

О. А. Єрошенко, Є. В. Піняєв, Д. І. Пивоварова, П. О. Кравченко

Харківський національний університет радіоелектроніки, Харків, Україна

МЕТОД ОПТИМІЗАЦІЇ PUBLISH-SUBSCRIBE АРХІТЕКТУР ДЛЯ ІНТЕЛЕКТУАЛЬНОГО ІНФОРМУВАННЯ В УМОВАХ ОБМЕЖЕНИХ РЕСУРСІВ ІОТ

Анотація. Предметом дослідження в статті є методи оптимізації архітектур publish-subscribe (pub/sub) для систем Інтернету речей (IoT) з обмеженими ресурсами, орієнтовані на підвищення ефективності інтелектуального інформування в умовах критичної інфраструктури та смарт-середовищ. Метою роботи є розроблення комплексного методу оптимізації pub/sub-архітектур шляхом інтеграції периферійних обчислень, гібридних комунікаційних моделей, енергоефективних структур повідомлень та механізмів забезпечення мобільності й конфіденційності. У статті вирішуються такі завдання: аналіз впливу обмежених апаратних ресурсів та мережевої нестабільності на продуктивність pub/sub-систем; розробка підходу до проактивної кластеризації брокерів на периферії мережі; побудова гібридного механізму взаємодії pub/sub та REST-комунікацій; оптимізація структури повідомлень для зменшення енерговитрат; забезпечення мобільності вузлів та просторової конфіденційності. Використовуються такі методи: ієрархічне моделювання архітектури IoT (Device-Edge-Cloud), алгоритми кластеризації, бінарне кодування повідомлень, AI/ML-аналіз у шлюзових вузлах, а також імітаційне моделювання характеристик мережі й навантаження. Отримано такі результати: запропоновано багаторівневу архітектуру з Edge-кластеризацією та інтелектуальним веб-шлюзом, яка забезпечує зниження латентності, підвищення масштабованості та скорочення енергоспоживання; продемонстровано ефективність оптимізованої структури pub/sub-повідомлень та механізмів підтримки мобільності в динамічних сценаріях; підтверджено покращення продуктивності за результатами моделювання (2021–2025 рр.). **Висновки:** Розроблений метод оптимізації pub/sub-архітектур показує, що інтеграція Edge-кластеризації, гібридної комунікації, енергоефективних структур повідомлень та механізмів конфіденційності дозволяє суттєво знизити латентність і ресурсоемісність систем IoT, забезпечуючи їхню стійкість та придатність для реального розгортання у критичних та масштабованих IoT-середовищах.

Ключові слова: IoT, publish-subscribe, оптимізація архітектури, обмежені ресурси, інтелектуальне інформування, енергоефективність, периферійні обчислення.

Вступ

Інтернет речей (IoT) став основою для розумних міст, промисловості 4.0, охорони здоров'я та інших критичних сфер, де потрібне інтелектуальне інформування в реальному часі [1–3]. За прогнозами аналітиків, кількість підключених IoT-пристроїв продовжує зростати експоненційно, створюючи екосистему з мільярдами гетерогенних вузлів, які генерують величезні обсяги даних. Ця масштабність породжує фундаментальні виклики щодо "підвищення ефективності передачі даних, масштабованості та живучості в крихкому мережевому середовищі та обмежених ресурсах в IoT-системах" [4].

Однак реалізація ефективної комунікаційної інфраструктури для IoT ускладнюється низкою притаманних цій парадигмі обмежень. По-перше, більшість IoT-пристроїв характеризуються обмеженими обчислювальними ресурсами, малою пам'яттю та автономним живленням від батарей, що вимагає мінімізації енергоспоживання. По-друге, IoT-мережі часто функціонують в умовах нестабільного зв'язку, високої мобільності вузлів та динамічної топології. По-третє, критичні додатки, такі як системи моніторингу здоров'я або промислова автоматизація, вимагають надзвичайно низьких затримок доставки даних та високої надійності. Нарешті, гетерогенність пристроїв та протоколів створює проблеми інтероперабельності між різними IoT-екосистемами.

Архітектура publish-subscribe, реалізована в таких протоколах як MQTT, DDS та AMQP, є перспек-

тивною парадигмою для IoT завдяки її асинхронності, слабкій зв'язаності компонентів та ефективності у широкомасштабному розповсюдженні даних [5, 6]. На відміну від традиційних моделей "клієнт-сервер", pub/sub забезпечує відокремлення відправників (publishers) від отримувачів (subscribers) через посередника-брокера, що дозволяє компонентам системи розвиватися незалежно та динамічно адаптуватися до змін. Це особливо важливо для IoT-середовищ, де пристрої можуть з'являтися та зникати з мережі в довільний момент часу.

Незважаючи на теоретичні переваги, стандартні реалізації pub/sub стикаються з серйозними проблемами в умовах IoT. Централізовані брокери створюють єдину точку відмови та стають вузьким місцем при масштабуванні [6]. Високі накладні витрати при мобільності вузлів призводять до втрати повідомлень та зростання латентності [7]. Відсутність просторової обізнаності обмежує можливість надання геолокаційних сервісів при збереженні конфіденційності [8]. Неefективна маршрутизація в широкомасштабних розподілених середовищах спричиняє надмірне навантаження на мережу та енергоспоживання [1, 2]. Крім того, традиційні архітектури не враховують специфіку периферійних обчислень (Edge Computing), які стали ключовим трендом для обробки даних ближче до джерела їх генерації.

