

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БУДІВНИЦТВА І АРХІТЕКТУРИ
КАФЕДРА КІБЕРБЕЗПЕКИ ТА КОМП'ЮТЕРНОЇ ІНЖЕНЕРІЇ

ДИПЛОМНА РОБОТА БАКАЛАВРА

Розробка алгоритму агрегації та попередньої обробки даних
на периферійних пристроях IoT з метою зменшення затримок
та навантаження на хмарну інфраструктуру

Виконала: Бондаренко С. О.
Науковий керівник: к.т.н. доцент Шабала Є.Є.

Мета роботи

Розробка алгоритму агрегації та попередньої обробки даних на периферійних IoT-пристроях з метою зменшення затримок та навантаження на хмарну інфраструктуру.

Об'єкт дослідження

Процес передачі та обробки даних у розподілених IoT-системах з використанням парадигми Edge Computing.

Предмет дослідження

Алгоритми агрегації та попередньої обробки даних на периферійних IoT-вузлах.

Актуальність теми

До 2030 р. понад 25 млрд IoT-пристроїв генеруватимуть петабайти даних. Обмежена пропускна здатність каналів та затримки вимагають обробки на периферії.

1. Аналіз архітектури IoT-систем та підходів до Edge Computing
2. Формалізація задачі дослідження та постановка критеріїв оцінювання
3. Розробка математичної моделі IoT-середовища (датчики, канали, edge-вузол)
4. Розробка алгоритму агрегації з ковзним вікном та адаптивним порогом
5. Реалізація симуляційного середовища (три сценарії навантаження)
6. Верифікація моделі на реальних даних Intel Lab Dataset (KS-тест)
7. Порівняльний аналіз ефективності алгоритму (з агрегацією / без)

Архітектура IoT (3 рівні)

- Devices

Датчики температури, вологості, вібрації;
збір первинних даних середовища.

- Edge

Попередня обробка, агрегація,
локальна фільтрація аномалій.

- Cloud

Довготривале зберігання,
аналітика.

Переваги Edge Computing

- Зменшення латентності

Обробка поблизу джерела даних.

- Економія пропускної здатності

Передача лише агрегованих пакетів.

- Автономна робота

Незалежність від з'єднання з хмарою.

Параметри моделі IoT-середовища

Edge-вузол

CPU: 4 ядра $\times$ 1.2 GHz
RAM: 512 MB
Черга: max 1 000 пакетів
Max response time: 100 ms
Service rate: 100 пкт/с

Канал edge $\rightarrow$ cloud

Bandwidth: 1 Mbps
RTT: 50 ± 10 ms
Packet loss: 1%
Cloud service rate: 150 пкт/с
Max queue: 10 000 пкт

Датчики (Intel Lab Data)

Температура:
 $N(22.35, 6.45^2)$, 1 Гц, 64 Б

Вологість:
 $Beta(37.2, 55.0) \times 100$, 0.5 Гц, 64 Б

Вібрація:
 $LogNormal(0, 0.5)$, 10 Гц, 128 Б

1. Вхід даних

Сенсор → локальний канал → Edge-вузол

2. Додавання до вікна

`W[sensor_id].append(x)`

Оновлення μ , σ вікна

3. Перевірка аномалії

$|x - \mu| > \alpha \cdot \sigma$?

→ ТАК: негайна передача anomaly-пакету

→ НІ: перехід до кроку 4

4. Перевірка розміру вікна

$|W| = N$ або $t > t_{\max}$?

→ ТАК: агрегований пакет (mean, std, count)

→ НІ: очікування нових даних

5. Надсилання до хмари

Пакет → канал edge→cloud

```
=== Scenario: stress - Peak load, 20% anomalies, 5% packet loss, sensor drift ===
Sensors: 50 total (25 temp, 15 hum, 10 vib)
Generated 79,500 readings
Dropped: 1602 (2.0%)
Avg processing: 0.0392 ms
Avg RTT: 50.91 ms
Avg queuing delay: 113.98 ms
Avg total cloud latency: 164.90 ms
Cloud utilization: 86.6%
CPU utilization: 2.1%
RAM usage: 8.4289 MB
```

```
=== Aggregated scenario: stress | N=10, a=3.0 ===
Generated 79,500 readings
Total readings:      78,689
Transmitted packets: 6,677
- aggregated:        50
- anomaly:           6522
- watchdog:          105
Compression ratio CR: 11.79
Dropped (local):     811
Avg processing:      0.0381 ms
Avg RTT:              50.88 ms
Avg queuing delay:   8.99 ms
Avg total cloud latency: 59.86 ms
Cloud utilization:    7.3%
CPU utilization:      0.17%
RAM usage:            0.6877 MB
```

