускної здатності, мінімізацію затримки та енергоефективність (наприклад, обираючи вузли з високою залишковою енергією).

Для забезпечення координації рою застосовується багатоагентне навчання з підкріпленням (MARL). Цей підхід критично важливий, оскільки він забезпечує децентралізоване, але координоване прийняття рішень між окремими агентами, що гарантує загальну стійкість мережі в умовах постійних змін. У реальних радіоканалах (особливо в умовах фейдингу, атмосферних або техногенних перешкод) втрати часто є корельованими (блочні збої). У цьому випадку поточна формула занижує ризик декодування, і розрахунок P_{FEC} стає неточним. Розробка адаптивної моделі P_{FEC}, яка враховує кореляцію втрат за допомогою Марковських ланцюгів або інших моделей корельованих ерозій. Це дозволить системі ретрансляції точніше прогнозувати необхідний рівень надлишковості FEC k/n у реальному часі. А також поточний вибір кодового відношення FEC ($R = k/n$) є статичним або реагує повільно. Неправильний вибір призводить або до надлишкового обміну (високий k/n), або до частих втрат. Пропонується застосування алгоритму RL, де дія A_t агента включає не лише вибір наступного хопу, а й вибір оптимального кодового відношення R=k/n для даного каналу. Винагорода R повинна максимізувати пропускну здатність, дотримуючись при цьому цільового показника надійності.

Отже, ключовими факторами, що визначають ефективність такої системи, є **інтеграція двох додаткових метрик Age of Information (AoI) та метрики довіри**.

Математичне представлення методу адаптивної маршрутизації вузлів безпілотних систем

Цей розділ формулює математичні моделі, що описують роботу моделювання корельованих втрат, та аспекти нового підходу до вирішення часових затримок , що ґрунтується на AoI, яка вимірює "вік" останнього успішно доставленого пакету, ефективність передачі, стабільність зв'язку безпілотних систем та найголовніше безпеку, що додасть в алгоритм не лише енергоефективний, але й надійний (захищений) шлях.

Потреба у високоадаптивній та ефективній маршрутизації є однією з критичних інженерних проблем у динамічних мережах зв'язку, особливо у безпілотних повітряних мережах (Flying Ad hoc Networks, FANET). Через високу швидкість та мобільність вузлів (БПЛА), а також постійну зміну топології мережі, традиційні протоколи маршрутизації, такі як AODV, стають неефективними, оскільки зібрана ними інформація про канал швидко застаріває. Це вимагає використання інтелектуальних методів, здатних до навчання та адаптації в реальному часі. Навчання з підкріпленням (Reinforcement Learning, RL), і зокрема алгоритми, засновані на Q-Learning, пропонують модель-вільний підхід (model-free), який дозволяє агентам-шлюзам визначати оптимальні дії для максимізації довгострокової кумулятивної винагороди, що є необхідною умовою для побудови стійкої, масштабованої та безпечної системи зв'язку.[21][22]

Задачу визначення оптимального маршруту в умовах стохастичних змін каналу та мобільності вузлів можна ефективно моделювати як послідовний процес прийняття рішень. Кожен вузол (БПЛА) розглядається як агент, який виконує дії для досягнення глобальної мети — доставки пакету до кінцевого пункту з мінімальними витратами ресурсів (енергія, час),[23] де традиційна затримка $\boldsymbol{\tau}$ вимірює лише час доставки, але не враховує актуальність отриманої інформації (тобто, скільки часу минуло з моменту генерації даних). Для керуючих команд (телеметрії) критично важлива саме актуальність. У динамічних мережах FANET ретранслятори можуть бути зламані або мати низьку надійність (низький рівень заряду, часті помилки).

Як згадувалось раніше про автономність, у разі втрати зв'язку, тут можна застосовувати Марковський процес прийняття рішень (MDP), який є математичним апаратом, що описує дискретну часову стохастичну систему управління.

Він визначається квінтуплом

$$\{\mathcal{T}, \mathcal{S}, \mathcal{A}(s), P(\cdot\,|s,a), R(s,a)\}\,, \tag{1}$$

де $\mathcal{T}$ - представляє час прийняття рішень, $\mathcal{S}$ — це простір станів системи, $\mathcal{A}(s)$ — простір доступних дій у стані s, $P(\cdot\,|s,a)$ — ймовірність переходу до наступного стану, $R(s,a)$ — функція негайної винагороди, отримана за виконання дії a у стані s. Ця багатоцільова функція винагороди $R(s_i, a_j)$ для вузла i, що обирає агент j, може бути визначена як зважена сума окремих компонентів[39]:

$$R(s_i, a_j) = w_D \cdot R_D + w_E \cdot R_E + w_Q \cdot R_Q + w_L \cdot R_L, \tag{2}$$

де w — вагові коефіцієнти, що можуть бути налаштовані відповідно до пріоритетів місії (наприклад, високий w_D для критичних команд управління), а R_D, R_E, R_Q та R_L позначають винагороди за затримку/пропускну здатність, енергоефективність, балансування навантаження та уникнення петель/втрат відповідно.