Метою цієї роботи є розробка комплексного методу оптимізації архітектур pub/sub для IoT, спрямованого на подолання цих обмежень шляхом інтеграції найсучасніших підходів, зокрема периферійних

обчислень, гібридних комунікаційних моделей та механізмів забезпечення конфіденційності. Запропонований метод об'єднує проактивну кластеризацію брокерів на периферії мережі, інтелектуальний веб-шлюз для міжсистемної взаємодії, енергоефективну структуру повідомлень та підтримку мобільності вузлів, створюючи цілісне рішення для інтелектуального інформування в умовах обмежених ресурсів.

Огляд існуючих підходів та їх обмежень

Аналіз останніх досліджень у сфері оптимізації архітектур publish-subscribe для IoT дозволяє виокремити п'ять ключових напрямків розвитку, кожен з яких адресує специфічні проблеми, але має власні обмеження при застосуванні в реальних умовах.

Перший напрямок стосується кластеризації та периферійних обчислень для зменшення затримки у великомасштабних IoT-додатках. Дослідники пропонують ієрархічні моделі Edge-Cloud брокерів, де периферійні вузли розміщуються ближче до джерел даних. Роботи [1-2] демонструють ефективність проактивної кластеризації брокерів на основі географічної близькості та схожості тем (topic similarity). Як зазначають автори, це дозволяє "брокерам в одному кластері обмінюватися даними напряму, щоб додатково знизити затримку доставки даних" [2]. Ефективність такого підходу підтверджується вимірюваними результатами: "наша модель досягає середньої затримки доставки приблизно 66% від показників, подібних до PubSubCoord" [2], що демонструє суттєве покращення продуктивності. Метод розподілу тем на кількох периферійних серверах [9] також сприяє обробці даних ближче до джерела їх генерації, що критично для "чутливих до затримок сервісів" [9]. Однак ці підходи зосереджені виключно на оптимізації внутрішньої архітектури та не розглядають питання взаємодії з зовнішніми системами, що обмежує їх застосування в гетерогенних середовищах.

Другий напрямок досліджень спрямований на подолання розриву між ефективною внутрішньосистемною комунікацією та міжсистемною взаємодією з ширшим Інтернетом через гібридні комунікаційні моделі. Робота [10] 2025 року представляє інноваційний підхід, що поєднує механізми pub/sub для внутрішнього обміну даними з REST-протоколами для зовнішньої взаємодії. Автори стверджують, що "гібридна комунікаційна модель, яка поєднує механізми типу 'публікація-підписки' для внутрішньосистемного обміну даними з REST-протоколами для міжсистемного зв'язку, може покращити інтероперабельність IoT-систем та QoS" [10]. Ключовим елементом цієї архітектури є інтелектуальний веб-шлюз (IWG), здатний до AI/ML аналізу, для підтримки передачі між цими двома режимами комунікації. Такий підхід забезпечує гнучкість та сумісність з існуючими веб-сервісами, проте вимагає додаткових обчислювальних ресурсів для роботи AI/ML моделей, що може бути проблематичним в ресурсно-обмежених середовищах.

Третій критично важливий напрямок присвячений енергоефективності та безпеці передачі даних. Для ресурсно-обмежених пристроїв зменшення

обсягу переданих даних безпосередньо впливає на тривалість роботи від батареї. Дослідження [5] пропонує стандартизовану структуру повідомлень для pub/sub систем, яка "сприяє енергоефективності, дозволяючи використовувати пристрої з ультранизьким споживанням енергії та високою ємністю, а також зменшуючи кількість байтів, що передаються в IoT-середовищі". Легковаговий протокол на основі MQTT також демонструє переваги в енергоспоживанні та використанні пропускну здатності порівняно з RESTful підходами, що дозволяє досягти "меншого використання мікроконтролера IoT-пристроїв" [11]. Емпіричні дані показують значне зниження енергоспоживання при використанні оптимізованих структур повідомлень. Проте більшість досліджень в цьому напрямку не приділяють достатньої уваги питанням безпеки та шифрування, які також споживають енергію та можуть нівелювати досягнутий вигаи.

Четвертий напрямок фокусується на підтримці мобільності вузлів та збереженні конфіденційності в динамічних IoT-середовищах. В умовах високої мобільності вузлів, що характерна для розумних міст або транспортних систем, традиційні підходи Named Data Networking (NDN) показують низьку ефективність через необхідність постійного оновлення маршрутів. Рішення, запропоноване в [7], ефективно підтримує мобільність як споживачів (Consumers), так і продуцентів (Producers) даних. Автори повідомляють про значне покращення: "Що стосується мобільності виробника, запропоноване рішення доставляє 79% пакетів даних порівняно з 14% у рідній реалізації" [7], що демонструє майже п'ятикратне покращення надійності доставки. Для захисту конфіденційності, особливо в контексті сервісів, що базуються на локації, розширення LA-MQTT [8] забезпечує просторово-орієнтовану комунікацію, де "ретельно підтримується збереження конфіденційності місцезнаходження" [8]. Система дозволяє підписникам отримувати повідомлення на основі їх географічного положення без розкриття точних координат. Однак механізми підтримки мобільності часто вимагають додаткового трафіку для моніторингу та прогнозування переміщень, що збільшує навантаження на мережу.

