**ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ**

НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Кафедра інженерії програмного забезпечення автоматизованих систем

# Пояснювальна записка

до магістерської роботи
на ступінь вищої освіти магістр

на тему: «**ДОСЛІДЖЕННЯ СТРУКТУР АГРЕГАЦІЇ ДАНИХ ДЛЯ МЕРЕЖ IoT З МЕТОЮ ПІДВИЩЕННЯ ЇХ ЕФЕКТИВНОСТІ»**

Виконав: студент 6 курсу, групи ІСДМ-61
спеціальності 126 Інформаційні системи та технології
освітня програма «Інформаційні системи та технології»
(шифр і назва спеціальності)
Постельников В.М. . .
(прізвище та ініціали)
Керівник Срібна І.М. . . .
(прізвище та ініціали)
Рецензент _____
(прізвище та ініціали)
Нормоконтроль _Чорна В.М.__
(прізвище та ініціали)

Київ – 2021

**НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

Кафедра Інженерії програмного забезпечення автоматизованих систем
Ступінь вищої освіти - «Магістр»
Спеціальність підготовки 126 Інформаційні системи та технології
Освітня програма «Інформаційні системи та технології»

**ЗАТВЕРДЖУЮ**
Завідувач кафедри ІПЗАС
_____ К.П.Сторчак
"____" _____2021 року

# З А В Д А Н Н Я
## НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

__Постельников В'ячеслав Миколайович
(прізвище, ім'я, по батькові)

1.Тема роботи: «Дослідження структур агрегації даних для мереж IoT з метою підвищення їх ефективності»
Керівник роботи: Срібна Ірина Миколаївна, доктор технічних наук, доцент кфедри ІПЗАС .
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені наказом вищого навчального закладу від _____року №
2. Строк подання студентом роботи _____
3. Вхідні дані до роботи :
1. Агрегація даних в IoT
2. Технології туманних, хмарних обчислень та машинние навчання
3. Аналітика та архітектура IoT та схеми маршрутизації даних

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити):
1. Розкриття поняття агрегації данних в IoT
2. Дослідження структур агрегації Інтернет речей
3. Приклади досліджень сценаріїв агрегаціх даних в IoT

5. Перелік графічного матеріалу
1.__
2. Перелік можливостей та проблем агрегації даних
3. Перегляд сценарію агрегації даних IoT

6. Дата видачі завдання _____

# КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів магістерської роботи | Строк виконання етапів роботи | Примітка |
|---|---|---|---|
| 1 | Підбір науково-технічної літератури | | |
| 2 | Вивчення матеріалів для подальшої взаємодії з ними | | |
| 3 | Огляд протоколів передачі даних | | |
| 4 | Визначення технічного завдання | | |
| 5 | Розробка сценарію взаємодії | | |
| 6 | Розробка бібліотеки | | |
| 7 | Вступ, висновки, реферат | | |
| 8 | Розробка демонстраційних матеріалів | | |
| 9 | Попередній захист роботи | | |

Студент_____ Постельников В.М..
( підпис ) (прізвище та ініціали)

Керівник роботи _____ Срібна І.М..
(підпис ) (прізвище та ініціали)

# РЕФЕРАТ

Текстова частина магістерської роботи складається з 83 сторінок, 27 рисунків, 3 таблиць, 23 джерел.

ІНТЕРНЕТ РЕЧЕЙ, АГРЕГАЦІЯ ДАНИХ, IOT-АНАЛІТИКА, D2D, PCA, МАШИННЕ НАВЧАННЯ, ОПТИМІЗАЦІЯ, ФЕДЕРАТИВНА ФІЛЬТРАЦІЯ

*Об'єкт дослідження:* процес дослідження структур агрегації данних в IoT-мережах.

*Предмет дослідження:* архітектури аланітики даних IoT, сценарії агрегації даних.

*Мета роботи:* дослідження ефективних механізмів агрегації даних для масивних мереж IoT у різних сценаріях, для підтримки належного функціонування рівня IoT.

*Методи дослідження:* методи наукового моделювання, методи дослідження інформаційних систем, методи теоретичного дослідження, евристичний метод.

У магістерській роботі запропоновано чотиришарову архітектуру аналітики даних, спрямовану на вирішення відкритих проблем, використовуючи агрегацію даних.

Було представлено три незалежні нові підходи для різних сценаріїв агрегації даних.

Було вирішено декілька проблем агрегації даних IoT, також було представлено імітаційні дослідження на підтвердження ефективності цих пропозицій.

Було представлено нову систему федеративної фільтрації для пристроїв IoMT.

# ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

| | |
|---|---|
| 3GPP | Проект партнерства третього покоління |
| ARIMA | Авторегресивний інтегрований фільтр |
| AR | Авторегрессія |
| AM-DR | Адаптивний метод скорочення данних |
| AOA | Кут прибуття |
| BS | Базова станція |
| BI | Бізнес інтелект |
| CQI | Індикатор якості каналу |
| CR | Діапазон зв’язку |
| CM | Член кластера |
| CH | Голова кластера |
| D2D | Зв’язок між пристроями |
| DCM | D2D матриця CQI |
| DS | Домінуючий підпростір FFF |
| FFF | Федеративна структура фільтрації |
| GPRS | Загальна служба пакетного радіозв’язку |
| GPS | Глобальна система позиціонування |
| IoT | Інтернет речей |
| IoMT | Меличний Інтернет речей |
| LMS | Найменший середній квадрат |
| MPT | Теорія збурень матриці |
| PCA | Аналіз основних компонентів |

**ЗМІСТ**

# ВСТУП

З поширенням Інтернету речей (IoT), спостерігається величезна кількість програм для аналізу даних IoT, таких як "розумні міста", моніторинг забрудненного повітря, промислових ланцюжків, поставок. Аналітика IoT може бути визначена як процес контролю та оптимізації прийняття рішень у режимі реального часу шляхом аналізу величезних масивів даних датчиків. Для правильного функціонування аналітики даних потрібні високоякісні дані, схеми маршрутизації з низькою затримкою, висока енергоефективність та розумна конфіденційність. Враховуючи вимоги, типова автономна система аналізу даних не справляється з поставленим завданням у практичному сценарії. Більше того, обмеженість ресурсів у мережах IoT ще більше погіршує ситуацію. Щоб забезпечити стійке рішення, була введена система агрегації даних перед рівнем аналізу даних, для підвищення загального рівня ефективності системи. Рівень агрегації даних відповідає за ефективну маршрутизацію та попередню обробку даних.

Метою даної магістерської роботи є розробка ефективних механізмів агрегації даних для масивних мереж IoT у різних сценаріях для підтримки належного функціонування рівня IoT аналітичного шару.

У цій магістерській роботі представлені три незалежні нові підходи для різних сценаріїв, щоб вирішити декілька відкритих проблем. Перший підхід фокусується на енергоефективну маршрутизацію; розглядається протокол кластеризації, заснований на зв'язку між пристроями для стаціонарних та мобільних вузлів IoT. Другий підхід зосереджений на обробці невизначених необроблених даних IoT; представлена схема агрегації даних IoT для покращення якості необроблених даних IoT. Зрештою, третій підхід фокусується на втраті потужності в результаті накладних витрат на зв'язок та проблем конфіденційності для медичних IoT-пристроїв (IoMT); описує механізм агрегації даних на основі прогнозів для масивних пристроїв IoMT.

Об’єктом дослідження являється процес дослідження структур агрегації

данних в IoT-мережах.

Предметом дослідження є архітектури аланітики даних IoT, сценарії агрегації даних.

В результаті повинні бути розглянуті три важливі проблеми, а саме обмеженість ресурсів пристроїв IoT, проблема достовірності вкрай невизначених даних IoT, мережна затримка та проблеми конфіденційності даних, продемонстровані сценарії агрегації даних, та архітектури аналітики даних.

В роботі були вирішені питання невизначенності, достовірності, конфеденційності та аналітики даних IoT, та використано сценарії використання агрегації даних, аналітичні платформи, федеративну фільтрацію.

Наукова новизна полягає в визначенні перспектив розвитку застосування структур агрегації даних в IoT.

Практична значущість полягає в передачі данних без розривів та затримок, збільшенні підсумкової пропускної здатності сигналу, скорочення на певний відсоток комунікаційних накладних видатків та збільшення рівня конфіденційності інформації.

# 1 РОЗГЛЯД ТАКСОНОМІЇ АРХІТЕКТУРИ ТА АНАЛІТИКИ МАСИВНИХ ДАНИХ IoT

## 1.1 Таксономія масивних даних IoT

Перший крок до створення будь-якої інфраструктури на основі IoT повинен починатися з визначення властивостей потужних даних IoT. На підставі емпіричних спостережень дані IoT класифікуються на дві категорії: дані генерації та якість даних. Наочне уявлення загальної таксономії представлено рис. 1.1.

