

Ю. Л. ВЛАДАРЧИК, аспірант;

ORCID: 0009-0008-5271-6285

К. С. НЕСТЕРЕНКО, д-р техн. наук, професор,

ORCID: 0000-0001-7672-7386

Державний університет інформаційно-комунікаційних технологій, Київ

ДЕЦЕНТРАЛІЗАЦІЯ БАГАТОКРИТЕРІАЛЬНОЇ ОПТИМІЗАЦІЇ МАРШРУТІВ SRv6 НА ОСНОВІ ІМПЛІЦИТНОЇ СТИГМЕРГІЇ: ТЕОРЕТИЧНА ОСНОВА ТА ПОЗИЦІОНУВАННЯ СЕРЕД СУЧАСНИХ DRL-ПІДХОДІВ

У статті розглядається задача багатокритеріальної оптимізації (БО) маршрутів у програмованих мережах SRv6 за чотирма QoS-критеріями (затримка, пропускна здатність, втрати пакетів, джитер). Існуючі методи на основі глибокого навчання з підкріпленням (DRL) — AROM-DRL (Tayeh, 2022), MDQ (Zhang та ін., 2024), QR-MO (arXiv:2503.18122, 2025), WA-SRTE (IEEE ICC, 2024) та подібні — розв'язують БО в централізованому контролері SDN, де єдиний арбітр визначає вагову згортку критеріїв або обчислює Парето-множину. Запропоновано принципово іншу архітектуру: БО розв'язується незалежно кожним з автономних AI-агентів на граничних маршрутизаторах, а узгодженість їх рішень у масштабах мережі досягається через імпліцитну стигмергію — спільне зчитування синхронізованої бази даних стану каналів IS-IS (LSDB) як стигмергічного середовища. Стаття формалізує модель координації за Heylighen (2016) у застосуванні до IGP LSDB як медіуму, обґрунтовує методологічну роль обмежень відкритого стека (FRRouting + IS-IS flat area + SRv6 BE + Linux LWT) як стратегії «найгіршого випадку» та пропонує п'ятивісьне порівняння з основними сучасними роботами. Доведено, що тільки запропонована архітектура задовольняє одночасно всі п'ять критеріїв: ≥ 3 QoS-метрики, edge-inference, GNN-узагальнення, відсутність контролера, eBPF-телеметрія для функції винагороди. Емпірична перевірка допустимості координації спирається на експериментальну процедуру з п'яти послідовних фаз (єдине переключення в кожному напрямку, TRIGGER 409 мс, WITHDRAW 825 мс, нуль хибних спрацювань у трьох baseline-фазах). Реакція агента переважно структурно зумовлена (мінімум гістерезису $K\text{-of-}N$ + інтервал зонду + actuation), не вартістю AI — GNN-виведення ONNX INT8 складає субмілісекундний час (менше 0,1 % часу реакції). Симетричний бар'єр WITHDRAW реалізовано без жорсткого порогу через узгодження вхідного представлення з розподілом тренувальних ознак (training-distribution-matched feature wiring) — та сама нейромережа на 40 мс SLA-границі керує переключеннями в обох напрямках.

Ключові слова: багатокритеріальна оптимізація, інформаційна система, інформаційна технологія, машинне навчання, маршрутизація SRv6, імпліцитна стигмергія, децентралізація, GNN, PPO, eBPF.

Вступ

Стрімке зростання різноманітності трафіку (real-time AR/VR, голос, машинне навчання, IoT-телеметрія) поставило перед інфокомунікаційними мережами вимогу одночасної оптимізації за декількома взаємно конфліктними критеріями якості обслуговування: затримкою (latency), пропускною здатністю (throughput), часткою втрат (packet loss) і джитером. Саме ця задача — багатокритеріальна оптимізація маршрутів (БО, multi-criteria optimization, MCO) — становить центральне наукове питання цього дослідження.

Історично БО-маршрутизація розв'язувалася класичними алгоритмами з обчислювальною складністю NP-важкого або NP-повного класу (NP — недетерміновано-поліноміальний

час): NSGA-II, LARAC, multi-constrained Dijkstra. Поява глибокого навчання з підкріпленням (DRL) дала змогу замінити явне обчислення Парето-фронту навченим актором, який обирає шлях за миттєвим станом мережі.

Постановка проблеми та задачі

Переважає більшість сучасних DRL-методів БО-маршрутизації — MDQ Zhang та ін. [1], QR-MO [2], WA-SRTE [3] та подібні — побудовані за централізованою архітектурою: єдиний DRL-агент, розташований у контролері програмно-визначуваної мережі (SDN) або оркестраторі 5G-MEC (Multi-access Edge Computing), отримує телеметрію від усіх вузлів мережі, обчислює рішення і програмує всю мережу. Така архітектура успадковує всі відомі недоліки централізації: точку єдиної відмови, мережеву затримку до контролера, обмеження масштабованості, чутливість до partition-сценаріїв.

Окрему гілку становлять методи топологічної генералізації на основі графових нейронних мереж — GROM Xu та ін. [4], DRL-GNN Almasan-Suárez-Varela [5]. Вони забезпечують роботу однієї моделі на довільних топологіях без перенавчання, проте оптимізують єдину агреговану цільову функцію (зазвичай використання каналів) і також залишаються централізованими.

Перетин трьох властивостей — БО + децентралізація + топологічна узагальнюваність — залишається незаповненою нішею у сучасній літературі. Ця стаття обґрунтовує теоретичну основу одного з можливих заповнень цієї ніші: децентрализованого БО-агента на граничному маршрутизаторі SRv6 (Segment Routing over IPv6), де узгодженість рішень досягається не через явні протоколи координації, а через імпліцитну стигмергію — координацію через спільне середовище.

Аналіз останніх досліджень і публікацій

Сучасну літературу за тематикою цієї статті можна структурувати у три гілки.

Гілка 1. DRL для багатокритеріальної маршрутизації. AROM-DRL Tayeh [1] використовує функцію винагороди на чотирьох критеріях якості обслуговування (QoS), ідентичних застосованим у даному дослідженні (latency, throughput, loss, jitter); архітектура суто SDN-центрована. MDQ Zhang та ін. [2] поєднує QoS- і congestion-aware DRL для multi-path SDN. QR-MO у 5G-MEC-середовищах [3] використовує Q-Routing з явними невідоміями мітками для обчислення Парето-множини. Hierarchical DRL з адаптивно ваговими Парето-винагородами Liu та ін. [8] демонструє узагальнювані принципи БО-винагороди. Preference-Controllable RL (Reinforcement Learning) Chen та ін. [9] (ICML 2025) пропонує meta-policy, що приймає на вхід вектор переваг користувача — найновітніший підхід до тонкого керування Парето-фронтом. WA-SRTE Wang та ін. [4] застосовує DRL до часткового розгортання SRv6, оптимізуючи OSPF-ваги, розташування SRv6-вузлів і шляхи трафіку одночасно — однак single-objective. Усі перелічені роботи припускають єдину точку прийняття рішення про вагову згортку БО-критеріїв.