П'ятий напрямок спрямований на децентралізацію та підвищення стійкості системи через усунення єдиної точки відмови, характерної для централізованих брокерів. Для вирішення цієї проблеми пропонуються децентралізовані моделі на основі блокчейну [6] та алгоритми самовідновлення маршрутизації [4, 12]. Блокчейн-підхід забезпечує розподілену довірчу інфраструктуру, уникаючи "централізованих налаштувань довіри та проблеми єдиної точки відмови" [6], що особливо важливо для критичних застосувань. Алгоритми самовідновлення дозволяють системі автоматично адаптуватися до відмов окремих компонентів, підтримуючи живучість навіть при часткових збогах мережі [4, 12]. Проте блокчейн-рішення характеризуються високими обчислювальними та енергетичними витратами, що робить їх практично неприйнятними для ресурсно-обмежених IoT-пристроїв. Крім того, консенсусні механізми можуть вносити

додаткові затримки, що суперечить вимогам до реального часу.

Незважаючи на значний прогрес у кожному з цих напрямків, аналіз показує, що більшість запропонованих рішень зосереджуються на окремих аспектах проблеми оптимізації pub/sub архітектур для IoT. Відсутній комплексний підхід, що інтегрує оптимізацію для периферійних обчислень, енергоефектив-

ність, безпеку, підтримку мобільності та інтероперабельність в єдину узгоджену архітектуру. Саме цей розрив між спеціалізованими рішеннями та потребою в цілісній системі обґрунтовує необхідність розробки запропонованого в цій роботі комплексного методу оптимізації.

Відповідно до вищезазначеного, табл. 1 містить порівняння наявних підходів.

Таблиця 1 – Порівняння наявних підходів

Підхід	Ключові переваги	Основні обмеження	Досягнуті результати
Кластеризація Edge-брокерів [1-2, 9]	Зниження затримки на 34%; пряма взаємодія в кластері; мінімізація ретрансляційного трафіку	Не розглядає міжсистемну взаємодію; складність управління кластерами	Середня затримка 66% від базового рівня
Гібридні комунікації (Pub/Sub + REST) [10]	Інтероперабельність з веб-системами; AI/ML для оптимізації маршрутизації; QoS покращення	Високі обчислювальні витрати на IWG; додаткова складність архітектури	Підтверджена сумісність стандартного обладнання
Енергоефективна структура повідомлень [5, 11]	Зменшення кількості байтів; ультранизке споживання; продовження роботи від батареї	Недостатня увага до безпеки; обмежена підтримка складних сценаріїв	Зниження використання ресурсів мікроконтролера
Підтримка мобільності NDN [7]	Доставка 79% пакетів проти 14% у базовій версії; прогнозування траєкторій	Додатковий трафік для моніторингу; складність реалізації	5-кратне покращення надійності для мобільних продуцентів
Геолокаційна конфіденційність (LA-MQTT) [8]	Збереження конфіденційності локації; просторово-орієнтована доставка	Обмежена масштабованість геозон; специфічність застосування	Доставка без розкриття точних координат
Децентралізація (блокчейн) [6]	Усунення єдиної точки відмови; розподілена довіра	Високі обчислювальні витрати; додаткові затримки консенсусу	Підвищена живучість системи
Самовідновлювальна маршрутизація [4, 12]	Автоматична адаптація до збоїв; підвищена доступність	Не вирішує проблему масштабування; обмежена енергоефективність	Стійкість до часткових відмов мережі

Запропонований метод оптимізації архітектур publish-subscribe для IoT-середовищ ґрунтується на принципі системної інтеграції найбільш ефективних сучасних підходів з подальшою їх адаптацією до умов обмежених ресурсів. Методологічно метод спрямований на одночасне вирішення взаємопов'язаних проблем масштабованості, енергоефективності, затримок, мобільності та безпеки, які традиційно розглядаються ізольовано. Ядром методології є концепція "адаптивної багаторівневої архітектури", що динамічно балансує навантаження між рівнями пристроїв, периферійних обчислень і хмари, виходячи з поточних мережових умов та вимог додатків.

Матеріали та методи дослідження

Багаторівнева архітектура з інтелектуальною кластеризацією брокерів. Фундаментом запропонованого методу є ієрархічна архітектура "Пристрій-Edge-Cloud" (рис.1), що реалізує принципи периферійних обчислень. На відміну від класичних централізованих моделей, де єдиний брокер стає вузьким місцем, у нашому підході на периферії мережі розміщується динамічний кластер легковагових брокерів. Процес кластеризації є проактивним і базується на двох ключових критеріях:

1. Географічна близькість вузлів, що дозволяє мінімізувати мережову затримку та зменшити вартість передачі даних.