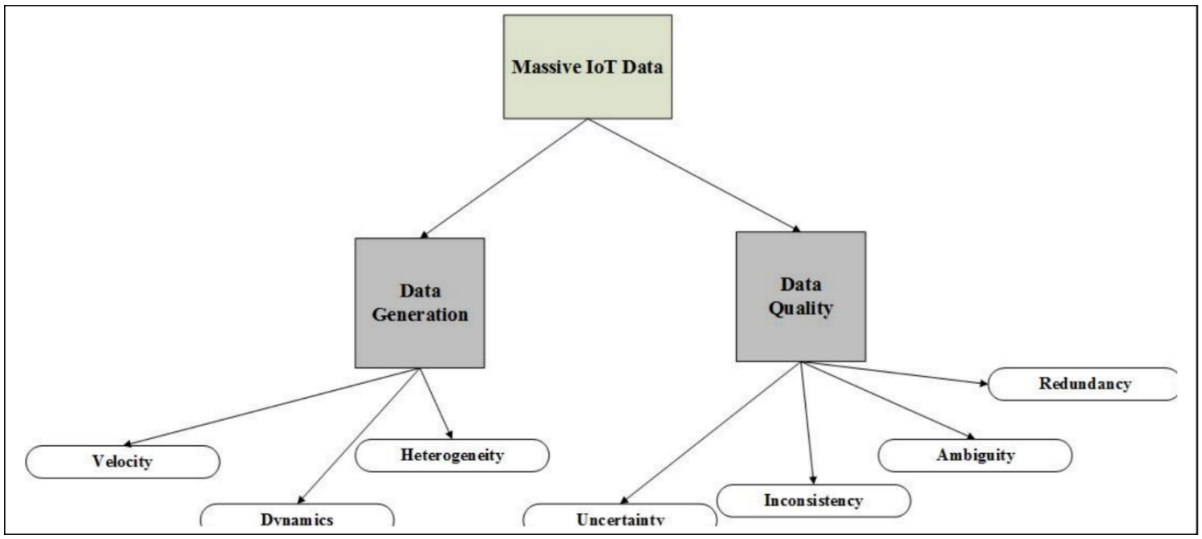



Рисунок 1.1 - Таксономія масивних даних IoT

Генерація даних:

- Швидкість: Дані IoT генеруються за допомогою різних пристроїв і всі пристрої генерують дані з різною частотою. Деякі датчики, такі як акселерометри в складних транспортних засобах генерують 10 000 показань кожну секунду. Інші пристрої, такі як датчики моніторингу води, встановлені в каналах, греблях та водосховищах, сканують довкілля зі швидкістю 10 показань на годину. У випадках коли швидкість генерації даних дуже висока, дані, що

надходять, можуть перевантажити IoT-пристрої, у той час як при низькій швидкості сканування IoT-пристроїв можуть бути втрачені важливі закономірності даних.

- Динаміка: Дані IoT можуть бути дуже динамічними за своєю природою залежно від джерела. Динамічне джерело даних генерує показання у різних місцях та у різний час. Джерело подорожує різними середовищами і виробляє дані в режимі реального часу. Більше важливо те, що джерело може переміщатися не добре пов'язаним шляхом, і в результаті переривчастого зв'язку дані IoT страждають від багатьох порушень.
- Гетерогенність: Парадигма IoT має потенціал підключення всього до Інтернету, це "всього" включає ювелірні вироби, взуття, автомобілі, класні кімнати і т.д. Тому дані від різних IoT-пристроїв мають різні формати, модальність. та розміри.

Кількість даних:

- Невизначеність: Невизначеність даних IoT пов'язана з тим, що необроблені дані датчиків IoT не заслуговують на довіру і безумовно не підходять для використання в системах прийняття рішень. Невизначеність у необроблених даних IoT обумовлена недостатньою точністю в IoT пристроях, що призводить до відсутності значень, викидів та неправильних значень. Шум також є однією з основних невизначеностей.
- Надмірність: Надмірність даних IoT відноситься до того факту, що існує декілька копій тих самих даних у наборі даних датчиків IoT. Надмірність потребує великих витрат обчислювальної потужності та зберігання, що знижує загальну ефективність системи. Надмірність може виникати в різних сценаріях, наприклад, коли декілька IoT-пристроїв розгорнуті по сусідству для спостереження за певним явищем.

- Неоднозначність: Набір даних IoT може сприйматися по-різному. залежно від сценарію та вимог споживача. Правильна інтерпретація даних є необхідною умовою підвищення можливості їх повторного використання. В деяких випадках, коли існує декілька можливих варіантів використання конкретного набору даних, двозначність стає серйозною проблемою для правильної інтерпретації даних.
- Неузгодженість: Неузгодженість поширена в масивних даних датчиків IoT. Неузгодженість викликана поганим калібруванням або пошкодженням пристроїв IoT, через шкідливий вплив навколишнього середовища, низької кваліфікації персоналу. IoT-пристрої генерують випадкові показання, які не є точним уявленням явища, якому воно відповідає.

## 1.2 Аналітика IoT

Величезний обсяг даних або просто масив даних, що генеруються додатками IoT, має бути проаналізований, щоб отримати інформацію для ухвалення рішень. Аналітика IoT може бути визначена як процес контролю та оптимізації прийняття рішень у режимі реального часу шляхом аналізу величезних шматків великих даних. IoT-аналітика необхідна для аналізу кожного сегмента великих даних IoT, даних для отримання найважливіших закономірностей у потоці даних. Це може допомогти промисловості у профілактичному обслуговуванні обладнання та інших інфраструктур, а також допомагає впроваджувати нові бізнес-моделі, оптимізовувати операційні процеси, генерувати нові продукти та послуги для користувачів. Варто зазначити, що існує чітка різниця між IoT аналітикою та аналітикою великих даних; аналітика IoT не має справу з усіма характеристиками великих даних (5 характеристик; обсяг, швидкість, правдивість, цінність та варіативність).

Зокрема, три сектори з більшою ймовірністю будуть порушені аналітичною парадигмою IoT, що насувається.

- Розумний спосіб життя: згідно з прогнозами Гартнер, до 2022 року кожна сім'я матиме понад 500 інтелектуальних пристроїв. Інший аналогічний прогноз свідчить, що до 2020 року підключена кухня сприятиме як мінімум 15% економії в харчовій промисловості та виробництві напоїв при використанні аналітики великих даних.
- Розумний транспорт: За прогнозами Гартнер, до 2020 чверть мільйона підключених транспортних засобів сприятимуть появі нових компетенцій у галузі автомобільного та автономного водіння. Підключені транспортні засоби генеруватимуть 30 ТБ даних щодня, спілкуючись із навколишнім середовищем та іншими транспортними засобами. Ця нова парадигма створить діловий потенціал вартістю 14 мільярдів доларів США у всьому світі.
- Розумні міста: Сан-Хосе, США, націлений на покращення якості життя за допомогою моніторингу якості повітря, транспортного потоку. У Пізі, (Італія) інтелектуальне управління допомагає водіям знайти вільне місце для паркування.

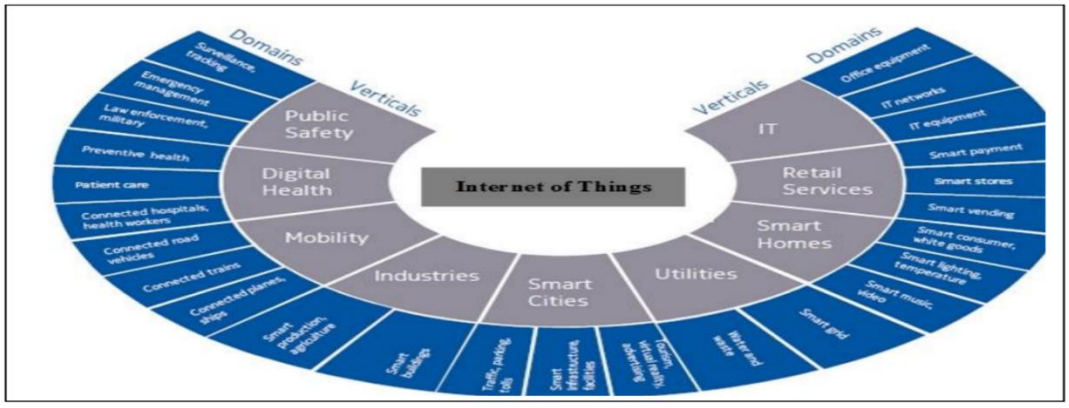



Рисунок 1.2 - Карта доменів та вертикалей IoT

IoT-аналітика довела свою цінність для суспільства, і нині вона привертає значну увагу як академічних, так і промислових кіл. Зростаючий інтерес до

аналітики IoT вимагає від зацікавлених сторін чіткого розуміння аналітичних підходів, будівельних блоків, технічних вимог та відкритих проблем.

### 1.3 Архітектура IoT-аналітики

Сучасна система аналітики IoT націлена на вилучення інформації з величезної кількості даних IoT, та виконання завдань щодо прийняття рішень на основі отриманої інформації. Всю аналітичну мережу IoT можна поділити на три основні частини, як показано на рис. 1.3. Мережі датчиків вимірюють навколишнє середовище та генерують показання. Ці мережі також відповідають за обмін даними з пристроєм, що маршрутизує, або шлюзом оптимально з погляду енергоспоживання, накладних витрат на зв'язок та найкоротшого можливого шляху. Крім того, мережа маршрутизації може бути шлюзовим пристроєм або IoT-пристроєм. Виступаючи як керівник кластера для кластера вузлів IoT, що доставляє агреговані дані на аналітичний сервер.