Параметри алгоритму:

$N = 10$

(розмір ковзного вікна)

$\alpha = 3.0$

(поріг аномалій у σ)

$t_{\max} = 60$ с

(watchdog-таймер)

Типи пакетів:

- aggregated
- anomaly
- watchdog

Три сценарії навантаження

Baseline:

10 сенсорів, 1 Гц
0% аномалій, Wi-Fi

Medium:

30 сенсорів, 1–10 Гц
5% аномалій, BLE

Stress:

50 сенсорів, 10–100 Гц
20% аномалій, Zigbee

Тривалість: 30 с (baseline/medium), 15 с (stress)

Верифікація (Intel Lab Dataset)

Датасет: 54 реальні сенсори,
~413 000 показань

KS-тест температура:

$p = 0.13 \checkmark (p > 0.05)$
 $\Delta_{\text{mean}} = 0.1\%$, $\Delta_{\text{std}} = 0.4\%$

KS-тест вологість:

$p = 0.46 \checkmark (p > 0.05)$
 $\Delta_{\text{mean}} = 0.0\%$, $\Delta_{\text{std}} = 0.5\%$

Висновок: модель статистично
відповідає реальним даним.

Baseline (10 девайсів)

Зчитувань: 15 900
Dropped: 179 (1.1%)

Proc time: 0.039 ms
RTT: 50.82 ms
Queuing: 12.7 ms
Total: 63.7 ms

Cloud util: 17.4%

Medium (30 девайсів)

Зчитувань: 42 000
Dropped: 461 (1.1%)

Proc time: 0.038 ms
RTT: 50.86 ms
Queuing: ~45 ms
Total: ~96 ms

Cloud util: ~48%

Stress (50 девайсів)

Зчитувань: 79 500
Dropped: 1 600 (2.0%)

Proc time: 0.039 ms
RTT: 50.87 ms
Queuing: 114.7 ms
Total: 165.6 ms

Cloud util: 86.6%

Без агрегації при stress-навантаженні: queuing delay 114.7 ms, cloud utilization 86.6% — система на межі насичення

Stress-сценарій (50 сенсорів, 20% аномалій)

Зчитувань: 79 500 → Передано: 7 373 (CR = 10.67×) · Агрег. пакетів: 50 · Anomaly: 7 215 · Watchdog: 108

RTT та Queuing delay

RTT: 50.9 ms
Queuing delay: 8.63 ms
Total latency: 59.5 ms

Cloud utilization

Без агрегації: 86.6%
З агрегацією: 8.0%
Зменшення: -78.6 п.п.