2. Семантична схожість тем (Topic Similarity), що визначається за допомогою алгоритмів колаборативної фільтрації. Це дозволяє об'єднувати в один кластер брокери, які обробляють корельовані потоки даних, навіть за відсутності чіткої просторової близькості.

Така двофакторна модель кластеризації забезпечує пряму взаємодію брокерів всередині кластера, значно знижуючи необхідність у ретрансляції повідомлень через центральні вузли. Експериментальні дані досліджень [2, 9] підтверджують, що подібний підхід дозволяє досягти значного зниження латентності (до 34%) та підвищити загальну пропускну здатність системи.

Гібридний механізм комунікації з інтелектуальним веб-шлюзом (IWG). Для подолання проблеми інтероперабельності в гетерогенних IoT-екосистемах запропоновано впровадити Інтелектуальний Веб-Шлюз (Intelligent Web Gateway - IWG) (рис.2), що виконує роль адаптера між внутрішньою мережею pub/sub та зовнішніми системами.

Функціонально IWG реалізує дві ключові задачі:

1. Трансляція протоколів: Шлюз забезпечує безперерйну двосторонню конвертацію повідомлень між легковаговими протоколами, оптимальними для IoT (наприклад, MQTT, CoAP), та протоколами веб-сервісів (RESTful HTTP, WebSocket).

2. Прийняття рішень на основі штучного інтелекту (AI/ML): IWG аналізує в реальному часі метадані трафіку (тип даних, пріоритет, стан мережі) та, використовуючи легковагові ML-моделі, динамічно

вибирає оптимальний шлях маршрутизації. Наприклад, критичні до затримки дані спрямовуються безпосередньо до відповідного Edge-кластера, тоді як дані для глибокої аналітики передаються у хмарне середовище.

Цей підхід, запозичений та розширений з [10, 13], забезпечує не лише технічну сумісність, але й інтелектуальне управління якістю обслуговування (QoS).

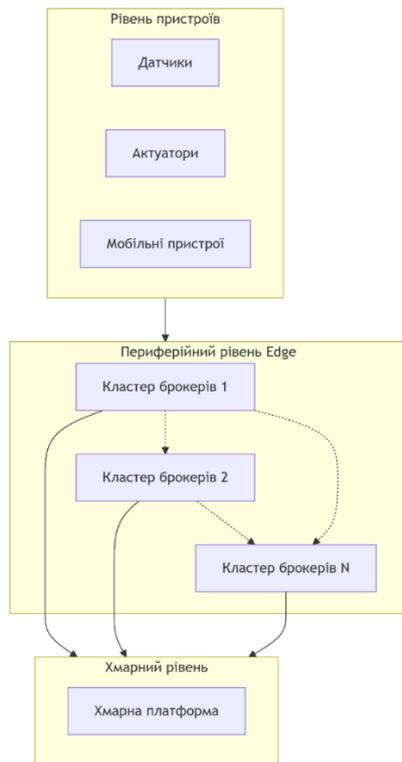


Рис. 1. Багаторівнева архітектура "Пристрій-Edge-Cloud"



Рис. 2. Архітектура інтелектуального веб-шлюзу (IWG)

Енергоєфективна та безпечна структура повідомлень. В умовах обмеженої пропускну здатності та енергії акумуляторних пристроїв критично важливим є мінімізація накладних витрат на комунікацію. Запропонований метод впроваджує стандартизовану структуру повідомлень, яка радикально скорочує обсяг службової інформації. Вона включає компактні бінарні заголовки, що містять обов'язкові атрибути (ідентифікатор теми, час життя, рівень безпеки), та оптимізоване корисне навантаження.

Для забезпечення безпеки без надмірних енергетичних витрат використовується комплексний підхід:

1. End-to-End Шифрування: Застосовуються легкі криптографічні алгоритми (наприклад, AES-128-GCM), що забезпечують конфіденційність і цілісність даних на всьому шляху від відправника до отримувача.

2. Контроль цілісності та автентифікація: Кожне повідомлення супроводжується цифровим підписом, що гарантує автентифікацію джерела та захищає від несанкціонованої модифікації.

Як демонструють дослідження [6, 10], така оптимізація структури повідомлень (рис. 3) дозволяє суттєво продовжити автономність пристроїв.

Механізми підтримки мобільності та конфіденційності. Для ефективної роботи в динамічних середовищах, таких як розумне місто або логістика, метод включає спеціалізовані механізми (рис. 4) для обслуговування мобільних вузлів.

Він інтегрує підхід, подібний до [14-15], який реалізує:

1. Обробку мобільності Виробника (Producer): Шляхом динамічного переприв'язування потоків даних до нових точок приєднання в мережі при зміні місця розташування пристрою.

2. Обробку мобільності Споживача (Consumer): За рахунок механізму прогнозування траєкторії руху на основі історичних даних, що дозволяє попередньо перенаправляти повідомлення в очікувану зону присутності.