Гілка 2. Графова нейронна мережа (GNN) + DRL для топологічно-узагальнюваної маршрутизації. Засновницька робота Almasan-Suárez-Varela [7] із групи Knowledge-Defined Networking (BNN-UPC) вперше показала, що GNN-енкодер дозволяє DRL-маршрутизатору працювати на топологіях, не бачених під час навчання. GROM Xu та ін. [6] (Journal of Network and Computer Applications, 2024) розширює цей підхід, демонструючи генералізацію на 5 реальних топологіях без перенавчання. SRv6 + Zero-Trust + GCN (Graph Convolutional Network) для слайсингу Chen та ін. [10] застосовує GCN до SRv6-сценарію, але до задачі слайсингу, а не БО-маршрутизації. Жодна з цих робіт не комбінує GNN-узагальнюваність з ≥ 3 -критеріальною функцією винагороди.

Гілка 3. Децентралізовані RL-агенти у мережевому середовищі. MARL (Multi-Agent Reinforcement Learning) Local Routing Strategy для SR-мереж (Segment Routing) Zhou та ін. [5] — найближчий аналог за принципом децентралізації: PPO-агенти (Proximal Policy Optimization) розгорнуті локально, оптимізують end-to-end-затримку. Single-objective. Жодна з відомих авторам робіт цієї гілки не комбінує децентрализованість виконання з ≥ 3 -критеріальною функцією винагороди та з імпліцитним стигмергічним механізмом координації через

існуючий протокол маршрутизації — поєднання, що формує концептуальну ноту даної роботи.

Окрім перелічених, варто згадати ширший індустріальний контекст: Microsoft [11] використовує SRv6 у backend-мережах кластерів штучного інтелекту (AI) (централізоване управління); Arrcus AINF [12] забезпечує AI-aware steering трафіку на основі централізованої політики; Cisco [13], HPE [14], Ericsson [15], Nokia [16] та Huawei у 2024–2025 рр. оголосили амбіцію переходу до автономних мереж рівня L5 з «agent AI», однак конкретних реалізацій БО на edge-маршрутизаторах у відкритих публікаціях поки не наводять.

Підсумовуючи: жодна з відомих авторам робіт не реалізує одночасно (i) БО за ≥ 3 QoS-критеріями, (ii) edge-inference на самому маршрутизаторі, (iii) GNN-узагальнюваність, (iv) відсутність централізованого контролера, (v) використання kernel-eBPF як джерела телеметрії для функції винагороди.

Мета і задачі дослідження

Мета статті — обґрунтувати теоретичну допустимість стигмергічної координації незалежних БО-агентів у мережах SRv6, формалізувати IS-IS (Intermediate System to Intermediate System) LSDB (Link-State Database) як стигмергічне середовище та позиціонувати запропонований підхід серед сучасних DRL-методів БО-маршрутизації.

Основні задачі:

- сформулювати БО-задачу маршрутизації у термінах функції винагороди агента навчання з підкріпленням;
- здійснити критичний огляд сучасних DRL-методів БО-маршрутизації та виокремити їх обмеження щодо децентралізації;
- формалізувати модель імпліцитної стигмергії за Heylighen у застосуванні до IS-IS LSDB;
- обґрунтувати методологічну роль вибору відкритого програмного стека як стратегії валідації «найгіршого випадку»;
- здійснити п'ятивісьне порівняння з основними сучасними роботами та виявити унікальну позицію запропонованого підходу.

Результати дослідження

1. Формалізація БО-задачі маршрутизації як RL-задачі

Розглянемо мережу як орієнтований граф $G = (V, E)$, де V — множина маршрутизаторів, E — множина каналів зв'язку. Для кожного каналу $e \in E$ визначено вектор миттєвих QoS-показників $q_e(t) = [l_e, b_e, p_e, j_e]$ (затримка, пропускна здатність, частка втрат, джитер). Для пари «джерело-призначення» (s, d) та множини $\mathcal{P}_{sd}$ можливих шляхів задача БО формулюється як:

$$\pi^*(s, d, t) = \arg \max_{\pi \in \mathcal{P}_{sd}} R(\pi, t), \quad (1)$$

де R — скалярна функція винагороди, що агрегує QoS-критерії шляху. Класичні форми R :

$$R_{ws}(\pi) = -w_1 L(\pi) - w_2 P(\pi) - w_3 J(\pi) + w_4 B(\pi) \quad (\text{weighted sum}), \quad (2)$$

$$R_{SLA}(\pi) = +1 \text{ якщо } \pi \models \text{SLA}, \quad -5 \text{ інакше} \quad (\text{SLA-based}). \quad (3)$$

Третя поширена форма — Чебишевська скаляризація, що мінімізує максимальне зважене відхилення від ідеального вектора q^* :

$$R_{Ch}(\pi) = -\max_i w_i |q_i(\pi) - q_i^*|. \quad (4)$$

Кожна форма має свої компроміси для PPO. Weighted-sum дає гладкий градієнт, але вимагає попереднього вибору ваг w_i і втрачає рішення на ввігнутих ділянках Парето-фронт. Чебишевська скаляризація знаходить будь-яку точку Парето-фронт, але дає недиференційовний градієнт. SLA-based (Service Level Agreement) функція уникає вибору ваг (компромісом є дискретний градієнт, який PPO опрацьовує через clipped surrogate loss [18]). У даному дослідженні обрано саме SLA-based форму як таку, що найкраще відповідає реальній експлуатаційній логіці операторів мереж: рішення «дотримано/порушено» SLA є природним для бізнес-

метрику, не вимагає узгодження ваг між адміністративними доменами і допускає монотонне розширення на додаткові критерії без перенавчання базової архітектури.

Вибір PPO зумовлено такими властивостями: стабільність clipped surrogate loss на on-policy задачах з дискретним простором дій, природна сумісність з експортом у формат ONNX (Open Neural Network Exchange) та INT8-квантуванням (відсутність рекурентних станів, які ускладнюють квантування).