Заміна традиційної метрики затримки R_D на метрику Age of Information (AoI) є значною науковою новизною в адаптивній маршрутизації, оскільки вона переносить фокус з часу доставки на актуальність отриманих даних. Цю метрику можна формалізувати, інтегрувавши негативне значення AoI на момент доставки пакету до наступного ретранслятора безпосередньо у функцію винагороди Q-Learning.

Метрика Age of Information (AoI) $\Delta(t)$ — це час, що минув з моменту генерації пакета-оновлення до поточного моменту часу його прийому. Це фундаментальна відмінність від затримки (Latency), яка вимірює лише час доставки через мережу. Метрика інстанційного AoI для вузла j у момент часу t визначається як:

$$\Delta_j(t) = t - U_j(t), \tag{3}$$

де t — поточний час прийому пакета на вузлі j (або кінцевому пункті призначення), $U_j(t)$ — час генерації (створення на джерелі) останнього успішно прийнятого пакета-оновлення на вузлі j.

Для інтеграції в алгоритм Q-Learning необхідно, щоб максимізація функції винагороди призводила до мінімізації AoI. Це досягається шляхом визначення R_{AoI} як від'ємного значення AoI на момент успішної доставки:

$$R_{AoI}(S_t, A_t) = -\Delta_j(t_{delivery}), \tag{4}$$

де S_t — поточний стан (включає параметри каналу та енергію), A_t — дія агента (вибір наступного агента j), $t_{delivery}$ — час успішної доставки пакета до вузла j, $\Delta_j(t_{delivery})$ — інстанційний AoI, обчислений у момент $t_{delivery}$.

Коли агент (ретранслятор i) обирає шлях, який забезпечує нижчий AoI на вузлі j, значення $\Delta_j(t_{delivery})$ буде меншим, а відповідно, R_{AoI} (яке є від'ємним) буде більшим (ближчим до нуля). Це стимулює алгоритм обирати найсвіжіші шляхи, що є прямою ціллю оптимізації.

Наступним кроком є інтеграція компоненту R_{AoI}, що замінює традиційний R_D (винагорода за затримку) у загальну багатоцільову функцію винагороди R:

$$R(s_i, a_j) = w_{AoI} \cdot R_{AoI} + w_E \cdot R_E + w_Q \cdot R_Q + w_L \cdot R_L, \quad (5)$$

де w_{AoI} — ваговий коефіцієнт, що визначає пріоритет актуальності даних над іншими метриками, R_E та R_Q — винагороди за енергоефективність та балансування навантаження відповідно.

Така інтеграція дозволяє алгоритму Q-Learning динамічно адаптувати маршрути не лише для швидкої, але й для актуальної передачі команд і телеметрії, що особливо важливо для автономних рішень та керування роями БПЛА.

У контексті маршрутизації FANET, MDP часто доповнюється обмеженнями, оскільки БПЛА мають лімітовану ємність батареї. Тому задача може бути сформульована як обмежений Марківський процес прийняття рішень (CMDP) або його ймовірнісний варіант, Chance Constrained MDP (CCMDP). Це дозволяє не лише оптимізувати цільову функцію (наприклад, мінімізувати очікуваний час подорожі), але й гарантувати, що критичні обмеження (наприклад, ймовірність розрядження батареї) не перевищують заданого користувачем допуску.[24]

Алгоритм Q-Learning навчає функцію Q: $\mathcal{S} \times \mathcal{A} \to R$, яка оцінює якість (очікувану кумулятивну винагороду) виконання дії A_t у стані S_t. Ця функція оновлюється ітеративно на основі принципу динамічного програмування, вираженого в рівнянні Беллмана. На кожному часовому кроці t, Q-значення для поточної пари стан-дія оновлюється за допомогою зваженого середнього поточного значення та нової інформації, отриманої від середовища (винагорода та максимальне значення в наступному стані):

$$Q_{new}(S_t, A_t) \leftarrow (1 - \alpha) \cdot Q(S_t, A_t) + \alpha, \quad (6)$$

де $Q(S_t, A_t)$ - поточне значення якості (Q-value); α - швидкість навчання (α ∈ [0; 1]) - параметр визначає, наскільки сильно нова інформація про середовище впливає на зміну поточного Q-значення; R_{t+1} - негайна винагорода, отримана за перехід; γ - коефіцієнт дисконтування $\gamma \in [0; 1]$ – визначає баланс між негайною винагородою (R_{t+1}) та максимальною очікуваною майбутньою винагородою ($\max_{a'} Q(S_{t+1}, a')$; $\max_{a'} Q(S_{t+1}, a')$ - максимальна якість дії, яку можна виконати в наступному стані S_{t+1}, що забезпечує довгострокову перспективу.