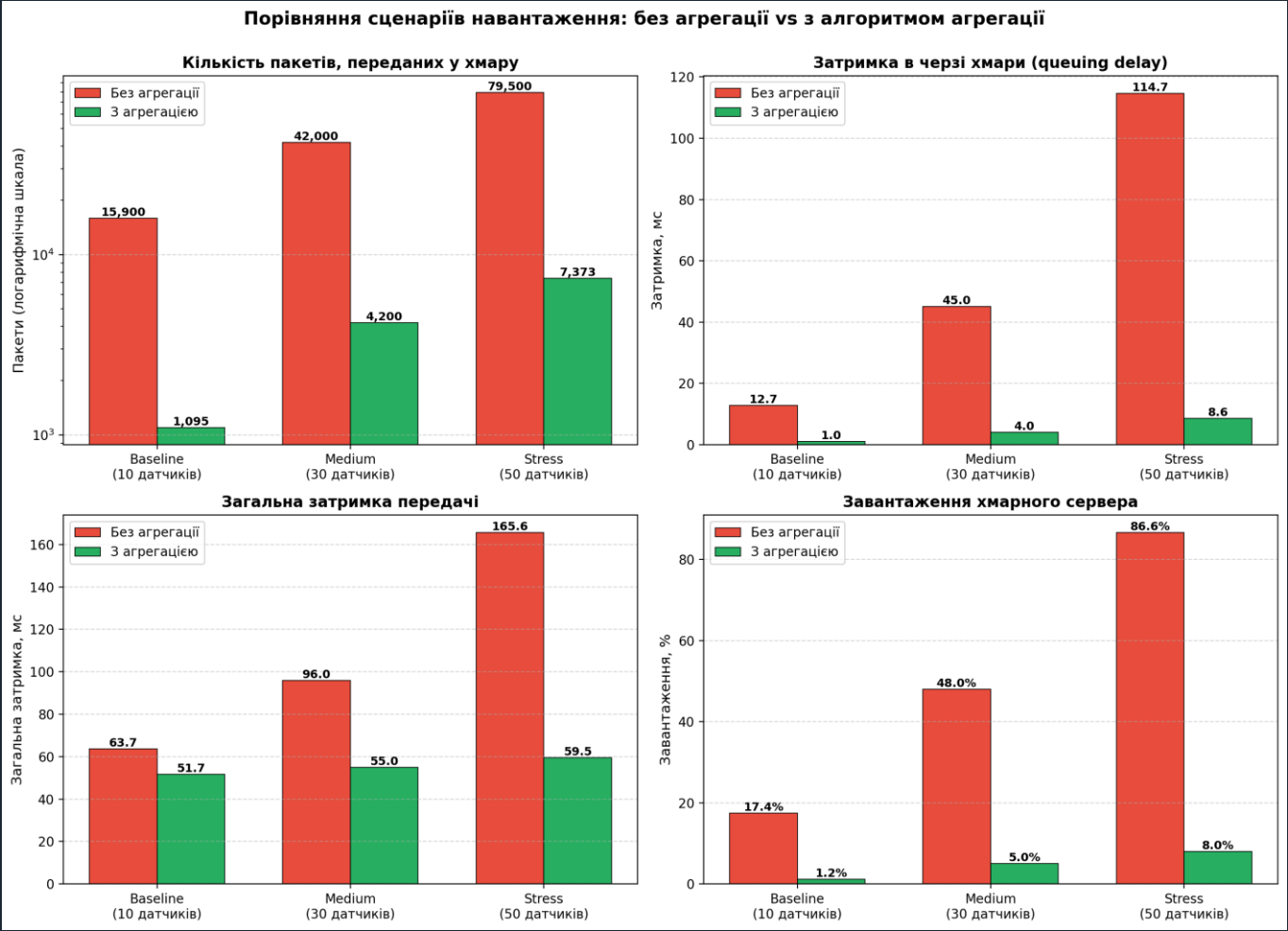
Compression Ratio

Baseline CR: 14.51×
Medium CR: ~10×
Stress CR: 10.67×
90.7% менше трафіку

Baseline з агрегацією

CR: 14.51×
Queuing: 1.0 ms
Total: 51.7 ms
Cloud util: 1.2%

Порівняльний аналіз: з агрегацією та без



1.

Розроблено алгоритм агрегації з ковзним вікном ($N=10$) та адаптивним порогом виявлення аномалій ($\alpha=3.0$), що реалізує три типи пакетів: aggregated, anomaly, watchdog.

2.

Верифіковано симуляційну модель на реальних даних Intel Lab Dataset: KS-тест $p > 0.05$ для температури та вологості, $\Delta\text{mean} < 0.1\%$.

3.

Алгоритм забезпечує коефіцієнт стиснення CR до $14.51\times$ (baseline) та $10.67\times$ (stress) — понад 90% зменшення трафіку до хмари.

4.

При stress-навантаженні queuing delay знижується з 114.7 ms до 8.6 ms (у 13 разів), cloud utilization — з 86.6% до 8.0%.

5.

Аномальні значення (до 20%) передаються негайно, що гарантує своєчасне реагування хмарної платформи на критичні події.

ДЯКУЮ ЗА УВАГУ!

Бондаренко С. О.
КНУБА, 2026