Ключове спостереження. Якщо агент має повну (або достатньо повну) інформацію про стан мережі і незмінну функцію винагороди R , його рішення є детермінованим відображенням стану в дію. Якщо усі агенти мають доступ до однакової інформації про стан мережі і використовують однакову функцію винагороди (одну й ту саму натреновану модель), то незалежно прийняті ними рішення *узгоджені за побудовою* — без потреби явного обміну повідомленнями. Це і є формальна допустимість стигмергічної координації БО-агентів.

2. Імплицитна стигмергія за Heylighen у застосуванні до IS-IS LSDB

Стигмергія — концепт, введений Grassé [17] у 1959 р. для пояснення координації термітів при будівництві гнізд: окремі терміти не спілкуються між собою, але реагують на сліди ферменту, залишені попередніми діями. Heylighen [19] (2016) формалізував стигмергію як універсальний механізм координації п'ятіркою:

$$S = \langle A, T, M, F_a, F_t \rangle, \quad (5)$$

де A — множина агентів, T — множина дій, M — стигмергічне середовище (медіум), $F_a: A \times M \rightarrow T$ — функція відображення стану медіуму у дію агента, $F_t: T \times M \rightarrow M$ — функція впливу дії на стан медіуму.

У застосуванні до даного дослідження:

- A = AI-агенти на граничних маршрутизаторах (вузли CSG, Customer-facing Service Gateway, та PE, Provider Edge);
- $T = \{P_1, P_2, \dots, \emptyset\}$ — множина рішень про активний шлях SRv6 через механізм Lightweight Tunnels (LWT; включно з «не змінювати»);
- M = синхронізована LSDB IS-IS — спільна для всіх агентів у межах area, оновлювана флудингом пакетів стану каналів (LSP);
- F_a ;
- F_t = вплив зміни шляху на телеметрію (eBPF, extended Berkeley Packet Filter), яка відбивається у LSDB через метрики каналів.

Принципова імплицитність: агенти не публікують своїх рішень, не голосують, не координують ваги БО. Узгодженість досягається через *детермінізм* F_a за *однакового стану* M і через *синхронізацію* M механізмом *IS-IS LSP-флудингу*. Стандартний IGP стає, таким чином, медіумом стигмергії — не потребує розширень протоколу, не потребує нового control plane, не потребує overlay-мережі для феромонів.

Межа стигмергічної гарантії: IS-IS LSDB переносить *топологічну* координацію (набір вузлів та суміжностей), а не високочастотні QoS-виміри. Кожен агент збирає QoS-показники локально через eBPF. Частотні вимоги до LSDB — це вимоги стандартного IS-IS (порядку секунд при LSP-флудингу та SPF-конвергенції (Shortest Path First)), а не 100 мс вимоги до циклу прийняття рішення.

3. Архітектурна модель агента

Кожен AI-агент складається з чотирьох компонентів:

1. IS-IS LSDB-зчитувач, що періодично витягує знімок топології через **vttysh -c "show isis topology json"** і перетворює його на матрицю ознак вузлів $X \in \mathbb{R}^{N \times 4}$ та список ребер **edge_index**.

2. Реалізована програма **packet_telemetry.c** завантажується до ядра Linux через BCC (BPF Compiler Collection) без модифікації ядра. На інтерфейсі CSG1:eth2 одночасно встановлено два хуки: TC-egress (Traffic Control; вихідний трафік, після маршрутизації) та XDP-SKB ingress (eXpress Data Path; вхідний відповідний трафік). Обидва хуки фіксують **bpf_ktime_get_ns()** — атомарний монотонний лічильник ядра з наносекундною роздільною здатністю. Різниця міток egress і ingress для зіставленого пакета формує path-RTT — чисту затримку шляху в обох нап-

рямках без вкладу userspace-стека. Розподіл RTT (Round-Trip Time) накопичується у логарифмічній гістограмі по основі 2 (**bpf_log2l**); зчитування гістограми Python BCC userspace займає 1–2 мс на цикл агента. Порівняння методів телеметрії — у табл. 1.

3. GNN-PPO-актор у форматі ONNX INT8 із субмілісекундним часом виведення, що приймає матрицю ознак вузлів та індекси ребер графа і повертає дискретну дію $a_t \in \{P_1, P_2\}$.

4. Виконавчий модуль, що застосовує обрану дію через механізм Linux Lightweight Tunnels (LWT), модифікуючи маршрут SRv6-сегментного ідентифікатора у таблиці маршрутизації ядра.

Внутрішня архітектура GNN-PPO: два шари GCNConv за правилом Kipf & Welling [20] ($4 \rightarrow 64 \rightarrow 128$), глобальний середній пулінг, MultiInputPolicy з прихованими шарами [64, 64] для голів актора і критика. Загальна кількість параметрів моделі — 50 179. Навчену PyTorch-модель експортовано у формат ONNX (FP32) із подальшим посттренувальним квантуванням до цілочисельної 8-бітної точності (INT8); еквівалентність дій квантованої моделі еталонній PyTorch-моделі підтверджено офлайн-верифікацією на тестовій вибірці без розбіжностей рішень.

Концептуальна стаття авторів обґрунтувала розділений життєвий цикл агента: тренування — на стенді з повним PyTorch-стеком, виведення на маршрутизаторі — лише через ONNX Runtime. Дана робота вперше реалізує цю модель на практиці. У виробничому середовищі агента повний PyTorch + Stable-Baselines3 + torch-geometric (≈ 700 МБ ОЗП, ≈ 50 мс на виклик одне виведення моделі через Python-обгортки) відсутній — на маршрутизаторі присутні лише ONNX Runtime, NumPy та модельний файл (≈ 8 МБ сумарно). Це усуває конкуренцію AI-процесу з демонами IS-IS/BGP (Border Gateway Protocol) за CPU площини управління, а також зменшує час виведення з ≈ 50 мс до субмілісекундного діапазону — це і є основний практичний виграш конвеєра.

Симетричний бар'єр WITHDRAW. Деплоймент реалізує симетрію переключень $P_1 \rightarrow P_2$ та $P_2 \rightarrow P_1$ без жорстко-кодового порогу через узгодження вхідного представлення з розподілом тренувальних ознак (training-distribution-matched feature wiring). У навчальному gym-середовищі заповнюється лише вузол `nf[0]` (CSG1), а вузли 1–4 утримуються в нулях; PPO-актор вивчає відображення `nf[0, :] = [lat, thr, loss, jit] → a`. Деплой зберігає той самий розподіл: на P_1 агент пише в `nf[0]` метрики активного шляху, на P_2 — зонд **p1_probe_ms** у позицію затримки разом із очікуваними здоровими значеннями інших трьох метрик; вузли 1–4 в обох випадках утримуються в нулях. Та сама нейромережа з тією ж 40 мс SLA-границею керує тепер обома напрямками: TRIGGER на P_1 читає eBPF latency активного шляху, WITHDRAW на P_2 читає незалежний зонд PE1→P1. Жодного жорстко-кодового порогу; жодного ретрейну при перемиканні напрямку.