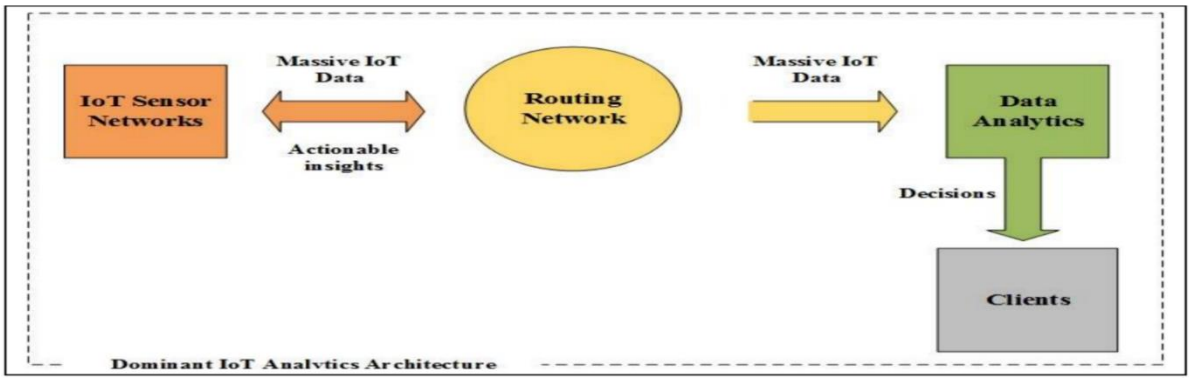



Рисунок 1.3 - Домінуючі архітектури IoT-аналітики

Як правило, завдання аналізу та прийняття рішень у масивній аналітичній платформі IoT виконуються централізовано усередині хмарних серверів. Централізована аналітична система IoT вимагає всі дані відразу, щоб робити висновки.

## 1.4 Таксономія IoT-аналітики

Вилучення бізнес-цінності з необроблених даних датчиків IoT – завдання не з простих. Для того щоб використовувати потрібні дані в потрібний момент для задоволення потреб клієнта дуже важливо класифікувати аналітику IoT на основі сценаріїв застосування та відповідних випадків використання. Всеосяжна таксономія аналітики IoT представлена на рис. 1.4. IoT аналітику можна розділити на дві категорії: історичний аналіз та проактивний аналіз. Історична аналітика забезпечує всебічну візуальну інтерпретацію даних. Зазвичай вона ґрунтується на традиційних методах видобутку даних. Історичні дані можна розділити на описову аналітику та діагностичну аналітику. описова аналітика генерує статистику та візуальну інтерпретацію даних, і вона зазвичай використовується в бізнесі для оцінки своїх продуктів і бази користувача. Діагностична аналітика виконує виявлення аномалій для перевірки несправностей в устаткуванні та забезпечує сигнал тривоги у разі виникнення будь-якої аномалії.

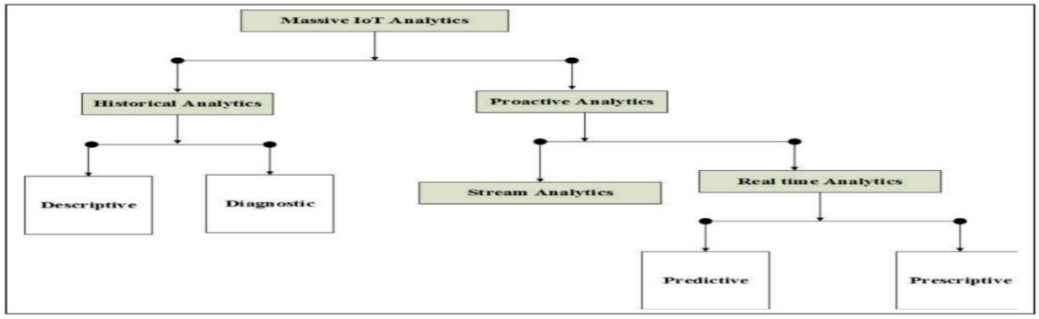



Рисунок 1.4 - Різні категорії IoT-аналітики на основі сценаріїв використання

З іншого боку проактивний аналіз повільно, але неухильно розвивається як нова тенденція для того, щоб генерувати та полегшувати отримання корисної інформації з величезного набору даних IoT. Ця категорія далі класифікується на потокову аналітику та аналітику в реальному часі. У разі потокової аналітики, в яку входять дані часових рядів аналізуються на основі партій чи потоків; ідея полягає в

тому, щоб знайти важливу закономірність усередині партії чи потоку даних. Аналітика у режимі реального часу потрібниа для отримання оптимального або неоптимального результату шляхом аналізу частини даних (не всього набору даних) протягом мінімального часу. Аналітика в реальному часі стає все популярнішою завдяки її широкому спектру застосування у громадянському суспільстві та промисловості.

### 1.5 Відкриті проблеми, пов'язані з рівнем аналітики даних

Проактивне прийняття рішень швидко стає основним напрямом, оскільки може забезпечити динамічні додатки для промисловості та сучасного суспільства. Однак, незважаючи на всю популярність аналітики даних, аналітика IoT в реальному часі все ще знаходиться в зародковому стані. IoT-аналітика у реальному часі стикається з низкою проблем і для належного функціонування системи вимагає високоякісних даних для прийняття рішень, ефективних схем маршрутизації з низькою затримкою, високу енергоефективність, оптимальні витрати на зв'язок та розумну конфіденційність. Більше того, обмеженість ресурсів мереж IoT ще більше посилює ситуацію. Домінуюча архітектура аналітики даних, як показано на рис. 1.3, не є компетентна для підтримки проактивних аналітичних програм, що вимагають аналізу в реальному часі. Очевидною проблемою домінуючох парадигми аналітики даних є те, що аналітичний рівень є автономний. Хоча ця парадигма добре працює для аналізу історичних даних, сучасний проактивний аналіз вимагає ухвалення рішень у режимі реального часу, проблемою є нестача часу, бюджет. Щоб обговорити та оцінити серйозність відкритих проблем, пов'язаних з кожним шаром домінуючої архітектури аналітики IoT (рис. 1.3), загальні недоліки класифікуються за трьома категоріями: проблеми пов'язані з сенсорними мережами, з мережею маршрутизації, та з аналітикою даних.

Проблема достовірності даних є одним із основних для будь-якої системи аналізу даних. Необроблені дані датчиків IoT мають високий рівень невизначеності. Невизначеність викликана наявністю викидів, відсутніх значень, надмірності, неточності та необ'єктивних показань. Викиди – це показання набору даних, які демонструють повне відхилення від встановленого зразка партії даних. Відсутні значення негативно впливають на алгоритм машинного навчання, який використовується аналітичними платформами для прийняття рішень, і в більшості випадків це призводить до випадкових результатів. Надмірність становить загрозу для ефективності системи. Як правило, система, перевантажена надмірними даними, втрачає значну частину своїх обчислювальних ресурсів на обробку даних, що дублюються. Отже, надмірність знижує загальну ефективність системи. Неточність і упередженість дуже важко виявити в масивному наборі даних IoT, оскільки вони залежать від контексту, і для виявлення аномалії система або людина повинні мати значний досвід у цій галузі. Загалом проблема достовірності даних є значною загрозою для всіх додатків, що використовують аналітичні платформи IoT. Зазвичай в аналітиці IoT/ухвалення рішень відбувається всередині віддалених серверів хмар. або у хмарних центрах обробки даних. У хмарні сервери вбудовано повністю централізовані алгоритми прийняття рішень які вимагають доступ до всього набору даних за один раз. Тому такі системи вимагають безперервної передачі всіх масивних даних, що генеруються мережами IoT у хмарні сервери, що викликає ряд критичних проблем, що стосуються масштабованості, мережевої затримки, енергоефективності мережі та конфіденційності.

IoT-пристрої окремо генерують невеликі пакети даних дуже часто враховуючи розмір мережі загалом, невеликі пакети даних разом стають величезними за обсягом. Більше того, величезний обсяг даних IoT доповнюється дуже високою швидкістю. генерації, що перетворює їх на потокові сенсорні дані часових рядів. Таким чином, масштабованість є найважливішою вимогою для аналітичних програм IoT в реальному часі. У практичних сценаріях потокові масивні дані IoT можуть

придушити/перевантажити будь-яку систему. В традиційних хмарних обчисленнях парадигма довела свою ефективність в аналітиці історичних даних понад десять років тому; однак вона не підходить для підтримки аналітичних програм IoT в реальному часі, таких як промислові та медичні програми, IoT-аналітика та аналітика в охороні здоров'я.

Проблеми мережевої затримки виникають через низку різних сценаріїв в аналітичній мережевій архітектурі. Мережеві затримки можуть виникати через багаторазові далекі передачі даних, розриву міжрівневого зв'язку в багаторівневій аналітичній архітектурі, переривчастий зв'язок та неефективні підпрограми зв'язку для ухвалення рішень. Основною проблемою, пов'язаною із затримками в мережі, особливо в приміських та сільських районах, є передача даних на великі відстані до віддалених хмарних серверів у великих містах. Наприклад, масивна мережа IoT із 100 000 датчиків, розгорнута на 14,00 км водного шляху з Пекіна до Тяньцзіня, постійно відстежує якість води. Датчик, розгорнутий поблизу Тяньцзіня, має передавати дані до Пекіна, оскільки Хмарні сервери розташовані в Пекіні, тому система працює зі значними затримками . Одним словом, затримки мережі можуть бути згубними для проактивної аналітики IoT.