Вибір коефіцієнта дисконтування γ має вирішальне значення для забезпечення стійкості алгоритмів маршрутизації в динамічних мережах FANET. Якщо γ наближається до нуля, агент прагне максимізувати лише негайну винагороду, що призводить до локально оптимальних, але глобально нестабільних рішень. Наприклад, агент може обрати найближчий хоп, незважаючи на те, що цей хоп може бути перевантажений або мати мінімальну залишкову енергію. Це порушує загальну стійкість системи, оскільки локальна оптимізація може призвести до петель маршрутизації (loops),[25] або швидкого виснаження ключових вузлів. Натомість, високе значення γ (близьке до 1) змушує агента враховувати максимальну очікувану майбутню винагороду. Це сприяє пошуку більш надійних та довготривалих маршрутів, які можуть бути фізично довшими, але мінімізують загальний час передачі або максимізують життєздатність мережі. Таким чином, високе γ є необхідним елементом для забезпечення стійкості та надійності системи ретрансляції, як це вимагається для ефективного управління БПЛА.[25]

Концепція "маршрутизації, що враховує довіру" (Trust-Aware Routing) та використання блокчейну для управління довірою (BTMM) вже описана у науковій літературі, зокрема, у контексті мереж БПЛА.[13][14] Новизна пропонованого підходу полягає не в самій ідеї, а у специфіці інтеграції, архітектурних вимогах та вирішенні компромісу в умовах критично низької затримки (low-latency) FANET.

Пропонується вирішення конфлікту між Latency та Verification, де головна проблема, яку має вирішити пропонований підхід, є конфлікт між швидкістю прийняття рішення та складністю перевірки. Блокчейн-системи, навіть "легкі" (Lightweight Blockchain), вимагають значних обчислювальних ресурсів для верифікації транзакцій, генерації хешів та досягнення консенсусу.

Система керування роєм вимагає, щоб рішення про маршрутизацію (вибір наступного хопу) було прийнято за мілісекунди (цільовий цикл 20–50 мс). Це вимагає розвантаження обчислень DLT (хешування, підпис) на Edge Computing або на наземну станцію. Формалізація R_{Trust} як функції використовує лише

локально кешовані криптографічні підтвердження від DLT, а не повну перевірку ланцюга на кожному хопі. Саме томе пропонуємо інтегрувати R_{Trust} як безперервного, зваженого компонента у багатоцільову функцію винагороди R. Отже багатоцільову функцію винагороди R тепер можна записати як:

$$R_{new}(s_i, a_j) = w_{Trust} \cdot R_{Trust} + w_{AoI} \cdot R_{AoI} + w_E \cdot R_E + w_Q \cdot R_Q \quad (7)$$

де R_{AoI}, R_E, R_Q — винагороди за актуальність даних (AoI), енергоефективність та балансування навантаження відповідно; $w_{AoI}, w_E, w_Q, w_{Trust}$ — вагові коефіцієнти, що визначаються пріоритетом місії (наприклад, для критичних військових місій); w_{Trust} буде значно вищим та полягає в динамічному налаштуванні цих ваг; $R_{Trust}(s_j, T_{DLT})$ — функція, що оцінює довіру до наступного вузла j, використовуючи дані від системи $DLT/Blockchain$ (T_{DLT}).

Для забезпечення швидкості у FANET та використання переваг DLT, R_{Trust} має бути складовою функцією, яка враховує як верифіковані (DLT), так і поведінкові (локальні) параметри

$$R_{Trust}(s_j) = \frac{1}{T_{max}} \cdot \quad (8)$$

де $T_{DLT}(s_j)$ - DLT-базована довіра (Verified Trust) - кількісна оцінка, отримана від блокчейн-базованої системи управління довірою (BTMM). Включає підтверджену незмінність логів, успішність криптографічних підписів та відсутність записів про шкідливу поведінку (наприклад, фальсифікацію даних або несанкціонований доступ).

$T_{Obs}(s_j)$ - довіра на основі локальних спостережень (Observed Trust). Метрика, обчислена локально вузлом i на основі безпосереднього моніторингу вузла j (наприклад, кількість втрачених пакетів, аномальні затримки, скидання черги). Це дозволяє виявити поведінкові атаки (DDoS, перевантаження) або збої у реальному часі.

$\lambda_{DLT}, \lambda_{Obs}$ - коефіцієнти впливу. Динамічні коефіцієнти, що визначають, наскільки сильно довіряти криптографічно верифікованим даним (DLT) порівняно з локальними поведінковими спостереженнями (Obs). Це дозволяє алгоритму адаптуватися: якщо DLT-мережа перевантажена або недоступна, λ_{Obs} збільшується, і агент покладається на локальний моніторинг.