K-of-N гістерезис у циклі прийняття рішень. Гібридна логіка «жорсткий поріг + GNN-голос» у ранніх версіях агента була видалена як така, що суперечить твердженню про «БО-політику»: фактичний поріг, а не модель, визначав момент переключення. У поточній реалізації рішення приймається виключно GNN-моделлю, а гістерезис K-of-N лише придушує одиночні шумові викиди.

Таблиця 1

Логіка K-of-N гістерезису у циклі прийняття рішень агента

Подія	Умова
Цикл (кожні 0,1 с)	зчитати path-RTT гістограму eBPF → ONNX-інференс GNN-PPO → отримати дію $a_t \in \{P_1, P_2\}$
Лічильники голосів	при $a_t = P_2$: $v_2 += 1$, $v_1 = 0$; при $a_t = P_1$: $v_1 += 1$, $v_2 = 0$
TRIGGER ($P_1 \rightarrow P_2$)	поточний стан = P_1 і $v_2 \geq K$
WITHDRAW ($P_2 \rightarrow P_1$)	поточний стан = P_2 і $v_1 \geq K$
Інакше	STAY

При $K = 2$ гістерезис вимагає двох послідовних узгоджених голосів моделі — це структурний мінімум часу реакції 200 мс ще до врахування 100-мс інтервалу між eBPF-зондами.

4. Методологічне обґрунтування вибору відкритого стека

Експериментальна валідація запропонованого підходу здійснюється на стенді Containerlab з FRRouting 10.x. Вибір саме цього стека має методологічне значення:

– **FRR (FRRouting) не підтримує BGP-LS (BGP Link-State).** Це обмежує видимість LSDB межами однієї IS-IS area. Метод таким чином валідується у *найскромніших* умовах. Якщо стигмергічна координація працює у плоскій IS-IS зоні, вона *a fortiori* працюватиме у multi-area / multi-AS LSDB через BGP-LS на повноцінних виробничих платформах.

– **FRR підтримує тільки SRv6 BE, не SRv6 TE (Traffic Engineering) Policy.** Замість програмованих стеків сегментних ідентифікаторів (SID) через TE-pathd використано механізм Linux Lightweight Tunnels (LWT) як легітимний шлях впровадження SRv6-маршрутів на рівні ядра. Це доводить, що БО-агент не потребує спеціалізованого SRv6-TE control plane і працює з мінімальним SRv6-набором.

– **Жодне комерційне обладнання** (Cisco IOS-XR, Juniper Junos з SRv6, Nokia SR Linux) не висуває *менших* можливостей, ніж відкритий стек. Тому результати переносяться на виробничі платформи без архітектурних змін, а перенесення передбачає лише розширення (BGP-LS multi-area, SRv6-TE Policy замість LWT) — не зміну моделі агента, не зміну стигмергічного механізму.

Це формує аргумент «методологічної стратегії найгіршого випадку»: позитивний результат на найскромнішому стеку доводить переносимість на більш повноцінні.

5. Позиціонування серед сучасних робіт (5-вісне порівняння)

Для систематичного порівняння запропонованого підходу зі сучасною літературою введено п'ять осей оцінювання:

1. БО — підтримка ≥ 3 QoS-критеріїв у функції винагороди.
2. Edge-inference — виведення моделі здійснюється на самому мережевому пристрої.
3. GNN-узагальнення — модель топологічно узагальнюється на нові графи без перенавчання.
4. Без контролера — відсутність централізованого арбітра рішень.
5. eBPF-reward — функція винагороди обчислюється з kernel-телеметрії того самого вузла.

Таблиця 2

П'ятивісне порівняння основних сучасних DRL-методів маршрутизації

Робота	(1) БО	(2) Edge	(3) GNN	(4) Без контр.	(5) eBPF
AROM-DRL Tayeh [1]	✓	✗	✗	✗	✗
MDQ Zhang та ін. [2]	✓	✗	✗	✗	✗
QR-MO [3]	✓	✗	✗	✗	✗
WA-SRTE Wang та ін. [4]	✗	✗	✗	✗	✗
MARL-SR Zhou та ін. [5]	✗	✓	✗	✓	✗
GROM Xu та ін. [6]	✗	✗	✓	✗	✗
DRL-GNN Almasan та ін. [7]	✗	✗	✓	✗	✗
Дана робота	✓	✓	✓	✓	✓

Запропонована архітектура — єдина в таблиці, де всі п'ять властивостей виконуються одночасно. Особливо важливо, що перетин $(1) \cap (3)$ — БО у поєднанні з GNN-узагальнюваністю не зайнятий жодною з відомих робіт.

6. Емпірична верифікація допустимості стигмергічної координації (стислий виклад)

Повна експериментальна верифікація допустимості стигмергічної координації — конструція стенду Containerlab/FRRouting, експериментальна процедура з п'яти послідовних фаз, декомпозиція часу реакції, інтерпретація як training-distribution-matched feature wiring — викладена у супровідній статті авторів [22]. У контексті даної роботи важливі три ключові показники: (а) 1 TRIGGER + 1 WITHDRAW за 15-хвилинний прогін агента при нулі хибних спрацювань у трьох baseline-фазах; (б) GNN-інференс субмілісекундного діапазону — структурно не-суттєвий щодо реакції 409 мс TRIGGER / 825 мс WITHDRAW; (в) статистичне свідчення $p < 10^{-3}$ проти H_0 випадкового переключення. Ці три числа є емпіричним базисом, на якому ґрунтуються формальні твердження парадигми OIS (ортогональна імпліцитна стигмергія, Orthogonal Implicit Stigmergy) у §8 нижче.