Парадигма аналітики IoT стає критично важливим фактором інтелектуальних рішень для охорони здоров'я. У сценаріях, пов'язаних із охороною здоров'я, дані дуже анонімні самі по собі, а анонімність та анонімність даних є головною турботою клієнтів. Домінуючі централізовані хмарні обчислення парадигма поєднує всі дані на публічних або приватних серверах, вразливих для кібер-атак; тому вона вкрай непридатна для таких критично важливих програм, оскільки може призвести до порушення конфіденційності.

## 1.6 Відкриті проблеми, пов'язані з сенсорними мережами IoT

Додавання напівпровідників та радіоприймачів до речей, які раніше не мали їх, різко скорочує термін служби товару. Як правило, пристрої IoT мають невеликі розміри, а батарея ще менше; через це обмеження багато IoT-пристроїв викидаються, як тільки в них закінчується заряд. Тому IoT-пристрої мають серйозні обмеження щодо ресурсів щодо потужності, обчислювальних можливостей та можливостей зберігання даних. Збільшення часу автономної роботи IoT є важливою проблемою. і цю проблему ще не вирішено. Тому при прийнятті рішень на основі машинного навчання IoT-пристроїв вимагають зовсім іншого підходу; традиційні алгоритми машинного навчання вимагають великих ресурсів, які недоступні для цих невеликих IoT пристроїв. Проблема виконання аналітики IoT в умовах сценарії з обмеженими ресурсами створює значні труднощі як для розробників вбудованого програмного забезпечення, так і для проектувальників мереж.

## 1.7 Відкриті проблеми, пов'язані з мережею маршрутизації

Кооперативні комунікації стануть важливим активом для 5G зв’язку, що насувається. Кооперативний зв'язок також відомий як зв'язок "пристрій-пристрій" (D2D). комунікація (D2D). Неефективний зв'язок D2D призводить до багатьох проблем, таких як високі втрати енергії, високі комунікаційні накладні витрати та висока втрата пакетів. Більш того, у динамічних сценаріях, де пристрої IoT нестаціонарні, мережа маршрутизації страждає від великих втрат енергії та втрати пакетів, що ще більше ускладнює проблему маршрутизації даних IoT. Тому мережі маршрутизації IoT потрібні кращі схеми, які можуть забезпечити високу продуктивність для стаціонарних та динамічних сценаріїв IoT.

Домінуюча архітектура аналітики IoT обтяжує рівень аналітики даних кількома обов'язками. Ці обов'язки включають завдання, необхідні для прийняття

рішень являються допоміжними завданнями. Однак у процесі балансування компромісу між загальною ефективністю системи та прийняттям рішень, рівень аналітики даних ставить під загрозу ефективність ухвалення рішень. Для забезпечення сталого рішення перед шаром аналізу даних вводиться шар агрегації даних, який збільшує загальну ефективність системи. Загалом рівень агрегації даних відповідає за ефективну маршрутизацію та попередню обробку даних. Він допомагає рівню аналізу даних повністю зосередитися на прийнятті рішень, виконуючи завдання, необхідні для підтримки аналітики IoT, такі як оптимальний збір схем даних, забезпечення якості даних, підтримання енергоефективного зв'язку з низькою затримкою та забезпечення безпеки ефективного зв'язку з низькою затримкою та забезпечення конфіденційності даних. Очевидно, що рівень агрегації завантажує допоміжні завдання рівня аналітики даних. Запропонована архітектура аналітики IoT допомагає рівню аналітики даних виконувати ефективне прийняття рішень, надаючи всі необхідні ресурси для підтримки спільної близької до оптимальної продуктивності на рівні системи. Запропонована архітектура включає в себе об'єднання сучасних технологій, таких як туманні обчислення, хмарні обчислення та машинне навчання.

Основною метою цієї магістерської роботи є розробка ефективних алгоритмів агрегації даних та фреймворків для масивних мереж IoT у різних сценаріях, для підтримки належного функціонування усієї аналітичної системи IoT. Для досягнення цієї мети у магістерській роботі досліджуються засновані на даних підходи до агрегації масивних даних IoT, які спираються на методи, засновані на неопуклій оптимізації, федеративній структурі навчання та машинному навчанні.

Агрегація даних залежить від операційного середовища; наприклад, динаміка агрегації даних у підземних шахтах має відмінну особливість від агрегації даних, яка лежить на поверхні землі. Для досягнення основної мети при вирішенні цього питання досліджуються різні сценарії, а саме:

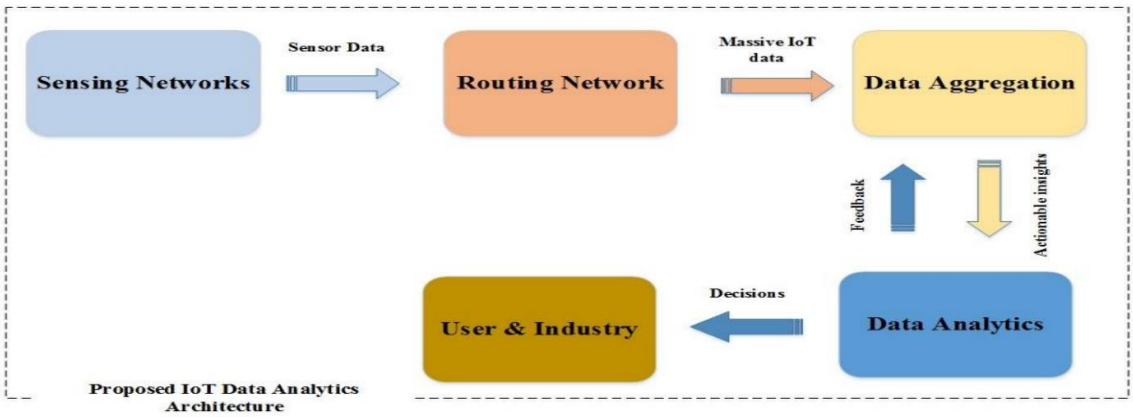



Рисунок 1.5 - Запропонована архітектура аналітики IoT

- Досліджується сценарій складної мережі доставки даних, що має стаціонарні та нестаціонарниі пристроїв IoT. Метою цієї частини дослідження є представлення та перевірка єдиної схеми агрегації на основі мобільності схеми, як заснована на D2D комунікації для стаціонарних та нестаціонарних IoT-пристроїв. Вказана схема також є дуже енергоефективною, оскільки пристрої IoT передбачають, що вони дуже обмежені у ресурсах.

- Ще один поширений сценарій, якому приділяється велика увага, – аналітична системf, яка одержує на вхід сильно спотворені потужні дані датчиків IoT. Метою ціэї частини дослідження є вирішення проблеми достовірності даних для необроблених даних IoT та покращити якість великих необроблених даних датчиків IoT.

- Досліджується дуже популярний та важливий сектор застосування IoT-аналітики, а саме "розумна" охорона здоров'я. Центральною метою цього дослідження є збільшити час автономної роботи пристрою за рахунок зниження накладних витрат на зв'язок та збереження розумної конфіденційності медичних даних.

# РОЗДІЛ 2 СХЕМА КООПЕРАТИВНОЇ АГРЕГАЦІЇ ДАНИХ НА ОСНОВІ МОБІЛЬНОСТІ

## 2.1 Алгоритм кооперативної агрегації даних IoT

Поява інтернету речей відкриває безліч нових можливостей та шляхів для досліджень для всього світу. Сьогодні IoT став улюбленою темою досліджень серед більшості провідних технологічних компаній, що входять до списку нових стартапів Forbes. Ця епоха запам'ятається сходинкою до більш пов'язаного та усвідомленого світу, де IoT застосовується від космічних кораблів до клітин тіла. Однак, щоб отримати вигоду з парадигми IoT, необхідно розробити ефективні методи маршрутизації даних, що генеруються величезною кількістю програм, додатків IoT. Співіснування різних технологій зв'язку (3G/4G/5G/Wi-Fi/GPRS) в той же час, не залишає іншого вибору, крім як знайти спільну технологію, яку можна використовувати для маршрутизації даних, технологію, яка може бути використана для маршрутизації даних за допомогою аналогів, що знаходяться поблизу. Проект «Партнерство 3-го покоління» визначив вищезгадану технологію як зв'язок між машинами та пристроями (D2D) замість MTC. Зв'язок між пристроями, та між користувальницьким обладнанням (UE), що знаходиться поблизу, підвищує спектральну та енергетичну ефективність.

У цьому розділі розглядається схема агрегації даних для стаціонарних та нестаціонарних IoT-вузлів з мінімальними ресурсами щодо потужності та обчислювальних можливостей, на основі запропонованої архітектури аналізу даних, розглянутої в розділі 1. Сенсорний рівень мережі IoT контролюється та оптимізується рівнем агрегації даних для ефективного об'єднання масивних IoT-даних. У розділіі визначено бюджет на електроенергію, динамічність пристрою та обчислювальну складність серед UE (пристроїв IoT) як основні відкриті проблеми. Щільність пристроїв UEs у певних місцях використовується у запропонованому

алгоритмі для D2D комунікації у взаємній близькості, ядром схеми є кооперативне завантаження вмісту UEs на віддалену базову станцію (BS) шляхом формування багатошляхової системи D2D. Спочатку пристрої формують кластер із розділом кластера на чолі. Голови кластера формують багато шляхів для завантаження даних на БС. Більш того, дані завантажуються тією головою кластера, яка знаходиться найближче до БС.