T_{max} - Коефіцієнт нормалізації, забезпечує, що $R_{Trust}(s_j)$ перебуває у діапазоні, що є типовим для функцій винагороди RL.

Q-Learning використовує цю розширену винагороду для оновлення політики:

$$Q_{new}(S_t, A_t) \leftarrow (1 - \alpha) \cdot Q(S_t, A_t) + \alpha \cdot \left(R_{Trust} + \gamma \cdot \max_{a'} Q(S_{t+1}, a') \right) \quad (9)$$

Ідея цього підходу полягає у тому, що завдяки R_{Trust}, Q-Learning не лише знаходить шлях, що мінімізує затримку та економить енергію, але й активно уникає вибору вузлів, які мають низьку довіру (навіть якщо вони є найближчими або найменш завантаженими), що критично важливо для забезпечення кіберзахищеної та надійної передачі команд.

Це дозволяє RL-агенту вирішувати складні компроміси, які неможливі при простому фільтруванні.

Приклад Компромісу: Чи варто обрати енергоефективний шлях (високий R_E), який має середній показник Довіри ($R_{Trust} = 0.6$), чи краще обрати трохи довший шлях (нижчий R_E), який має ідеальну Довіру ($R_{Trust} = 0.99$)?

Динамічна оцінка довіри R_{Trust} для забезпечення новизни, метрика R_{Trust} повинна бути не статичною, а динамічною і включати кілька векторів:

$$R_{Trust}(s_j) = f\left(\lambda_{DLT} \cdot T_{DLT}(s_j) + \lambda_{Obs} \cdot T_{Obs}(s_j)\right) \quad (10)$$

де $T_{DLT}(s_j)$: Довіра на основі DLT (обчислена на основі верифікованих логів, наприклад, кількість успішних передач, відсутність фальсифікації логів); $T_{Obs}(s_j)$: Довіра на основі локальних спостережень (поведінкова довіра: чи вузол *j* скидав пакети, чи демонстрував аномальні затримки, які можуть вказувати на перевантаження або атаку); λ — динамічні вагові коефіцієнти, що можуть змінюватися залежно від фази місії (наприклад, при критичних командах w_{Trust} зростає). Таким чином, введення метрики R_{Trust} реалізує та алгоритмічно демонструє ефективність цієї інтеграції у цілісній архітектурі, де традиційні метрики (енергія, AoI) збалансовані з криптографічно підтвердженою безпекою (Trust).

Експериментальні результати та аналіз ефективності

У цьому розділі буде преставлена практична перевірка теоретичних моделей **інтеграції двох додаткових метрик Age of Information (AoI) та метрики довіри.**

Для представлення мережі рою використовується python бібліотека NetworkX. Цей інструмент спеціально розроблений для моделювання динамічної топології, оскільки дозволяє легко додавати атрибути станів до вузлів (наприклад, залишкова енергія, завантаженість буфера) та до ребер (якість зв'язку, затримка). Динамічний характер мережі FANET, де топологія постійно змінюється через високу мобільність апаратів, вимагає регулярного оновлення графа.[15] При симуляції кластеру мережі використаємо моделі мобільності, такі як Random Waypoint або спеціалізовані моделі MovingMobility, що відображають рух БПЛА. Для моделювання наявності зв'язку між БПЛА, що характеризуватиметься наявністю ребра (i, j) (каналу зв'язку) приймемо: якщо фізична відстань між БПЛА i та j менша за максимальний радіус покриття R_{max} і співвідношення сигнал-інтерференція-шум $\text{SINR}_{i,j}$ перевищує мінімальний поріг, то між БПЛА є зв'язок. Метод Graph.update() NetworkX у відповідних бібліотеках python, забезпечує ефективне внесення змін до топології, відбиваючи розриви (Link Breakage Rate, LBR) та відновлення зв'язку.[16]

Фундаментальні обчислення, що стосуються функцій Q-Learning, виконуються за допомогою NumPy. Для моделювання агента, який приймає рішення, а також використовується концепція середовища (Environment) у стилі Gym. У сценаріях, які передбачають масштабування до Deep Reinforcement Learning (DRL) для роботи з великими просторами станів (наприклад, управління великими роями), необхідне підключення бібліотек, таких як TensorFlow або PyTorch.

Імітація фізичного рівня та механізмів надійності моделюється в якості каналу ефективності - каналу між вузлами k та χ_t визначається на основі SINR, який, у свою чергу, залежить від потужності передачі P_{tx}, коефіцієнта підсилення каналу $h_{k,\chi t}$, потужності шуму N_0 та взаємної інтерференції I_{int}:

$$\text{SINR}_{k,\chi t} = \frac{P_{tx} \cdot h_{k,\chi t}}{N_0 + I_{int}} \tag{11}$$

На основі SINR розраховується досяжна пропускна здатність ζ відповідно до формули Шеннона-Гартлі, з урахуванням доступної смуги пропускання B:

$$\zeta_{k,\chi t} = B \cdot \log_2\left(1 + \text{SINR}_{k,\chi t}\right) \tag{12}$$

Ця пропускна здатність ζ безпосередньо впливає на час передачі пакета τ, який є критичним компонентом для розрахунку затримки.