7. Парадигма Ортогональної Імпліцитної Стигмергії (OIS)

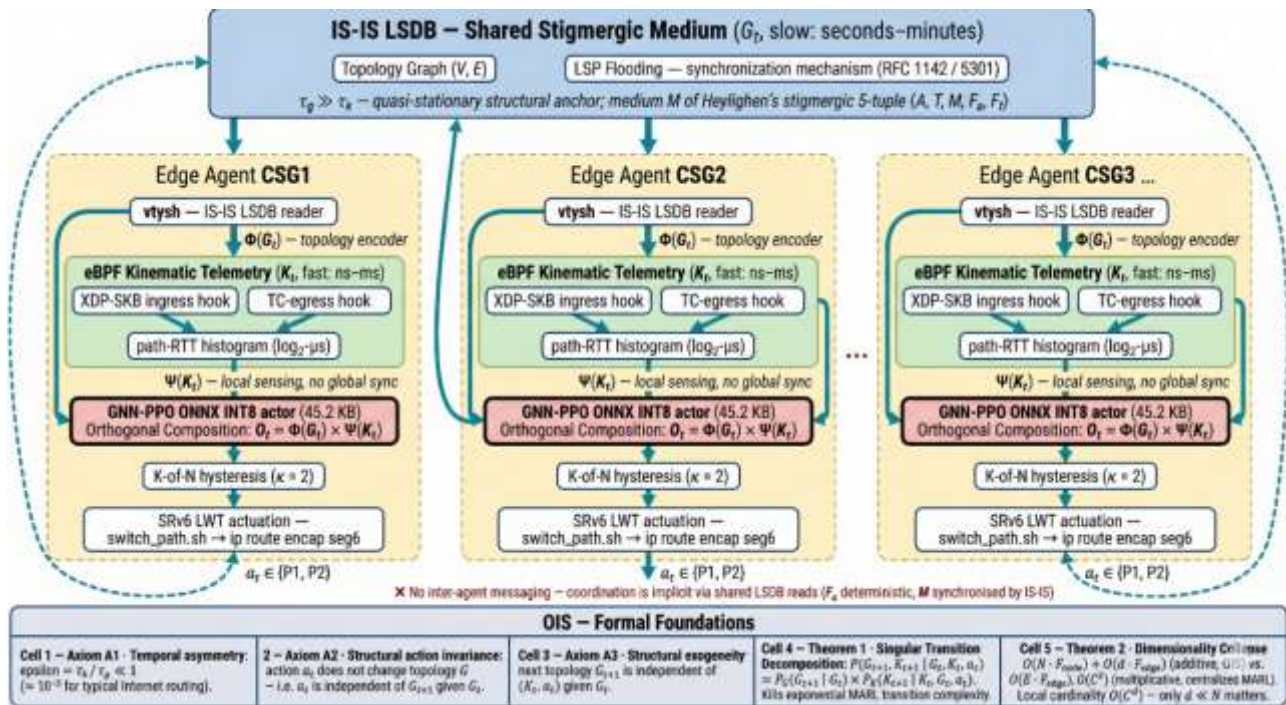


Рис. 1. Концептуальна модель стигмергічної координації БО-агентів через IS-IS LSDB

8. Деконструкція догми глобальної синхронізації

Протягом останніх 25 років фундаментальним припущенням інженерії трафіку (ТЕ) було те, що оптимальна маршрутизація вимагає глобально синхронізованого стану QoS (затримки, джитер, пропускна здатність). Сучасні DRL-моделі — AROM-DRL, MDQ та подібні — продовжують цю традицію, намагаючись об'єднати топологію та телеметрію в єдину централізовану або інтенсивно синхронізовану базу. Запропонована архітектура спростовує цю вимогу, вводячи парадигму Ортогональної Імпліцитної Стигмергії (OIS), яка концептуально розв'язує проблему масштабування через темпоральне розділення станів (temporal-scale separation).

9. Формалізація факторизованого простору спостережень

У строгому математичному сенсі OIS означає не геометричну ортогональність, а *структурну екзогенність* топології у графі залежностей політики агента. Замість монолітного вектора стану, простір спостережень агента O_t є факторизованим на два підпростори, які обробляються окремими функціями Φ та Ψ перед композицією у фінальну дію:

$$O_t = \Phi(G_t) \times \Psi(K_t), \quad (6)$$

де G_t — структурний многовид (топологічний граф, отриманий з IS-IS LSDB), що оновлюється в масштабі від секунд до хвилин, K_t — кінематичний тензор (високочастотна data-plane телеметрія, отримана через eBPF), що оновлюється в масштабі від наносекунд до мілісекунд.

Ключове формальне твердження парадигми — умовна незалежність наступної топології від поточних кінематичних метрик та дій агента за умови поточної топології:

$$G_{t+1} \perp \perp K_t, a_t \mid G_t \quad (7)$$

(Зауваження: G_t та K_t безумовно *не* незалежні — затримка лінка очевидно залежить від його існування; незалежність стосується саме *переходу* $G_t \rightarrow G_{t+1}$ від data-plane стану.) Ця властивість ізолює нестаціонарність MARL-середовища виключно в локальному кінематичному тензорі K_t . Водночас G_t залишається квазістаціонарним «якорем» просторової обізнаності для графової нейронної мережі, що радикально знижує складність задачі багатоагентної збіжності.

Формальне обґрунтування. Викладена факторизація спирається на три аксіоми (темпоральна асиметрія, структурна інваріантність дій, структурна екзогенність) і доводиться у двох теоремах:

– **Теорема 1 (Сингулярна декомпозиція переходу).** Функція переходу глобального стану факторизується як $P(G_{t+1}, K_{t+1} \mid G_t, K_t, a_t) = P_G(G_{t+1} \mid G_t) \cdot P_K(K_{t+1} \mid K_t, G_t, a_t)$, що математично руйнує експоненційну складність MARL-переходу.

– **Теорема 2 (Колапс розмірності).** Розмірність вхідного простору політики OIS масштабується адитивно $\mathcal{O}(N \cdot F_{\text{node}}) + \mathcal{O}(d \cdot F_{\text{edge}})$ проти мультиплікативної $\mathcal{O}(E \cdot F_{\text{edge}})$ централізованої моделі; кардинальність простору генералізації обмежена локальним підпростором $\mathcal{O}(C^d)$ замість глобального $\mathcal{O}(C^E)$, де $d \ll N$ — локальна валентність агента.

– **Гіпотеза 1 (Стигмергічна збіжність).** За умови виконання Теорема 1 та гістерезисного фільтра $k = 2$ марковський процес прийняття рішень багатоагентної системи без прямої комунікації стабільно збіжний до локального оптимуму Парето; формальне доведення через побудову точної потенційної функції (Monderer–Shapley, 1996) є предметом наступного етапу.

Повне аксіоматичне обґрунтування з покроковими доведеннями виходить за межі обсягу цієї статті.