У цьому розділі робота зосереджена на розробці енергоефективного та менш складного у обчислювальному відношенні алгоритму кооперативної агрегації даних IoT. Нижче наведено основні результати роботи:

- Через обмежений бюджет на електроенергію використання GPS не є розумним варіантом. Тому було запропоновано новий альтернативний варіант для обчислення швидкості нестаціонарного IoT-пристрю.
- Алгоритм I є полегшеною локальною підпрограмою для формування кластера, цей евристичний підхід відрізняється високою енергоефективністю.
- Алгоритм II є енергоефективним алгоритмом для агрегації даних. агрегація даних заснована на багатошляховому міжкластерному D2D-зв'язку.
- Запропонований підхід враховує як стаціонарні, так і нестаціонарні IoT-пристрої.

## 2.2 Історія питання

Зв'язок між пристроями був вперше представлений в 3GPP у релізі 11, пізніше для зміцнення основи були опубліковані додаткові технічні деталі у релізі 12. Безконтактний зв’язок, підтримуваний D2D, отримав величезний імпульс як спосіб подолання недоліків традиційної стільникової системи. Основними перевагами D2D є: (a) низьке енергоспоживання, завдяки взаємодії пристроїв, що знаходяться в безпосередній близькості один від одного; (b) передача даних з високою швидкістю,

завдяки взаємному співробітництву між пристроями; (c) надійний зв'язок; (d) менш накладні витрати для БС та (e) гетерогенне підключення пристроїв, оскільки пристрої, що використовують різні технології, такі як Wi-Fi, LTE-A можуть бути розташовані.

D2D можна класифікувати двома різними способами, один з яких реалізується з використанням ліцензованого стільникового спектру (внутрішньосмуговий зв'язок D2D), а інша використовує неліцензований спектр (несмуговий зв'язок D2D). Внутрішньосмуговий зв'язок класифікується на дві категорії: D2D у підсмуговому режимі де використовується стільниковий зв'язок і зв'язок D2D, який використовують одні й самі ресурси одночасно, тоді як у оверлей D2D, для роботи D2D виділяються спеціальні ресурси. Позасмуговий зв'язок D2D спрямований на усунунення проблеми перешкод, але оскільки БС не бере участі у сценарії, їй необхідно покладатися на такі інтерфейси, як WiFi, Bluetooth, інфрачервоний порт. Технологія D2D була прочитана і досліджується вже досить давно, багато опублікованих дослідних статей успішно відображають загальну ідею та сучасні реалізації технології.

У зв'язку з великою кількістю пристроїв IoT, D2D є фаворитом для використання спільної агрегації даних. А. Орсіно у своїй роботі представив захоплюючий сценарій агрегації даних IoT для розумних міст; у цій роботі розглянуто широке коло проблем із докладним рішенням. Тим не менш, деякі припущення, такі як стаціонарні UE, розподіл ресурсів за круговою системою, обмежують сферу застосування рішення. Г. Раззі також запропонував відмінну альтернативну схему розподілу ресурсів, яка є практичною та легко реалізованою. Діяльність Мілітано представила спільне формування коаліції на основі соціальної довіри та близькості.

Рівень агрегації даних також може бути використаний для управління блоком ресурсів (РБс) та потужністю. Шуберт запропонував новий метод агрегації даних для пристроїв IoT він використовував агрегацію даних, як інструмент для зменшення

енергоспоживання. Загалом, дослідницьке братство в галузі бездротових технологій намагається впоратися з передачею даних IoT так само, як вони справляються із передачею даних у бездротових сенсорних мережах. Передача даних IoT має певні відмінності з бездротовою передачею даних у бездротових сенсорних мережах (БСС): (a) вміст, що генерується пристроями IoT відносно менше, ніж загальні дані WSN, (b) зазвичай очікується, що канал буде для передачі даних у WSN, але в IoT це не має великого значення, (c) в IoT передача даних має бути мінімальною, на відміну від WSN, де в цілому немає такого зобов'язання. Як підсумок, вищезазначені відповідні роботи зосереджені лише на зниження енергоспоживання під час передачі даних IoT; однак рішення, запропоноване у цьому розділі, враховує практичний компроміс. У цьому розділі розглядається практичний компроміс між обчислювальною складністю та споживанням енергії під час передачі даних IoT з використанням зв'язку D2D.

### 2.3 Модель системи

Запропонована система сформульована з використанням мережі Long Term Evolution Advanced (LTE-A) мережу, яка у разі потреби надає виділені ресурси для D2D зв'язку та система використовує внутрішньосмуговий протокол оверлейного зв'язку D2D. Для збереження простоти у цьому розділі показана демонстація пропозиції на прикладі одного осередка, як показано на рис. 2.1.

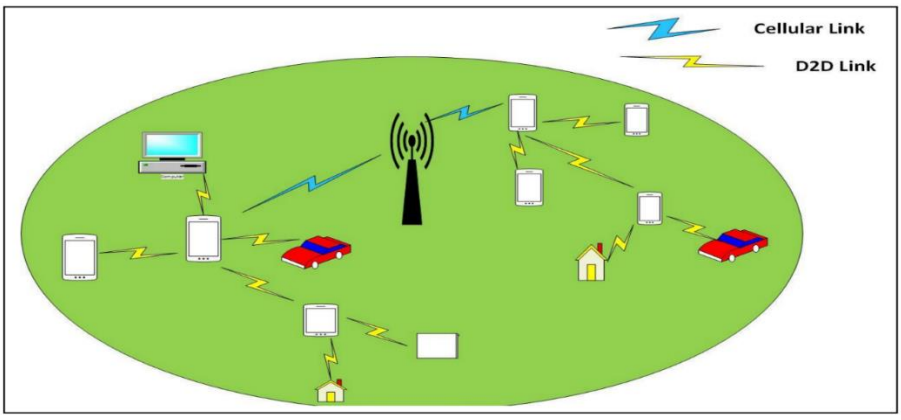



Рисунок 2.1 - Запропонована аналітична архітектура IoT

Дотримуючись стандартної процедури завантаження даних у стільникову мережу, в першу чергу БС збирає зворотний зв'язок CQI (індикатор якості каналу) у режимі стільникового зв'язку від усіх пристроїв IoT (вузлів), які бажають завантажити деякі дані. У другу чергу БС збирає значення CQI режиму D2D та формує матрицю CQI режиму D2D (DCM). Припустимо, що кожен вузол $n_i = \{n_{1,\ldots\ldots,} n_{|n_i|}\} \in N$ є частиною кластера $s_i = \{s_{1,\ldots\ldots,} s_{|s_i|}\} \in S$ у комірці, керованої БС. Як стаціонарні, так і нестаціонарні пристрої IoT (вузли) передають сигнальні повідомлення, щоб знайти сусіда з відривом 1-хоп, тут учасники використовують метод, що базується на діапазоні, для оцінки кута прибуття (AOA) та індикатора рівня сигналу (RSSI) для визначення відстані до невідомого сусіднього вузла, беремо відстань невідомого сусіднього вузла та кут по відношенню до опорного вузла, (вузол, який потребує ретрансляції). якому необхідний ретрансляційний вузол передачі даних. Спільна мобільність (розглянута у розділі 2.6) між вузлами розраховується та зберігається у кожному вузлі у вигляді матриці. Далі всі вузли вибирають відрив 1-хоп, береться сусідній вузол з урахуванням відносної мобільності між сусідніми вузлами, для передачі своїх даних. Цей процес здійснюється всіма вузлами в осередку, вузли групуються між собою в динамічні чи статичні кластери.

У цьому розділі передбачається, що кожен пристрій є інтелектуальним пристроєм із акселерометром або GPS, встановленим усередині нього. На основі принципу внутрішньосмугового накладання зв’язку «пристрій-пристрій», БС надає виділені ресурси. Більше того, використовуючи DCM (D2D CQI матриці), згаданої раніше, БС відстежує та контролює весь сценарій D2D. Зв'язок, який не є доцільним відповідно до DCM, не підключається.

У міру виконання алгоритму, БС надає ідентифікатори вузлів кожному вузлу та створює кільця вузлів з урахуванням відстані до нього. Потім кожен вузол у осередку розсилає маяки, щоб визначити місцезнаходження ретрансляційного вузла

у безпосередній близькості від нього. Релейний вузол - це IoT-пристрій, який допомагає вузлу-джерелу спільно завантажувати свої дані на БС, використовуючи D2D-зв'язок. Тут пристрої використовують метод дальності заснований на методі оцінки кута прибуття (AOA) і рівня прийнятого сигналу (RSSI) для визначення місцезнаходження індикатора (RSSI), для визначення відстані до невідомого сусіднього вузла та кута щодо опорного вузла, якому необхідний ретрансляційний вузол передачі даних. Кожен вузол обчислює динамічну матрицю N x N, яка схожа на матрицю DCM, проте замість значення CQI містить співвідносну мобільність опорного вузла, по відношенню до всіх інших вузлів. Ця матриця надалі відіграє важливу роль у пошуку гідного сусіднього вузла передачі даних. У цьому розділі також передбачається, що всі UE є інтелектуальними пристроями і можуть надсилати або отримувати дані автономно. Перед тим, як перейти до основної пропозиції, необхідно визначити деякі ключові поняття. які використовуються у всій пропозиції.