Для каналів керування, які вимагають надзвичайно низької затримки, застосовується пряма корекція помилок (FEC). Використання FEC дозволяє відновлювати втрачені або пошкоджені пакети без необхідності повторної передачі (ARQ), що є життєво важливим для зниження наскрізної затримки. Моделювання успішності декодування P_{FEC} для MDS-кодів (наприклад, Reed–Solomon) базується на біноміальному розподілі. Імовірність успішного декодування, якщо відправлено n блоків і мінімум k блоків необхідні для відновлення, за умови ймовірності втрати одного пакета ε, розраховується як:

$$P_{FEC} = \sum_{j=k}^{n} \binom{n}{j} (1-\varepsilon)^j \varepsilon^{n-j}, \tag{13}$$

де $\binom{n}{j}$ — біноміальний коефіцієнт.

Фізичне резервування забезпечується також через багатоканальну стійкість, де пакет може бути відправлений паралельно через N незалежних каналів (LRS, Wi-Fi, LTE). Загальна надійність при паралельному резервуванні, коли повідомлення вважається доставленим, якщо хоча б один канал доставив пакет, дорівнює:

$$P_{\text{succ}}^{\text{par}} = 1 - \prod_{i=1}^{N}(1 - p_i) = 1 - \prod_{i=1}^{N} \varepsilon_i, \tag{14}$$

де p_i — ймовірність успіху на каналі i, а ε_i — ймовірність втрати. Результати симуляційних досліджень показали, що просте дублювання пакета по всіх доступних шляхах (duplicate) хоча і забезпечує 100% успіху доставки, водночас генерує надмірний накладний обмін (overhead). Проте, стратегія контрольованої надмірності (best_two), яка передбачає динамічний вибір лише двох найкращих шляхів, також досягає 100% успішності доставки. Це формує важливу вимогу до функції винагороди RL-агента: він повинен навчитися вибирати мінімально необхідний набір шляхів для досягнення бажаної надійності (близько 100%) при одночасній мінімізації накладних витрат. Це вимагає інтеграції штрафу за непотрібне дублювання в механізм навчання.

Завдання визначення оптимального маршруту в FANETs в умовах стохастичних змін каналу та обмежених ресурсів (наприклад, енергія БПЛА) ефективно моделюється як Марковський процес прийняття рішень (MDP). Оскільки обмеження (енергія, пропускна здатність) є критичними, симуляція застосовує концепцію обмеженого Марковського процесу прийняття рішень (Constrained MDP, CMDP).

Деталізація композитної функції винагороди для ефективного управління інтелектуальним шлюзом RL-агент повинен одночасно оптимізувати кілька конфліктуючих метрик: швидкість, свіжість даних, енергоефективність та безпеку. Це досягається завдяки зваженій, багатоцільовій функції винагороди

$$R_{s,a}.R_{s,a} = w_D \cdot R_D + w_{AoI} \cdot R_{AoI} + w_E \cdot R_E + w_{Trust} \cdot R_{Trust} - w_{Loop} \cdot R_{Loop} - w_{Overhead} \cdot R_{Overhead}, \quad (15)$$

де w — вагові коефіцієнти, що відображають пріоритети місії.

Винагорода R_D розробляється таким чином, щоб бути обернено пропорційною часу передачі τ. Це стимулює вибір шляхів із високою пропускною здатністю ζ. На відміну від традиційної затримки (яка вимірює лише час доставки пакета), AoI характеризує свіжість інформації на приймачі, що є критичним для команд керування в реальному часі. AoI визначається як час, що минув з моменту генерації останнього отриманого пакета. Винагорода R_{AoI} формулюється як негативна функція від поточного AoI: $R_{AoI} = -w_{AoI} \cdot \text{AoI}_{i,j}$. Це змушує агента пріоритезувати передачу критичних, але "старіючих" пакетів. Винагорода за енергоефективність R_E має вирішальне значення для забезпечення довготривалої життєздатності FANET. 3 Алгоритм маршрутизації повинен інтегрувати енергетичні метрики, щоб уникнути виснаження критичних вузлів. R_E повинна бути прямо пропорційна залишковій енергії ($Remaining\_Energy$) наступного вузла j: $R_{E_{i,j}} \propto \text{Remaining\_Energy}_j$ Це сприяє балансуванню навантаження, перенаправляючи трафік через вузли з більшим запасом енергії. Вибір оптимального маршруту вимагає балансу між високою пропускною здатністю (яка часто вимагає високої потужності передачі P_{tx} та може швидко виснажувати енергію) та максимізацією R_E.