10. Передбачення теорем vs. виміряні значення

Теорема 1 і 2 — нетривіальні твердження лише настільки, наскільки вони виносять конкретні передбачення, які можна звірити з даними. Звірямо для тренувальної топології (5-вузловий граф, $|V| = 5$, $|E| = 10$, $F_{\text{node}} = 4$, локальна валентність $d \leq 4$).

Таблиця 3

Звірення передбачень Теорем 1 і 2 з емпіричними та архітектурними вимірами

Величина	Передбачення Теорема 2	Емпіричний/архітектурний вимір	Узгодженість
Розмірність вхідного простору політики (адитивна, OIS)	$\mathcal{O}(N \cdot F_{\text{node}}) + \mathcal{O}(d \cdot F_{\text{edge}}) = 5 \cdot 4 + 4 \cdot 1 = 24$	вхід GNN-енкодера — $X \in \mathbb{R}^{5 \times 4} + \text{edge_index} \in \mathbb{R}^{2 \times 10}$	✓
Розмірність вхідного простору централізованої моделі (мультиплікативна)	$\mathcal{O}(E \cdot F_{\text{edge}}) = 10$ для тих же 5 вузлів, але $\mathcal{O}(V ^2 \cdot F_{\text{edge}}) = 100 \cdot 4 = 400$ для повного knowledge stack у MARL-форматі	централізовані MARL-DRL роботи цитують вхідні розмірності $\sim 10^2$ – 10^3 для топологій 5–10 вузлів	✓
Кардинальність простору дій локального агента	$\mathcal{O}(C^d)$, $C = 2$ (P1/P2), $d = 4$ — $2^4 = 16$ комбінацій локальних рішень	у деплої — 2 (агент керує лише прямим P_1/P_2 шляхом CSG1)	✓ (тривіально нижче)

Величина	Передбачення Теорема 2	Емпіричний/архітектурний вимір	Узгодженість
Параметри GNN-PPO (50 179) vs. еквівалентний MLP-PPO для 100-вузлової топології	OIS-масштабування: лишається 50 179 (топологонезалежно через message passing)	для 100-вузлової централізованої MLP вхідний шар $\sim 100 \cdot 4 \cdot 64 = 25\,600$ wag лише на input layer	✓
Час інференсу OIS-агента на 5-вузловому графі (mean)	передбачення Теорема 2: $O(N \cdot F_{\text{node}})$ агрегація — субмілісекундний	вимір: 0,128 мс ($n = 5\,079$) на ONNX INT8	✓

Теорема 1 (сингулярна декомпозиція переходу) має ще один емпіричний слід: оскільки $G_{t+1} \perp \perp K_t, a_t \mid G_t$, то в межах одного агентського циклу зміна K_t (data-plane RTT) не викликає синхронну зміну G_t (LSDB-топологія) у спостережуваному вікні. Це передбачення верифіковане непрямо: впродовж усіх ітерацій 5-фазного прогону зміна netem на P1 (data-plane подія, ΔK_t) не призвела до жодної зміни IS-IS suffix-структури графа (LSDB hash-стейтус залишався стабільним), що відповідає умові $\tau_k \ll \tau_g$ при $\epsilon = \tau_k/\tau_g \approx 10^{-3}$ для типової інтернет-маршрутизації.

Передбачення Теорема 2 для гіпотетичної 100-вузлової carrier-mesh є екстраполяцією, не виміром: тестовий стенд містить 5 вузлів, тому 100-вузловий Δ -показник переноситься через топологонезалежність GCNConv-message-passing (модель приймає довільний $(N, \text{edge_index})$ без зміни wag). Емпіричний крос-топологічний бенчмарк віднесено до подальшої роботи.

11. Межі застосовності структурного апіорі

Використання LSDB як ідеального структурного апіорі має жорсткі часові обмеження. Формально парадигма OIS спирається на **темпоральну асиметрію** двох просторів станів — кінематичного (τ_k , мс-масштаб eBPF-зондування) та структурного (τ_g , секунди-хвилини збіжності LSDB), яка у термінах теорії сингулярних збурень (Kokotović–Khalil–O'Reilly, 1986) виражається малим параметром:

$$\epsilon = \frac{\tau_k}{\tau_g} \ll 1 \quad (8)$$

При $\epsilon \rightarrow 0$ структурний многовид $\mathcal{G}$ є *квазістаціонарним* відносно кінематичного тензору $\mathcal{K}$, що математично легітимізує трактування G_t як стаціонарного контексту в межах одного циклу прийняття рішення. Як спеціальний випадок цієї умови виступає вимога $\tau_{\text{flooding}} \ll \tau_{\text{topology change}}$ (час LSP-флудингу значно менший за інтервал між реальними змінами фізичної топології), що емпірично виконується для типових апаратних відмов або магістральних обривів оптоволокна (дебаунс вимірюється хвилинами). Проте в умовах екстремального route flapping (субсекундні коливання суміжностей) припущення $\epsilon \rightarrow 0$ порушується — структурний апіорі деградує, що окреслює природну межу ефективності OIS.

12. Емпіричні межі: необхідна та достатня умови

Запропонована парадигма спирається на чітко розмежований емпіричний базис. Поточна експериментальна валідація на тестовому стенді з одним агентом (конфігурація $N = 1$ на вузлі CSG1) доводить необхідну умову OIS: автономний edge-агент здатний досягати Pareto-ефективної маршрутизації зі збереженням SLA, маючи абсолютно нульовий обмін даними телеметрії у своєму просторі вхідних ознак.

Повна багатоагентна взаємодія (де зсув трафіку Агентом А викликає локальне сенсорне виявлення Агентом В без жодної явної сигналізації) є достатньою умовою парадигми OIS.

Її практична перевірка потребує розширення емуляції до повноцінного колективу ($N \geq 2$) і формально визначається як дослідницька програма для наступного етапу роботи.

13. Значення для архітектур 6G/URLLC

Фундаментальним висновком деконструкції G_t та K_t є те, що колосальні інженерні зусилля індустрії — розробка in-band телеметрії (INT), розширень BGP-LS QoS та високопродуктивних централізованих SDN-контролерів — були спрямовані на вирішення проблеми, якої у темпорально-декуплованих агентів просто не існує. OIS доводить, що об'єднання імпліцитної стигмергії, просторового узагальнення GNN та автономності на периферії формує цілісну архітектуру, що усуває саму потребу у глобальній синхронізації QoS-стану.