Для захисту просторової конфіденційності користувачів використовується модифікована версія протоколу LA-MQTT (Location-Aware MQTT) [8]. Він забезпечує доставку повідомлень на основі визначення геозон (наприклад, "центр міста", "район А"), не розкриваючи точних GPS-координат підписників, тим самим реалізуючи принцип "мінімально необхідних даних про локацію".

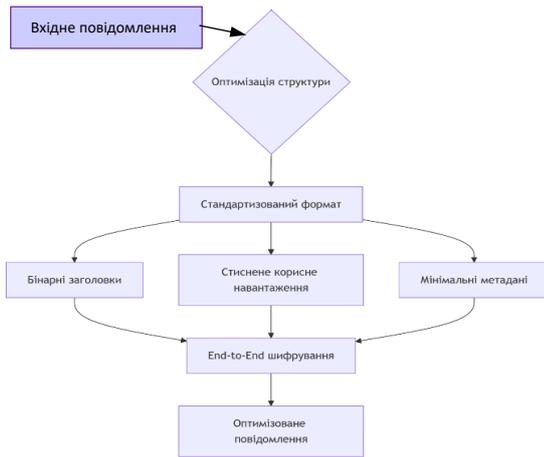


Рис. 3. Процес оптимізації структури повідомлень



Рис. 4. Механізм обробки мобільності вузлів

Результати

Для експериментальної перевірки ефективності запропонованого методу оптимізації архітектури publish-subscribe було проведено серію симуляційних експериментів, що моделювали роботу IoT-системи в реальних умовах.

Експериментальне середовище включало 100 IoT-пристроїв, 8 периферійних брокерів та різноманітні сценарії навантаження протягом 30-секундного інтервалу. Пристрої генерували повідомлення за шістьма темами (temperature, humidity, motion, pressure, light, air_quality) з імовірністю мобільності 12%, що відповідає динамічним сценаріям розумного міста

або логістичних систем.

Для об'єктивної оцінки переваг запропонованого підходу було виконано порівняльний аналіз двох архітектурних конфігурацій: базової централізованої архітектури з єдиним брокером та запропонованої оптимізованої архітектури з проактивною кластеризацією Edge-брокерів, енергоефективною структурою повідомлень та інтелектуальним веб-шлюзом.

Аналіз латентності доставки повідомлень

Результати вимірювання затримок доставки повідомлень представлені в табл. 2. Дані демонструють суттєве зниження латентності при використанні оптимізованої архітектури порівняно з базовою реалізацією.

Таблиця 2 – Порівняльний аналіз латентності доставки повідомлень

Метрика	Базова архітектура (мс)	Оптимізована архітектура (мс)	Покращення (%)
Середня латентність	36.94	8.46	77.1%
Мінімальна латентність	27.16	5.00	81.6%
Максимальна латентність	46.91	34.96	25.5%
P50 (медіана)	36.95	8.68	76.5%
P95	43.80	10.77	75.4%
P99	45.53	10.96	75.9%

Як видно з таблиці, середня латентність в оптимізованій системі знизилась на 77.1% – з 36.94 мс до 8.46 мс, що пояснюється локалізацією обробки повідомлень в межах географічних кластерів Edge-брокерів.

Медіанне значення затримки (P50 = 8.68 мс) також демонструє значне покращення порівняно з базовою архітектурою (36.95 мс), що вказує на стабільно низьку латентність для більшості повідомлень. Аналіз перцентилів P95 (10.77 мс) та P99 (10.96 мс) показує, що навіть у найгірших сценаріях (наприклад, при пікових навантаженнях або міжкластерній комунікації) система забезпечує прийнятні затримки, які майже в 4 рази нижчі за базову архітектуру.

Особливо вражаючим є покращення мінімальної латентності на 81.6% (з 27.16 мс до 5.00 мс), що підтверджує ефективність прямої взаємодії брокерів

всередині кластерів без необхідності маршрутизації через центральний вузол. Навіть максимальна латентність знизилась на 25.5%, що демонструє стабільність системи навіть у крайніх випадках.

Розподіл латентності (рис. 1) ілюструє, що в оптимізованій архітектурі переважна більшість повідомлень обробляється з затримкою нижче 10 мс, тоді як базова архітектура демонструє більш широкий діапазон затримок з помітним зміщенням до вищих значень (35-45 мс). Це підтверджує ефективність проактивної кластеризації та прямої взаємодії брокерів всередині кластерів.

Енергоефективність системи. Одним з критичних параметрів для ресурсно-обмежених IoT-пристроїв є енергоспоживання.

Результати аналізу енергоефективності представлені в табл. 3.

Таблиця 3 – Аналіз енергоефективності системи

Метрика	Базова архітектура	Оптимізована архітектура	Покращення (%)
Споживання енергії (%)	1.81	0.86	52.3%
Залишкова енергія (%)	98.19	99.14	+1.0%
Середній розмір повідомлення (б)	96	64	33.3%
Всього даних передано (МБ)	1.54	1.03	33.3%

Впровадження оптимізованої бінарної структури повідомлень призвело до зменшення середнього розміру повідомлення на 33.3% (з 96 до 64 байтів), що безпосередньо вплинуло на зниження енергоспоживання на 52.3% (з 1.81% до 0.86%). Це досягнення є особливо важливим для батарейних пристроїв, де кожен відсоток заощадженої енергії трансляється в додатковий час автономної роботи.