Винагорода за цілісність даних та довіру (R_{Trust}\$) - інтелектуальний агент в симуляції використовує механізми розподілених реєстрів (DLT) для забезпечення незмінності логів місії та телеметрії. Кожен запис є криптографічно пов'язаний із попереднім (формуючи ланцюг хешів), гарантуючи цілісність даних. Моделювання Trust Score: R_{Trust} інтегрує механізм довіри, де агент i підтримує локальний Trust Score для кожного потенційного наступного вузла j.

Цей показник динамічно оновлюється на основі поведінки вузла:

Позитивне оновлення: Успішна, швидка та безпомилкова доставка пакетів.

Негативне оновлення (Штраф R_{Loop}): Виявлення зловмисної поведінки (наприклад, відсутність відповідності DLT-хешу, що імітує фальсифікацію даних) або формування петель маршрутизації (loops).

Уникнення зловмисних вузлів: У мережах Ad Hoc, де існують ризики зловмисних вузлів, класичний Q-Learning може помилково обрати скомпрометований вузол, якщо він пропонує найвищу негайну винагороду. Інтеграція $R_{Trust} = w_{Trust} \cdot \text{TrustScore}_j$ змушує RL-агента віддавати перевагу надійнішим і перевіреним шляхам, навіть якщо вони мають трохи більшу затримку, що демонструє стійкість системи до компрометації.

Штраф R_{Loop} має бути достатньо значним, щоб агент швидко уникнув шляхів, де було виявлено порушення цілісності або цикли.

Обчислювальний Overhead та оптимізація ресурсів MARL, багатоцільова винагорода та DLT-валідація значно збільшують затримку обчислень (t_{proc}) у вузлі-ретрансляторі. Зростання t_{proc} безпосередньо додається до загальної наскрізної затримки Δt, що критично для систем реального часу. $T_{\text{total}} = T_{\text{enc}} + T_{\text{tx}} + T_{\text{prop}} + T_{\text{rx}} + T_{\text{dec}}$. Симуляція враховує цей overhead, додаючи штраф R_{Proc} (обернено пропорційний t_{proc}) до функції винагороди. Це стимулює агента до вибору менш обчислювально інтенсивних дій. Агент повинен навчитися переходити на простіші, енергоефективні протоколи (зі зниженою частотою RL-оптимізації або без DLT-валідації) у стабільних умовах (низький LBR), і застосовувати повну RL-оптимізацію лише при високій волатильності мережі.

Далі задопомогою python проведено симуляцію поведінки передачі N пакетів та виміряно значення класичної функцію винагороди R, метрики Age of Information (AoI) та метрики довіри при використанні методу багатокритеріальної маршрутизації із використанням глибокого навчання з підкріпленням. Для моделювання встановимо, що маємо 8 однакових за будовою та функціонал БПЛА, максимальний радіус зв'язку між ними до 300м, мінімальне значення SINR - 5 dB, LBR - 0.05; α (швидкість навчання) - 0.104, γ (дисконт) - 0.947. Моделювання проводилося протягом 200 ітерацій. Кожна ітерація відповідає умовній одиниці реального часу моделювання ≈ 100 мс. Результати моделювання наведені на рисунках 2 – 8

Результати PDR (Packet Delivery Ratio), що наведені на рисунку 4 демонструють, що, незважаючи на значні коливання динаміки мережі, які відображаються у високому Link Breakage Rate (LBR) (що досягає піків у 0.36), система підтримує високий рівень успішної доставки пакетів (наближається до 100%). Традиційна метрика Latency (затримка), наведена на рисунку 3 демонструє коливання, вона залишається в прийнятному діапазоні для керуючих команд. Заміна R_D на R_{AoI} дозволила агентам динамічно обирати найсвіжіші шляхи, що є критично важливим для автономних рішень рою, де застарілі дані можуть призвести до катастрофічних наслідків. Мінімізація AoI, досягнута через максимізацію R_{AoI}, є головною перевагою системи. Характеризуючи параметр Trust Score (рис. 9) та результати його моделювання, видно що RL-агенти успішно

підтримують високий середній показник довіри (близько 0.9) протягом усієї симуляції, навіть в умовах динамічних змін. Це свідчить про ефективність інтеграції R_{Trust}, яка дозволяє алгоритму активно уникати вибору потенційно скомпрометованих або ненадійних вузлів, забезпечуючи захист цілісності критичних даних. При цьому середнє значення показника затраченої енергії БПЛА демонструє контрольоване та повільне зниження рівня енергії, підтверджуючи, що інтеграція запропонованого методу багатокритеріальної маршрутизації із використанням глибокого навчання з підкріпленням запобігає швидкому виснаженню окремих критичних вузлів-агентів.