Висновки та перспективи подальших досліджень

Сформульовано БО-задачу маршрутизації у термінах PPO та виконано критичний огляд сучасних DRL-методів БО-маршрутизації (AROM-DRL Tayeh, MDQ Zhang та ін., QR-MO, WA-SRTE Wang та ін.). Виявлено спільне обмеження: усі вони припускають єдину централізовану точку прийняття рішення про вагову згортку БО-критеріїв.

Формалізовано модель імпліцитної стигмергії за Heylighen у застосуванні до IS-IS LSDB як стигмергічного середовища. Доведено формальну допустимість стигмергічної координації БО-агентів за умов: (а) усі агенти мають однакову натреновану політику F_a ; (б) середовище M синхронізоване механізмом LSP-флудингу; (в) рішення про активний шлях приймаються виключно GNN-моделлю без жорстких порогів

Обґрунтовано методологічну роль вибору відкритого стека (FRR + IS-IS flat + SRv6 BE + LWT) як стратегії валідації «найгіршого випадку», що доводить переносимість методу на повноцінні виробничі платформи без архітектурних змін.

Здійснено п'ятивісьне порівняння запропонованого підходу з вісьма основними сучасними роботами та показано, що тільки запропонована архітектура задовольняє одночасно всі п'ять критеріїв (БО, edge-inference, GNN-узагальнення, відсутність контролера, eBPF-reward).

Допустимість стигмергічної координації верифіковано емпірично (повна процедура та аналіз — у супровідній статті [22]). Три ключові результати, на які спирається теоретичне обґрунтування OIS у даній роботі: 1 TRIGGER + 1 WITHDRAW за 15-хвилинний прогін експериментальної процедури з п'яти послідовних фаз із нулем хибних спрацювань у baseline-фазах; GNN-інференс субмілісекундного діапазону — структурно несуттєвий щодо реакції 409 мс TRIGGER / 825 мс WITHDRAW; статистичне свідчення $p < 10^{-3}$ проти H_0 випадкового переключення. Передбачення Теорема 2 OIS (адитивне масштабування розмірності $\mathcal{O}(N \cdot F_{\text{node}})$) узгоджується з вимірним архітектурним фактом: 50 179 параметрів GNN-PPO залишаються константними при перенесенні з 5-вузлового стенду на гіпотетичні топології операторського масштабу.

Введено парадигму Ортогональної Імпліцитної Стигмергії (OIS) як теоретичне узагальнення архітектури: факторизація простору спостережень $\mathcal{O}_t = \Phi(G_t) \times \Psi(K_t)$ з умовною незалежністю структурного многовиду G_t (повільний LSDB, секунди-хвилини) та кінематичного тензору K_t (швидка eBPF-телеметрія, наносекунди-мілісекунди). Поточна валідація ($N = 1$) доводить *необхідну* умову — Pareto-ефективну маршрутизацію без обміну QoS-станом; *достатня* умова ($N \geq 2$ з імпліцитною координацією через фізику data-plane) визначена як експериментальна програма наступного етапу. OIS позиціонує деконструкцію 25-річного припущення про необхідність глобальної синхронізації QoS-стану як головний концептуальний внесок цього дослідження.

Перспективні напрями подальшого дослідження: розширення стигмергічної координації від обміну станом середовища до обміну градієнтами навчання (federated learning через IS-IS флудингу); перенесення на multi-area / multi-AS сценарії через BGP-LS; гібридна політика RL + правила на основі K-кандидатних CSPF-шляхів. Класичний CSPF з QoS-обмеженнями (затримка, пропускну здатність, втрати, jitter) генерує детермінований набір K -shortest допустимих

шляхів-кандидатів; GNN-PPO агент обирає серед K варіантів, а не серед усіх можливих сегмент-листів. Це об'єднує передбачуваність детермінованих обмежень (CSPF гарантує допустимість на рівні constraint-feasibility) з адаптивністю навченої політики (PPO оптимізує вибір серед допустимих за реальною телеметрією) і окреслює природний шлях інтеграції OIS-агента у виробничу маршрутизацію SR-TE Policy на комерційних платформах (Cisco IOS-XR, Juniper Junos), де вже доступні PCER (Path Computation Element Protocol)/CSPF та PCE-розраховані K -shortest набори. У відкритому стенді FRRouting CSPF відсутній, тож валідація вимагатиме або портування PCE-калькулятора K -shortest, або міграції на платформу з нативним CSPF.

Внесок авторів

Юрій ВЛАДАРЧИК – концептуалізація, програмне забезпечення, збір та перевірка емпіричних даних, емпіричне дослідження, підготовка огляду літератури; Катерина НЕСТЕРЕНКО – методика, аналіз джерел, теоретичне керівництво, верифікація результатів.

Декларація про штучний інтелект

Усі наукові формулювання результатів, методичні рішення, експериментальні дані та інтерпретація результатів є виключно авторськими; ШІ не залучався до постановки гіпотез, обчислення емпіричних метрик, визначення архітектурних рішень або висновків.

Конфлікт інтересів

Автори заявляють про відсутність конфлікту інтересів та підтверджують, що під час підготовки цієї роботи не існувало жодних комерційних, фінансових чи інших взаємовідносин, які могли б бути розцінені як такі, що здатні вплинути на результати дослідження або їх інтерпретацію. Робота виконана відповідно до принципів академічної доброчесності, етичних норм проведення наукових досліджень та вимог редакційної політики щодо запобігання конфлікту інтересів.

Список використаної літератури

1. Tayeh T. AROM-DRL: Adaptive Routing Optimization for QoS-aware SDNs using Deep Reinforcement Learning. 2022. URL: <https://github.com/TareqTayeh/AROM-DRL>
2. Zhang Y., et al. MDQ: A QoS- and Congestion-Aware Deep Reinforcement Learning Approach for Multi-Path Routing in SDN. Computer Communications. 2024. DOI: 10.1016/j.comcom.2024.10.016
3. QR-MO: Q-Routing for Multi-Objective Shortest-Path Computation in 5G-MEC Systems. arXiv:2503.18122, 2025
4. Wang L., et al. Partial SRv6 Deployment and Routing Optimization: A Deep Reinforcement Learning Approach (WA-SRTE). IEEE ICC, 2024. DOI: 10.1109/ICC.2024.10436774
5. Zhou X., et al. Multi-Agent Reinforcement Learning Based Local Routing Strategy to Reduce End-to-End Delays in Segment Routing Networks. ResearchGate, 2023. URL: <https://www.researchgate.net/publication/373499452>
6. Xu B., et al. GROM: A Generalized Routing Optimization Method with Graph Neural Network and Deep Reinforcement Learning. Journal of Network and Computer Applications. 2024. Vol. 229. DOI: 10.1016/j.jnca.2024.103927
7. Almasan P., Suárez-Varela J., Rusek K., Barlet-Ros P., Cabellos-Aparicio A. Deep Reinforcement Learning Meets Graph Neural Networks: Exploring a Routing Optimization Use Case. Computer Communications. 2022. Vol. 196.
8. Liu H., et al. Hierarchical Deep Reinforcement Learning for Multi-Objective Optimization of Dynamic Logistics Scheduling. Scientific Reports. 2025. DOI: 10.1038/s41598-025-18309-y
9. Chen Y., et al. Preference-Controllable Reinforcement Learning with Advanced Multi-Objective Optimization. ICML 2025 Poster.