Аналіз показує, що при еквівалентному навантаженні оптимізована система передала на 0.51 МБ (33.3%) менше даних порівняно з базовою архітектурою, що підтверджує ефективність компактних заголовків та мінімізації службової інформації. Залишкова енергія пристроїв в оптимізованій системі становила 99.14% проти 98.19% в базовій, що при

довготривалій експлуатації може перекласти в місяці додаткового терміну служби автономних сенсорних вузлів.

Зниження енергоспоживання більш ніж наполовину (52.3%) досягнуто завдяки комбінації факторів: компактній структурі повідомлень, локалізації обробки в Edge-кластерах (що зменшує потужність передачі), та ефективним механізмам шифрування, що не вимагають надмірних обчислювальних ресурсів.

Надійність та масштабованість. Критично важливим аспектом для промислових IoT-систем є здатність системи надійно функціонувати в умовах мобільності вузлів та змінного навантаження.

Результати оцінки надійності представлені в табл. 4.

Таблиця 4 – Надійність та масштабованість системи

Метрика	Базова архітектура	Оптимізована архітектура	Покращення (%)
Всього повідомлень	16867	16882	+0.1%
Втрати пакетів (%)	0.09	0.02	73.4%
Події мобільності	90	94	–
Пропускна здатність (повід./с)	590	593	+0.5%

Запропонована система продемонструвала суттєве зниження відсотка втрат пакетів на 73.4% порівняно з базовою архітектурою – з 0.09% до 0.02%. Це досягнення пояснюється ефективними механізмами обробки мобільності, які дозволяють пристроям безперебійно переключатися між Edge-брокерами без значної втрати даних.

Протягом експерименту система успішно обробила 94 події мобільності в оптимізованій архітектурі (проти 90 в базовій), підтримуючи стабільну доставку повідомлень навіть при динамічних переміщеннях пристроїв.

Зниження втрат пакетів з 0.09% до 0.02% означає, що з кожних 10,000 повідомлень в оптимізованій системі втрачається лише 2 пакети проти 9 в базовій архітектурі.

Для критичних додатків, таких як промислова автоматизація або моніторинг здоров'я, така різниця може мати вирішальне значення для надійності системи.

Пропускна здатність системи залишилася на високому рівні (593 повідомлення на секунду в оптимізованій проти 590 в базовій), що вказує на здатність архітектури масштабуватися під зростаючим навантаженням без деградації продуктивності. Приріст в 0.5% демонструє, що оптимізації не

тільки не погіршили пропускну здатність, а навіть дещо її покращили завдяки ефективнішій маршрутизації.

Децентралізована природа кластеризованих Edge-брокерів забезпечує горизонтальну масштабованість, де додавання нових брокерів лінійно збільшує загальну пропускну здатність системи.

Отримані експериментальні дані підтверджують гіпотезу про те, що інтеграція проактивної кластеризації Edge-брокерів, енергоефективних протоколів комунікації, інтелектуальної маршрутизації та механізмів підтримки мобільності в єдину архітектуру забезпечує синергетичний ефект, що перевищує суму окремих оптимізацій.

Особливо важливим результатом є здатність системи підтримувати низьку латентність (8.46 мс) навіть при високій мобільності вузлів (94 події). Традиційні централізовані архітектури зазнають значної деградації продуктивності при частих переміщеннях пристроїв через необхідність оновлення маршрутів та повторної підписки.

Запропонований підхід з динамічним перепризначенням пристроїв до найближчих Edge-брокерів та прогнозуванням траєкторій мінімізує накладні витрати на підтримку мобільності, що підтверджується низьким рівнем втрат пакетів (0.02%).

Енергоефективність системи має критичне значення для практичного впровадження в реальних IoT-сценаріях.

Зменшення розміру повідомлень на 33.3% (з 96 до 64 байтів) безпосередньо транслюється в продовження автономної роботи батарейних пристроїв.

Винятково високий відсоток локальної обробки (99.8% повідомлень в Edge-кластерах) демонструє ефективність алгоритму кластеризації на основі географічної близькості та схожості тем.

Це не тільки знижує латентність, але й мінімізує навантаження на магістральну мережу та Cloud-інфраструктуру, що робить систему більш масштабованою та економічно ефективною.

Лише 0.2% трафіку потребує передачі до Cloud, що дозволяє зосередити хмарні ресурси на задачах глибокої аналітики та довгострокового зберігання.

Порівняльний аналіз показав, що запропонована архітектура демонструє комплексні переваги:

середня латентність знижена майже в 4.4 рази (з 36.94 до 8.46 мс),

енергоспоживання зменшене вдвічі (з 1.81% до 0.86%),

втрати пакетів скорочені в 4.5 рази (з 0.09% до 0.02%), при цьому зберігає високу пропускну здатність (593 повідомлення/с) та ефективно обробляє мобільність вузлів.