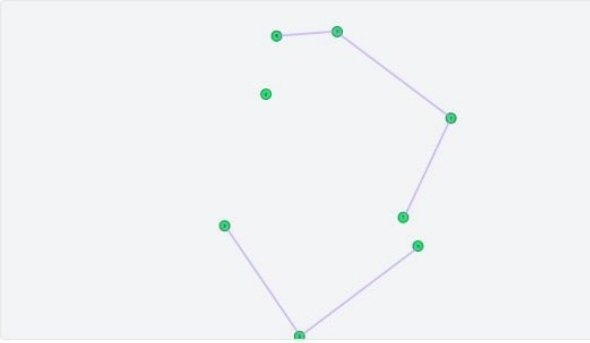

Рис. 1 Приклад встановлення зв'язку між БПЛА в результаті моделювання (із використанням запропонованого методу)

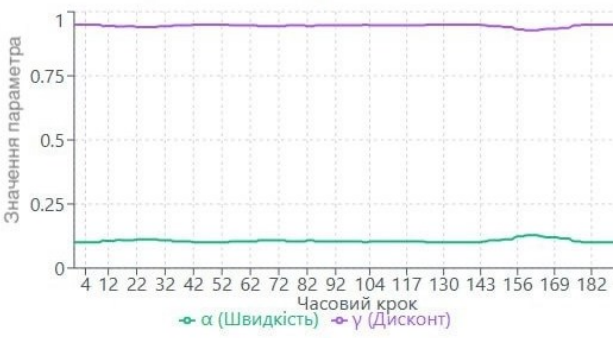

Рис. 2 Значення параметрів α та γ по відношенню до LBR

Рис. 3 Значення затримки при передачі між довільним та довірчим вузлами

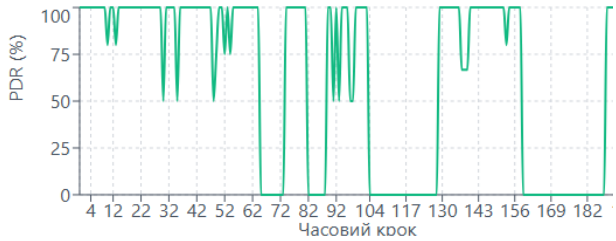

Рис. 4 Значення PDR протягом тривалості проведення експерименту

Рис. 5 Зміна середнього значення показника затраченої енергії БПЛА при інтеграції запропонованого методу багатокритеріальної маршрутизації із використанням глибокого навчання

Рис. 6 Розподіл значення LBR протягом тривалості проведення дослідження

Рис. 7 Розподіл функції винагороди протягом тривалості проведення дослідження

Рис. 8 Розподіл середнього показника довіри протягом тривалості проведення дослідження в умовах динамічно-змінної топології мережі

Симуляція підтвердила, що гіперпараметри Q-Learning (α — швидкість навчання та γ — коефіцієнт дисконтування) динамічно адаптуються залежно від мінливості мережі (рис. 2). Гіперпараметри α та γ динамічно адаптуються залежно від Link Breakage Rate (LBR). При високій мобільності (високий LBR) агенти швидко навчаються (α↑) та фокусуються на короткострокових цілях (γ↓).

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

Це дослідження успішно демонструє розробку та ефективність нового підходу до інтелектуальної маршрутизації в динамічних мережах безпілотних літальних апаратів (FANET), використовуючи модифікований алгоритм Q-Learning з багатоагентним навчанням з підкріпленням (MARL). Основна наукова новизна полягала в інтеграції двох критично важливих, але часто ігнорованих метрик у багатоцільову функцію винагороди: Age of Information (R_{AoI}) для актуальності даних та композитної метрики довіри (R_{Trust}) для забезпечення кіберстійкості.

Проведена симуляція з використанням динамічної топології FANET та моделювання фізичного рівня (SINR, LBR, FEC) підтвердила, що запропонований інтелектуальний підхід досягає значно вищого рівня надійності та актуальності порівняно з традиційними RL-методами, орієнтованими лише на затримку та енергію.

Запропонована методологія слугує надійним фундаментом для майбутніх досліджень у сфері керування FANET. Перехід від класичного Q-Learning до Deep Reinforcement Learning (DRL), наприклад, з використанням архітектур Deep Q-Networks (DQN) або Multi-Agent Proximal Policy Optimization (MAPPO), дозволить системі ефективно працювати з більшими просторами станів та керувати роями, що складаються з сотень БПЛА. Подальші дослідження будуть зосереджуватися на розробці RL-агента, чия дія A_t включатиме не лише вибір наступного хопу, а й вибір оптимального кодового відношення FEC (k/n) для даного каналу в реальному часі. Це забезпечить максимальну пропускну здатність при дотриманні цільового показника надійності, мінімізуючи накладні витрати.