### 2.4 Енергетична ефективність

У цьому розділі зосереджені всі зусилля щодо оптимізації енергоефективності з урахуванням обчислювальної складності, що в даному випадку є компромісом між цими двома параметрами в даному реченні. Більше того, кількість енергії, яка буде використана для даної пропозиції, ґрунтується на кількість пакетів, що завантажуються кожним вузлом. Оскільки енергетична ефективність ( $\eta$ ) досліджується протягом тривалого часу, у реченні використовується стандартне рівняння з літератури. Енергоефективність ( $\eta$ ) системи, яка має N кількість користувачів, може бути обчислена як:

$$\eta = \sum_{N} \frac{d_N}{E_N \cdot r_N \cdot TTI} \quad (2.1)$$

де $d_N$ загальний обсяг завантажуваних даних (у бітах), $E_N$ це середня потужність споживана для доставки одного пакета даних, $r_N$ загальна кількість пакетів даних, які повинні бути завантажені всіма вузлами, а TTI - інтервал часу передачі, фіксований для всіх пакетів.

### 2.5 Акселерометр

Акселерометр це пристрій, який вимірює прискорення і є скрізь виявленим у пристроях IoT по всьому світу. Триосьовий акселерометр в IoT-пристроях забезпечує координати X, Y, Z значення, які використовуються для виміру положення та прискорення пристрою. Обертання, напрямок та положення вимірюються за допомогою датчиків гіроскопу. Ця пропозиція цікавити тільки лінійне прискорення пристрою. Показання акселерометра фільтруються для отримання корисних результатів. Функцію, яка облислює швидкість, показано нижче.

$$v(t) = v(0) + \sum a \times \delta t \tag{2.2}$$

Де *v(t)* - миттєва швидкість у момент часу *t*, *v(0)* - початкова швидкість, а - миттєве прискорення в момент часу і $\delta t$ цей час, витрачений для прискорення. Тут цікавить пропозиція - горизонтальної складової швидкості вузлів. $n_x$

$$V_x(t) \text{ c } \alpha\theta s \tag{2.3}$$

Де θ кут швидкості щодо іншого вузла. Пропозиція спрямована на зниження енергоспоживання, що підвищує час автономної роботи пристрою. Тому на чолі виступає використання акселерометрів як спідометрів, для обчислення швидкості вузлів IoT з обмеженими ресурсами. Ця концепція дозволяє економити енергію та обчислення, які в іншому випадку використовуються GPS для отримання координат

розташування та подальшого обчислення швидкості на основі руху. Показання акселерометра можуть бути неточними, але оскільки пропозиція стосується відносної швидкості, воно підходить для цієї мети.

### 2.6 Спільна відносна мобільність

По відносній мобільності є багато літератури, і дослідницька компанія використовувала термін відносна мобільність у різних сценаріях. Більше того, у запропонованій схемі термін «співвідносна мобільність» використовують як функцію, тобто перехресна кореляція відносної мобільності між двома вузлами $n_x$ та $n_y$ де x,y ∈ і $n_x, n_y \in N$ у час t. Як уже згадувалося, для обчислення відстані між двома вузлами, пропозиція використовує RSSI та AOA методи.

Відстань між двома вузлами $n_x$ та $n_y$ де $x,y \in i$ та $n_x, n_y \in N$ на в момент часу $t$ може бути визначена як $D_{XY}(t)$ . Нормовану відстань $\bar{D}_{XY}(t)$ можна визначити як:

$$\bar{D}_{XY}(t) = {}^{D_{XY}}\!/_{CR} \tag{2.4}$$

Де CR – максимальна дальність зв'язку маяка. Горизонтальна швидкість вузлів може бути визначена як:

$$V_x(t)\cos\theta \tag{2.5}$$

$$V_y(t)\cos\theta \tag{2.6}$$

Тут θ кут нахилу вектора швидкості, крім того, орієнтація руху щодо опорного вузла може бути виміряна θ.

$$\bar{V}_x(t) = \left|V_x(t)\cos\theta\right| \tag{2.7}$$

Де $\bar{V}_x(t)$ це тільки величина горизонтальної складової швидкості вузла $n_X$ .

$$RM_X(t) = \alpha \bar{D}_{XY}(t) + \beta \bar{V}_X(t) \quad (2.8)$$

Де $RM_X(t)$ відносна мобільність вузла $n_X$. Значення α і β є схильні до змін залежно від сценаріїв руху. Використовуючи рівняння (2.4), (2.7) та (2.8), можна визначити співвідносну мобільність між вузлом X та вузлом Y, вона може бути визначена як перехресна кореляція між відносною мобільністю двох вузлів, як показано нижче:

$$cr_{XY}(t) = Correlation(RM_X(t), RM_Y(t)) \quad (2.9)$$

Де $cr_{XY}(t)$ спільна мобільність двох вузлів $n_X$ *і* $n_Y$ де x,y ∈ I та $n_x$, $n_y \in N$ у момент часу t, а n – загальна кількість вузлів.

## 2.7 Запропонована схема формування кластера

Формування кластера можна узагальнити таким способом:

- БС випадково надає унікальний ідентифікатор вузла всім вузлам.
- БС (розташована в центрі осередку) групує вузли IoT у декілька кругових кільця на основі радіальної відстані від центру. БС також надає кільцеві ідентифікатори кожному IoT-вузлу відповідно до його положення.
- Вузли, що беруть участь (опорні вузли) передають рекламні повідомлення у межах свого радіального діапазону передачі. Одночасно сусідні вузли одержують повідомлення в радіальному діапазоні, обчислюють співвідносну мобільність із вузлом-рекламодавцем на основі рівняння 2.9.
- Сусідний вузол обчислює компетентність опорного вузла (вузол-рекламодавець) для приєднання до кластера на основі критерію співвідносної мобільності. сусідній вузол зберігає це значення компетентності кожного вузла-рекламодавця в масив, відомий як масив компетенцій. Цей масив містить значення компетентності вузлів, зацікавлених у передачі даних.

- Сусідний вузол (поточний вузол) приєднується до кластера, який має найвищу співвідносну мобільність серед інших вузлів-рекламодавців, який також є найбільш компетентним вузлом ретрансляції для вузла рекламодавця. Важливим обмеженням щодо Угруповання вузлів IoT є те, що вузол може приєднатися до кластера тільки в тому випадку, якщо його голова кластера має той самий ідентифікатор кільця.

- Коли вищезгаданий процес здійснюється над великою групою вузлів, вони утворюють декілька кластерів різного розміру.

Вищезгаданий процес можна легко зрозуміти, використовуючи блок-схему, показану на рис. 2.2. У наступному розділі представлена ідея завантаження даних у БС на основі міжкластерної багатоцільової передач даних.

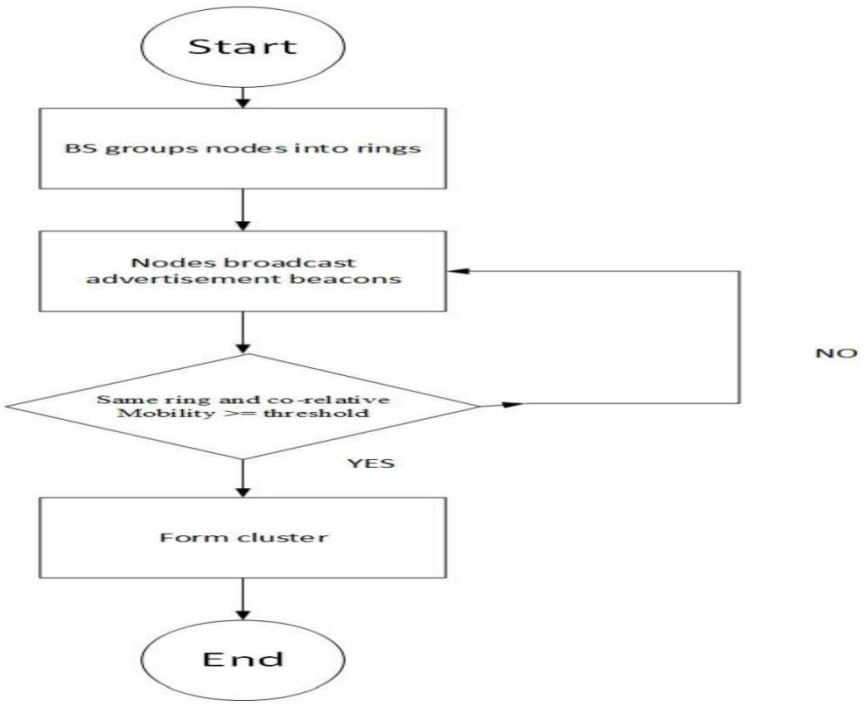



Рисунок 2.2 - Схема формування логічного потоку кластера

### 2.8 Схема завантаження даних з реквізитами

Як згадувалося, обсяг даних, генерованих пристроями IoT, відносно малий, тому це вимагає високого CQI; користуючись цією особливістю, IoT-пристрої

формують багатошляхову маршрутизацію даних від ЦМ до БС (поглинача). Більш того, тільки CH(s) бере участь у багатошляхової передачі даних, також відомої як міжкластерна передача даних. Після формування кластера всі КМ завантажують свої дані на CH, використовуючи D2D-зв'язок. CH агрегує всі дані, завантажені КМ, і починає пошук наступного хопа CH передачі даних у сусідньому кільці, розташованому ближче до БС.

Нижче описана схема завантаження даних:

- CM(и) завантажують дані в CH.