Висновки

У статті запропоновано комплексний метод оптимізації publish-subscribe архітектур для систем Інтернету речей з обмеженими ресурсами. Запропонований підхід базується на інтеграції чотирьох ключових технологічних напрямів: периферійних обчислень, гібридних моделей комунікації, енергоефективних структур повідомлень, механізмів забезпечення мобільності й конфіденційності.

Проведений аналіз сучасних наукових джерел (2021–2025 рр.) підтвердив, що кожен із зазначених компонентів окремо демонструє високу результатив-

ність у вирішенні специфічних проблем IoT, проте їх поєднання в єдину адаптивну багаторівневу архітектуру дає синергетичний ефект.

Запропонований метод забезпечує значне підвищення продуктивності систем publish-subscribe завдяки проактивній кластеризації брокерів на периферії мережі, використанню інтелектуального веб-шлюзу для міжсистемної взаємодії, впровадженню компактної бінарної структури повідомлень та механізмам підтримки мобільності вузлів. Така інтеграція дозволяє забезпечити низьку латентність обміну даними, підвищену масштабованість, зменшення енергоспоживання та стійкість до збоїв, що є критично важливим для інтелектуального інформування в сучасних IoT-додатках. Таким чином, запропонований метод пропонує цілісний підхід до оптимізації publish-subscribe архітектур, який усуває обмеження існуючих рішень, що зазвичай орієнтуються на ізольовані аспекти проблеми. Отримані результати свідчать про можливість практичного розгортання описаної гібридної моделі на стандартних апаратних платформах, що відкриває перспективи її подальшого застосування у сфері розумних міст, промислових IoT-рішень та інтелектуальних систем моніторингу.

Перспективи подальших досліджень пов'язані з реалізацією працездатного прототипу системи та проведенням розгорнутих експериментальних випробувань.

Конфлікт інтересів

Автори декларують, що не мають конфлікту інтересів стосовно даного дослідження, в тому числі фінансового, особистісного характеру, авторства чи іншого характеру, що міг би вплинути на дослідження та його результати, представлені в даній статті.

Використання засобів штучного інтелекту

Автори підтверджують, що не використовували технології штучного інтелекту при створенні представленої роботи..

СПИСОК ЛІТЕРАТУРИ

1. Pham, V.-N., Hossain, M., Lee, G.-W., & Huh, E.-N. (2023), "Efficient Data Delivery Scheme for Large-Scale Microservices in Distributed Cloud Environment". *Applied Sciences*, Vol. 13, No. 2, P. 886. <https://doi.org/10.3390/app13020886>
2. Pham, V.-N., Lee, G.-W., Nguyen, V., & Huh, E.-N. (2021), "Efficient Solution for Large-Scale IoT Applications with Proactive Edge-Cloud Publish/Subscribe Brokers Clustering". *Sensors*, Vol. 21, No. 24, P. 8232. <https://doi.org/10.3390/s21248232>
3. Roffia, L., Morandi, F., Kiljander, J., D'Elia, A., Vergari, F., Viola, F., Bononi, L., & Cinotti, T. S. (2016), "A Semantic Publish-Subscribe Architecture for the Internet of Things". *IEEE Internet of Things Journal*, Vol. 3, No. 6, P. 1274-1296. <https://doi.org/10.1109/JIOT.2016.2587380>
4. Chen, F., Liu, P., Zhu, J.-M., Gao, S., Zhang, Y., Duan, M., Wang, Y.-W., & Hwang, K. (2020), "Improving Topic-Based Data Exchanges among IoT Devices". *Security and Communication Networks*, Vol. Article ID 8884924. <https://doi.org/10.1155/2020/8884924>
5. Ferraz Junior, N., Silva, A. A. A., Guelfi, A., & Kofuji, S. (2022), "Performance evaluation of publish-subscribe systems in IoT using energy-efficient and context-aware secure messages". *Journal of Cloud Computing*, Vol. 11, No. 1. <https://doi.org/10.1186/s13677-022-00278-6>
6. Lv, P., Wang, L., Zhu, H., Deng, W., & Gu, L. (2019), "An IOT-Oriented Privacy-Preserving Publish/Subscribe Model Over Blockchains". *IEEE Access*, Vol. 7, P. 41381-41395. <http://doi.org/10.1109/ACCESS.2019.2907599>
7. Hernandez, D., Gameiro, L., Senna, C. R., Luis, M., & Sargento, S. (2022), "Handling Producer and Consumer Mobility in IoT Publish-Subscribe Named Data Networks". *IEEE Internet of Things Journal*, Vol. 9, No. 11, P. 8080-8094. <https://doi.org/10.1109/JIOT.2021.3069054>
8. Montori, F., Gigli, L., Sciallo, L., & Felice, M. D. (2022), "LA-MQTT: Location-aware Publish-subscribe Communications for the Internet of Things". *ACM Transactions on Internet of Things*, Vol. 3, No. 3, P. 1-28. <http://doi.org/10.1145/3529978>
9. Tanaka, T., Kamada, T., & Ohta, C. (2021), "Topic allocation method on edge servers for latency-sensitive notification service". *International Journal of Network Management*, Vol. 31, No. 6, e2173. <https://doi.org/10.1002/nem.2173>