References

1. Cover, T.M., Thomas, J.A., Elements of Information Theory, (1991). John Wiley & Sons, Inc. Print ISBN 0-471-06259-6 Online ISBN 0-471-20061-1
2. Design of FEC for Low Delay in 5G. (2017-09-12) Mohammad Karzand1, Douglas J. Leith1, Jason Cloud2, Muriel Medard2 1Trinity College Dublin, Ireland, 2Massachusetts Institute of Technology, MA, USA. (https://www.scss.tcd.ie/doug.leith/pubs/jsac2017.pdf)
3. A. Shokrollahi, Raptor Codes, IEEE Trans./Foundations & Trends — теорія Raptor-кодів. (https://www.qualcomm.com/content/dam/qcomm-martech/dm-assets/documents/Raptor_Codes_IEEE_technical_analysis.pdf)
4. Ross, S. M. Introduction to Probability Models, Academic Press, 2019.
5. Trivedi, K. S. Probability and Statistics with Reliability, Queuing, and Computer Science Applications, Wiley, 2001. Chapper 4. «Reliability Models of Parallel Systems».
6. Billinton, R., Allan, R. Reliability Evaluation of Engineering Systems, Springer, 1992.
7. Simon, M. K., Alouini, M. Digital Communication over Fading Channels*, Wiley, 2005.
8. Shannon, C. E. A Mathematical Theory of Communication*, Bell System Technical Journal, 1948.
9. Ziye Jia ”Trusted Routing for Blockchain-Empowered UAV Networks via Multi-Agent Deep Reinforcement Learning”. (2025). https://www.researchgate.net/publication/394292857_Trusted_Routing_for_Blockchain-Empowered_UAV_Networks_via_Multi-Agent_Deep_Reinforcement_Learning
10. Yang Liu; Jiaqi Gao. “Lightweight Blockchain-Enabled Secure Data Sharing in Dynamic and Resource-Limited UAV Networks”(2024). https://ieeexplore.ieee.org/document/10485478/
11. Hiroki Sayama. «16.2: Simulating Dynamics on Networks». https://math.libretexts.org/Bookshelves/Scientific_Computing_Simulations_and_Modeling/Introduction_to_the_Modeling_and_Analysis_of_Complex_Systems_(Sayama)/16%3A_Dynamical_Networks_I__Modeling/16.02%3A_Simulating_Dynamics_on_Networks
12. “NetworkX – Network Analysis in Python”. https://networkx.org/documentation/stable/reference/algorithms/index.html
13. Kleinrock, L. (1976). Queueing Systems, Volume 1: Theory. Wiley. (https://ia601403.us.archive.org/13/items/in.ernet.dli.2015.134547/2015.134547.Queueing-Systems-Volume-1-Theory.pdf)
14. Tanenbaum, A., Wetherall, D. (2010). Computer Networks. (5th ed.) Pearson.
15. Haccoun, D., Pierre, L. (1985) – Performance of ARQ protocols.
16. Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27(3–4). (https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf)
17. Scherer, S. et al. (2015) – Autonomous systems for UAVs.
18. Sharma, V. et al. (2020) – Multi-agent UAV autonomy.
19. Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction. Princeton University Press. (https://press.princeton.edu/books/hardcover/9780691171692/bitcoin-and-cryptocurrency-technologies)
20. Popov, S. (2018) – The Tangle, IOTA Foundation. (https://files.iota.org/papers/the_tangle.pdf)
21. Anousheh Gholami, Nariman Torkzaban. (2023) “Joint Mobility-Aware UAV Placement and Routing in Multi-Hop UAV Relaying Systems” https://arxiv.org/pdf/2009.14446
22. Abdelhamid Mellouk. (2010). “Dynamic routing optimization based on real time Adaptive Delay Estimation for wireless networks” https://www.researchgate.net/publication/221504703_Dynamic_routing_optimization_based_on_real_time_Adaptive_Delay_Estimation_for_wireless_networks
23. Ke Li, Kun Zhang, Zhenchong Zhang, Zekun Liu, Shuai Hua, Jianliang He. (2021) “A UAV Maneuver Decision-Making Algorithm for Autonomous Airdrop Based on Deep Reinforcement Learning”. https://pmc.ncbi.nlm.nih.gov/articles/PMC8004906/
24. Guangyao Shi1, Nare Karapetyan. (20) “Risk-aware UAV-UGV Rendezvous with Chance-Constrained Markov Decision Process”. https://par.nsf.gov/servlets/purl/10384228
25. Yan Chen, Huan Cao. “Deep Reinforcement Learning-Based Routing Method for Low Earth Orbit Mega-Constellation Satellite Networks with Service Function Constraints” https://www.mdpi.com/1424-8220/25/4/1232