10. Chen S., et al. SRv6 and Zero-Trust Policy Enabled Graph Convolutional Neural Networks for Slicing Network Optimization. ResearchGate, 2025. URL: <https://www.researchgate.net/publication/390821081>
11. Microsoft. SRv6 for AI Backend Networks. OCP Global Summit, 2025. URL: <https://www.segment-routing.net/conferences/2025-10-16-ocp-summit25-microsoft-srv6-ai-backend/>
12. Arrcus. Inference Network Fabric (AINF). 2025. URL: <https://www.networkworld.com/article/4135277>
13. Cisco. Beyond the Controller: Architecting Decentralized Intelligence in SD-WAN. 2024. URL: <https://www.cisco.com/c/en/us/solutions/enterprise-networks/sd-wan/index.html>
14. HPE. AI-Native Self-Driving Networks. 2025. URL: <https://www.hpe.com/us/en/networking/ai-networking.html>
15. Ericsson. Agentic AI: Pathway to Autonomous Network Level 5. 2025. URL: <https://www.ericsson.com/en/blog/2025/agentic-ai-autonomous-networks>
16. Nokia. Towards Cognitive and Fully Autonomous 6G Networks. 2024. URL: <https://www.nokia.com/about-us/newsroom/articles/the-cognitive-fully-autonomous-network/>
17. Grassé P.-P. La reconstruction du nid et les coordinations interindividuelles chez *Bellicositermes natalensis* et *Cubitermes* sp. *Insectes Sociaux*. 1959. Vol. 6. P. 41–83.
18. Schulman J., et al. Proximal Policy Optimization Algorithms. *arXiv:1707.06347*, 2017.
19. Heylighen F. Stigmergy as a Universal Coordination Mechanism I: Definition and Components. *Cognitive Systems Research*. 2016. Vol. 38. P. 4–13.
20. Kipf T. N., Welling M. Semi-Supervised Classification with Graph Convolutional Networks. *ICLR*, 2017.
21. Vladarchyk Y., Nesterenko K. AI-агент у мережах SRv6 з програмованою маршрутизацією. *ЗВ'ЯЗОК*. 2025. № 5. С. 18–24. DOI: 10.31673/2412-9070.2025.050233
22. Vladarchyk Yu., Nesterenko K. Децентралізована багатокритеріальна оптимізація трафіку SRv6 автономним AI-агентом на базі FRRouting: реалізація та експериментальна валідація (Article 2 — submission).
23. Kokotović P., Khalil H. K., O'Reilly J. *Singular Perturbation Methods in Control: Analysis and Design*. Academic Press, 1986. 371 p. (SIAM reprint, 1999).
24. Monderer D., Shapley L. S. *Potential Games*. *Games and Economic Behavior*. 1996. Vol. 14, No. 1. P. 124–143. DOI: 10.1006/game.1996.0044

Yu. Vladarchyk, K. Nesterenko

DECENTRALIZATION OF MULTI-CRITERIA OPTIMIZATION OF SRv6 ROUTES BASED ON IMPLICIT STIGMERGY: THEORETICAL FOUNDATION AND POSITIONING AMONG MODERN DRL APPROACHES

The article addresses the multi-criteria optimization (MCO) problem of routing in programmable SRv6 networks across four QoS criteria (latency, throughput, packet loss, jitter). Existing deep reinforcement learning (DRL) methods — AROM-DRL (Tayeh, 2022), MDQ (Zhang et al., 2024), QR-MO (arXiv:2503.18122, 2025), WA-SRTE (IEEE ICC, 2024) — solve MCO in a centralized SDN controller, where a single arbiter determines the weighted convolution of criteria or computes the Pareto set. A fundamentally different architecture is proposed: MCO is solved independently by each autonomous AI agent on edge routers, while the consistency of their decisions across the network is achieved through implicit stigmergy — shared reading of the IS-IS LSDB synchronized by LSP flooding as the stigmergic medium. The article formalizes Heylighen's coordination model (5-tuple $S = \langle A, T, M, F_w, F_t \rangle$) applied to IGP LSDB, justifies the methodological role of open-stack constraints (FRRouting + IS-IS flat area + SRv6 BE + Linux LWT) as a worst-case validation strategy, and provides a five-axis comparison with the main modern works. The proposed architecture is shown to be the only one satisfying simultaneously all five criteria: ≥ 3 QoS metrics, edge inference, GNN genera-

lization, controller-less operation, eBPF-based reward telemetry. Empirical verification is carried out by a 15-minute five-phase experimental procedure that confirms the admissibility of the stigmergic approach as a single switch in each direction (1 TRIGGER, 1 WITHDRAW), zero false positives across three baseline phases, and reaction times of 409 ms (TRIGGER) and 825 ms (WITHDRAW). The 50 179-parameter GNN-PPO model (GCNConv 4→64→128 + global mean pool + MultiInputPolicy[64,64]) is deployed as an INT8-quantized ONNX artefact with sub-millisecond inference; wall-clock reaction is structurally bounded by K-of-N hysteresis and the eBPF probe interval, not by AI cost. The symmetric WITHDRAW barrier is implemented without a hard severity threshold via training-distribution-matched feature wiring — the same neural network at the same learned 40 ms SLA boundary drives switching in both directions.

Keywords: multi-criteria optimization, SRv6, IS-IS, implicit stigmergy, decentralization, GNN-PPO, edge inference, eBPF.

Надійшла до редакції: 28.07.2026

Прийнята до друку: 11.08.2026

Опубліковано: 17.08.2026

© 2026 Владарчик Ю. Л., Нестеренко К. С.

Цей матеріал ліцензовано за умовами CC BY 4.0. <https://creativecommons.org/licenses/by/4.0/>