- Поточний CH агрегує дані та розсилає рекламні повідомлення, щоб знайти потенційного наступного CH у сусідньому кільці ближче до БС.

- Зацікавлені потенційні ретрансляційні вузли підтверджують рекламні повідомлення на поточному CH.

- Поточний CH вибирає наступний хоп CH, який має найвищу співвідносну мобільність.

- Вищезгадані кроки повторюються до тих пір, поки всі пакети даних не досягнуть БС.

Схема етапів завантаження даних представлені на блок-схемі на рис. 2.3.

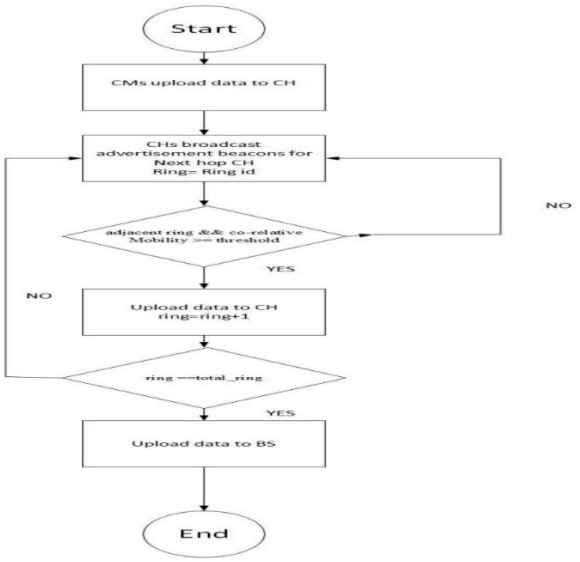



Рисунок 2.3 - Схема логічного потоку завантаження даних

## 2.9 Реалізація та результати

Запропонована схема агрегації даних IoT на основі D2D змодельована у Matlab 2012, Для уявлення конкурентоспроможності пропозиції в порівннянні з іншими подібними схемами, було зіставлено отримані результати з результатами нещодавно опублікованого дослідження, відомого як схема D2D-EE. Схема D2D-EE також базується на D2D комунікаціях всередині кластера пристроїв та агрегує дані в кластерних головах перед завантаженням у БС. З огляду на подібність із запропонованою схемою вона обрана для порівняння. Під час моделювання було розглянуто відстань та швидкість як випадкові змінні. Інтуїтивно зрозуміло, що запропонована робота заснована на співвідносній мобільності, тому очевидно, що в сценарії масового IoT швидкість та відносна відстань окремих вузлів мають мінімальний вплив на загальну продуктивність. Отже, твердження про те, що запропонована схема включає як стаціонарні, так і нестаціонарні вузли, підтверджується з урахуванням теоретичного аналізу, і небуло надано жодних графіків для спостереження впливу мобільності вузлів на запропоновану систему. Більше того, експерименти, проведені під час симуляції, досліджують енергоефективність усієї системи у різних сценаріях. У ході імітаційного дослідження було розроблено систему, що складається з N випадкових вузлів та БС у центрі. Деякі параметри залишалися постійними під час проведення експериментів, як показано у таблиці 2.1. Формування кластерів на основі спільної мобільності зображено на рис. 2.4.

Під час першого експерименту було побудувано графік енергоефективності (біти/джоуль) як в залежності від кількості пристроїв IoT, де довжина пакета фіксована та становить 10 байт, як показано на рис. 2.5, енергоефективність зростає із збільшенням числа пристроїв обох схем. Більше того, запропоноване рішення дуже перевищує у всіх випадках, і воно приблизно вп'ятеро більш енергоефективне, ніж схема D2D-EE на основі кількості пристроїв, що беруть участь. У ході другого

експерименту було збудувано графік енергоефективності, при зміні розмір пакета від 0 до 100 байт при постійній кількості пристроїв 50, як показано на рис. 2.6.

Таблиця 2.1 - Параметри значень

| Параметри | Значення |
|---|---|
| Кількість раундів повторення процедури | 2 |
| $\alpha$ | 0,5 |
| $\beta$ | 0,5 |
| Дальність передачі маяка | 300м |
| Інтервал часу передачі | 1м/с |
| Довжина пакета керування | 200біт |
| Початкова енергія кожного вузла | 0,5 джоуль |
| Параметри | Значення |
| Максимальна кількість кілець | 5 |

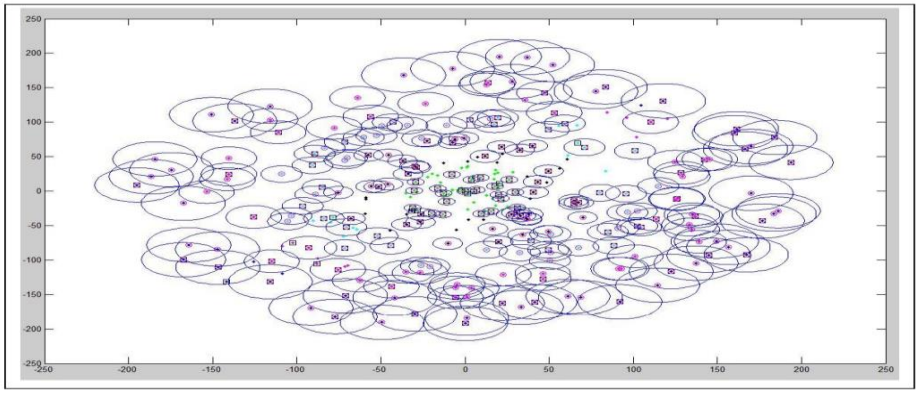



Рисунок 2.4 - Формування змодельованих кластерів на основі співвідносної мобільності

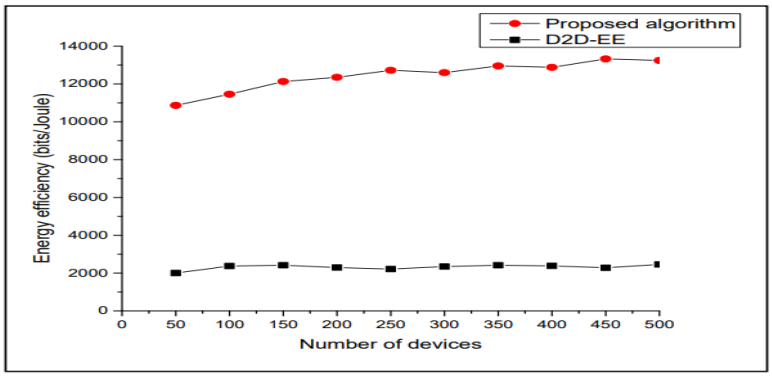



Рисунок 2.5 - Енергоефективність в залежності від кількості пристроїв

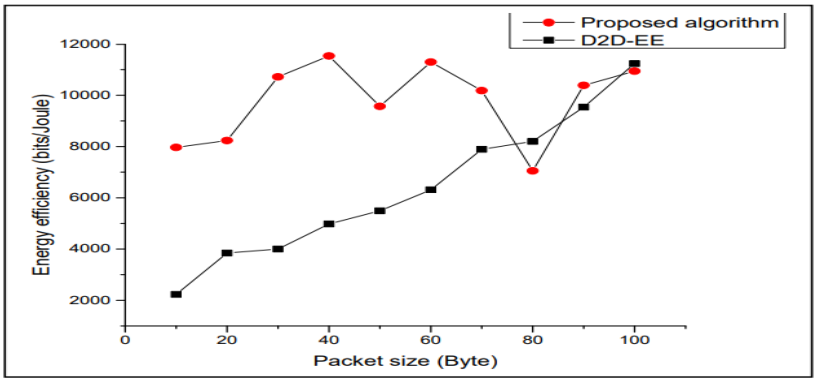



Рисунок 2.6 - Енергоефективність в залежності від розміру пакета

Як видно з графіка, енергоефективність запропонованого рішення збільшується до 40 байт, досягаючи максимальної енергоефективності, а після цього вона відхиляється від графіка. У разі рішення D2D-EE, вона збільшується з збільшенням розміру пакета. Більше того, запропоноване рішення зберегло свою перевагу над D2D-EE майже в 80 % експериментальних результатів.