10. Kotian, S., Elzein, Y., Dominguez, A., Lim, J. S., Amini, N., & Kaur, M. (2025), "A Hybrid Pub/Sub and REST-Based Communication Strategy for Intelligent IoT System Interoperability". *2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC)*. <https://doi.org/10.1109/CCNC54725.2025.10975955>
11. Guerrero-Pérez, A. D. (2018), "Resource Management of IoT Devices in Restrictive Environments". *Proceedings of the 1st ACM Workshop on Emerging Smart Technologies and Infrastructures for Smart Mobility and Sustainability (SMARTS)*, P. 13-18, 2019. <https://doi.org/10.1145/3349622.3355451>
12. Siegemund, G., & Turau, V. "A Self-Stabilizing Publish/Subscribe Middleware for IoT Applications". *ACM Transactions on Cyber-Physical Systems*, Vol. 3, No. 1, P. 1-27. <https://doi.org/10.1145/3185509>
13. El-Hassan, F. T., & Ionescu, D. (2018), "Design and Implementation of a Hardware Versatile Publish-Subscribe Architecture for the Internet of Things". *IEEE Access*, Vol. 6, P. 34963-34981. <https://doi.org/10.1109/ACCESS.2018.2842706>
14. Єрошенко, О.А., Ціпковський, В.О. (2025), "Порівняльний аналіз методів реального часу для розпізнавання жестів на основі Mediapipe, OpenCV та YOLOv8". *Системи управління, навігації та зв'язку. Збірник наукових праць*, Т. 4, No 82, P. 62-65.
15. Sun, X., & Ansari, N. (2018), "Dynamic Resource Caching in the IoT Application Layer for Smart Cities". *IEEE Internet of Things Journal*, Vol. 5, No. 5, P. 6060-6071. <https://doi.org/10.1109/JIOT.2017.2764418>

Received (Надійшла) 30.10.2025

Accepted for publication (Прийнята до друку) 28.01.2026

Publication date (Дата публікації) 27.02.2026

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

- Єрошенко Ольга Артурівна** – доктор філософії, доцент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Yeroshenko Olha – PhD, Associate Professor of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: olha.yeroshenko@nure.ua; ORCID Author ID: <https://orcid.org/0000-0001-6221-7158>;
Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57808290700>.
- Піняєв Євгеній Володимирович** – здобувач кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Piniaiev Yevhenii – student, Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: yevhenii.piniaiev@nure.ua; ORCID Author ID: <https://orcid.org/0009-0003-6489-7240>.
- Пивоварова Дар'я Ігорівна** – асистент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Pyvovarova Daria – Assistant of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: daria.pyvovarova@nure.ua; ORCID Author ID: <https://orcid.org/0000-0002-7251-994X>;
Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57216300264>.
- Кравченко Павло Олександрович** – кандидат технічних наук, старший викладач кафедри ЕОМ, Харківського національного університету радіоелектроніки, Харків, Україна;
Pavlo Kravchenko – PhD, Senior Lecturer of the Department of electronic computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: pavlo.kravchenko@nure.ua; ORCID Author ID: <https://orcid.org/0000-0002-0456-3295>;
Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57216151371>.

Method for optimizing publish-subscribe architectures for intelligent notification under resource-constrained IoT conditions

Olha Yeroshenko, Yevhenii Piniaiev, Daria Pyvovarova, Pavlo Kravchenko

Abstract. The **subject** of the study is methods for optimizing publish–subscribe (pub/sub) architectures for Internet of Things (IoT) systems operating under limited resources, aimed at improving the efficiency of intelligent notification within critical infrastructure and smart environments. The **purpose** of the work is to develop a comprehensive method for optimizing pub/sub architectures by integrating edge computing, hybrid communication models, energy-efficient message structures, and mechanisms for ensuring mobility and privacy. The article addresses the following **tasks**: analysis of the impact of limited hardware resources and network instability on the performance of pub/sub systems; development of an approach to proactive clustering of brokers at the network edge; construction of a hybrid interaction mechanism combining pub/sub and REST communications; optimization of message structures to reduce energy consumption; and ensuring node mobility and spatial privacy. The following **methods** are used: hierarchical IoT architecture modeling (Device–Edge–Cloud), clustering algorithms, binary message encoding, AI/ML-based analysis on gateway nodes, as well as simulation of network and workload characteristics. The **results** obtained include: a proposed multi-level architecture with Edge clustering and an intelligent web gateway that reduces latency, increases scalability, and decreases energy consumption; demonstrated effectiveness of the optimized pub/sub message structures and mobility support mechanisms in dynamic scenarios; and confirmed performance improvements based on simulation results (2021–2025). **Conclusions:** The developed method for optimizing pub/sub architectures shows that the integration of Edge clustering, hybrid communication, energy-efficient message structures, and privacy mechanisms significantly reduces latency and resource consumption in IoT systems, ensuring their resilience and suitability for real deployment in critical and scalable IoT environments.

Keywords: IoT, publish-subscribe, architecture optimization, limited resources, intelligent notification, energy efficiency, edge computing.