# РОЗДІЛ 3 ЕВРИСТИЧНЕ ВИРІШЕННЯ ПРОБЛЕМИ ДОСТОВІРНОСТІ ДАНИХ ДЛЯ МАСИВНИХ ОБСЯГІВ НЕОБРОБЛЕНИХ ДАНИХ IoT

## 3.1 Датчики IoT

Необроблені дані датчиків IoT, зібрані за допомогою зв'язку "пристрій-пристрій" (D2D) від нерівномірно розподілених мереж датчиків IoT, мають високу невизначеність через наявність шуму, викидів, відсутніх показань та надмірності. Ці невизначеності, якщо їх не усунути, можуть поширитися по всій системі та погіршити загальну продуктивність, що формально називається проблемою достовірності даних. У цьому розділі розглядається проблема достовірності даних у необроблених даних датчиків IoT, зібраних за допомогою D2D-зв'язку. Ефективність досліджень, проведених Балзано і Новаком та Хаге і Клейнстубером забезпечують нам сильну мотивацію для вирішення проблеми достовірності даних IoT. У цій магістерській роботі використовуються такий підхід, як вбудований алгоритм навчання для локальної обробки даних, що може зруйнувати глобальну кореляцію і, отже, негативно вплинути на прийняття рішень. Більше того, так звані надійні алгоритми аналізу даних фактично ставлять під свою загрозу точність рішень для досягнення високої стійкості до високо невизначених даних датчиків IoT. Запропонований підхід заснований на архітектурі агрегації даних, розглянутої в розділі 1 рис. 1.5. У цьому розділі розглядається застосування запропонованої архітектури. де вона виконує агрегацію даних для покращення якості даних шляхом відновлення відсутніх значень, зіпсованого читання, викидів та надмірності, не впливаючи на глобальні внутрішні закономірності вихідних даних. Інформаційний зміст внутрішньої закономірності вищий, ніж у інших даних і може бути використаний для оцінки кращої версії (істинних даних) даних, яка є більш надійною і може бути використана безпосередньо для аналізу даних. Більшість існуючої літератури спрямовано на скорочення сенсорних даних, як ліки від невизначеності в

сенсорних даних. Ці підходи можна згрупувати в вигляді таких випадкових методів як: a) вибірка та апроксимація. Ці випадкові методи ефективні з обчислювальної точки зору, однак, спричиняють великі втрати важливих даних, що збільшує невизначеність. b) методи спостереження, такі як регресія, потребують великих обчислювальних та комунікаційних витрат. Регресія також вимагає попередньої інформації, що робить її непрактичним для реалізації. в) неконтрольовані методи, такі як аналіз основних компонентів (PCA), не можуть впоратися з невизначеністю. аналіз (PCA) також не може впоратися із невизначеністю. Інші існуючі рішення надто складні для практичної реалізації. Тому існує гостра необхідність у практичній уніфікованій парадигмі, яка може перетворити вкрай невизначені необроблені дані датчиків IoT на надійні дані датчиків IoT.

У цьому розділі пропонується альтернативна парадигма, у якій механізм доставки даних IoT та функціональні можливості агрегації даних IoT стають незалежними від платформи аналізу даних. Запропонована парадигма дозволяє розробникам алгоритмів, додатків для аналізу даних повністю зосередитися на точності прийняття рішень, а не на стійкості програми. Механізм доставки базується на передачі даних від пристрою до пристрою комунікації, аналогічно розділу 2, хоча в цьому розділі IoT-пристрої підключені до туманного сервера за допомогою WiFi-маршрутизаторів, які виступають як базова станція. Крім того, туманний сервер також виступає як препроцесор даних для агрегованих необроблених даних датчиків IoT. У цьому розділі пропонується новий підхід до агрегації даних для покращення якості агрегованих необроблених даних IoT на туманному сервері. Запропонований підхід оцінює справжню матрицю сенсорних матрицю даних із необроблених даних датчиків IoT. Справжня матриця даних не залежить від невизначеності із-за шуму, викидів, значень і надмірних даних. Більше того, вона також містить внутрішні характеристики, які справді відображають динаміку необроблених даних IoT і, отже, являється більш надійною, ніж необроблені дані датчиків IoT, для аналітичних додатків. Запропонований підхід складається з трьох етапів: спочатку відновлюється

груба оцінка вихідного підпростору, яка відновлюється за допомогою вибіркових даних, потім цей підпростір оптимізується для відстеження домінуючого підпростору; по-друге, домінуючий підпростір використовується для знаходження вектора посилення пристроїв IoT; і, по-третє, вектор посилення використовується для оцінки істинних даних датчика матриці. Цей підхід є міждисциплінарним за своєю суттю, оскільки відстеження підпростору та сліпе урівнювання широко використовуються в комп'ютерному зорі та обробці сигналів, відповідно.

### 3.2 Підходи, спрямовані на очищення необроблених сенсорних даних

У цьому розділі обговорюються деякі з відомих підходів, схожих на данну пропозицію, які спрямовані на очищення необроблених сенсорних даних, у хронологічному порядку, а також порівняння їх з данним запропонованим підходом. Алгоритм SMURF є одним із перших спроб очищення необроблених даних RFID даних за допомогою адаптивного фільтра, що згладжує, на основі вікна. Використовується теорія статистичної вибірки для безперервного вивчення розміру вікна. Більш того, різні підходи до вибірки, такі як випадкова вибірка, систематична вибірка, кластерна вибірка, вибірка за квотами, також використовуються у літературі, для очищення необроблених сенсорних даних. Основним недоліком підходів до вибірки є те, що вони зазвичай зберігають локальну кореляцію та руйнують глобальну внутрішню кореляцію необроблених даних, що підвищує випадковість. Хоча пропозиція агрегує дані у вигляді матриці та виконує очищення даних на основі глобально представленних даних.

Апроксимаційні підходи часто використовуються для оцінки більш надійних даних із необроблених даних. У загальному випадку апроксимаційні підходи поєднують дані в межах точності, обмеженої похибкою незалежно від фактичних умов системи. Апроксимаційні підходи мають обмежену практичну реалізацію у сценаріях, де точність є критичною вимогою, а також складний характер

теоретичного аналізу апроксимаційних підходів робить їх непривабливими для розробників. Хоча запропоноване рішення вимагає оцінки, однак, запропонований підхід немає меж похибки точності.

Ймовірні підходи, що базуються на Байєсовській статистиці, генерують модель на основі визначених користувачем ознак для виділення важливих даних серед вихідних даних, відбувається генерація моделі необроблених даних тимчасового ряду мобільної радіомітки, використовуються такі важливі характеристики, як мобільність користувача, динаміка об'єкта, зіпсовані свідчення. Фільтр частинок розроблений для відстеження бажаної інформації із необроблених даних. Довірчий інтервал обчислюється на основі багатовимірної моделі нормального розподілу ймовірності потоку необроблених даних IoT для фільтрації небажаних даних. Цей підхід вимагає попередньої інформації про дані для створення моделі. Запропонований підхід не вимагає жодної попередньої інформації та дає задовільні результати.

Стійкий аналіз головних компонентів (RPCA) є домінуючою парадигмою для вилучення бажаного підпростору з вкрай невизначених необроблених показань датчиків. Переслідування викидів RPCA дозволяє отримати оптимальну низькорангову матрицю без шуму. Цей підхід був розширений як ітеративний підхід, заснований на пороговому обчисленні, з меншою складністю, ніж перший, навіть в умовах шуму. Деякі інші помітні підходи для RPCA можна знайти у представленому кластерному аналізу даних, це підхід з використанням рекурсивного аналізу основних компонентів для виявлення провалів даних IoT, які обмежуються оцінкою підпростору, проте запропонований підхід виконує надійну оцінку підпростору, як перший крок данної пропозиції і далі використовує оптимальний підпростір для генерації надійних сенсорних даних. У випадку данного підходу, метою якого є виявлення викидів, він націлений на виявлення та усунення невизначеності.

### 3.3 Модель системи

У цьому розділі представлена чотиришарова архітектура для масової IoT аналітики. Приклад масового розгортання IoT можна знайти у практичних сценаріях, публічні та приватні хмари можуть знаходитися дуже далеко від фактичного географічного розташування розгорнутої мережі датчиків. У випадку вищезгаданого прикладу, хмарні сервери розташовані в місті Пекін, віддалені хмарні сервери можуть викликати високу затримку та погіршити загальну продуктивність системи. Туманний сервер, який розміщений відносно ближче, може забезпечити платформу для фільтрації та аналізу сенсорних даних поблизу мережі датчиків даних. Це зменшує загальну передачу даних у хмару і тим самим покращує загальну продуктивність системи. У цьому розділі туман виступає як препроцесор даних та базової станції, яка керує та контролює зв'язок між пристроями IoT. Сервер туману також агрегує дані датчиків та генерує локальну матрицю трафіку даних датчиків, потім ця матриця трафіку обробляється за допомогою запропонованого алгоритму (розділ 3.8). для зменшення невизначеності, викликаної зашумленими показаннями, відсутніми даними, викидами та надмірністю. Хмарний сервер завантажує оброблені дані з серверів туману та генерує центральну матрицю трафіку даних для централізованої аналітики та прийняття рішень. Модель системи зображено на рис. 3.1.

Передбачається, що всі вузли IoT з'єднані із сервером туману за допомогою маршрутизаторів Wi-Fi. Передбачається, що рівень маршрутизації також є сценарієм зв'язку між пристроями, сценарій в якому кожен IoT-вузол $\{n_1 \ldots n\} \in N$ формує кластери $\{1_1 \ldots 1\} \in L$ с сусідніми вузлами, і кожен кластер має голову кластеру. У даній головній частині передбачається, що вузли IoT є стаціонарними інтелектуальними пристроями з обмеженими можливостями живлення, зберігання та обчислення. Враховуючи той факт, що вузли IoT обмежені в ресурсах, їхня роль обмежена зондуванням та передачею даних. Вектор даних датчиків IoT в певний

момент часу, коли кожен датчик генерує свої показання, може бути представлений як $y_{\{1\ldots n\}} = [y_1 \cdots y_n]^T$. Туманний сервер надає унікальний ідентифікатор кожного IoT-пристрою, а також поділяє IoT-пристрої на основі обслуговуючого головну частину кластеру. Головна частина кластера відповідає за агрегацію даних, що генеруються членами кластеру. Голова кластера далі передає дані в наступний кластерний вузол, який знаходиться поруч з туманним вузлом, тому дані доставляють на туманний сервер за допомогою міжкластерного багатошляхового зв'язку між пристроями.

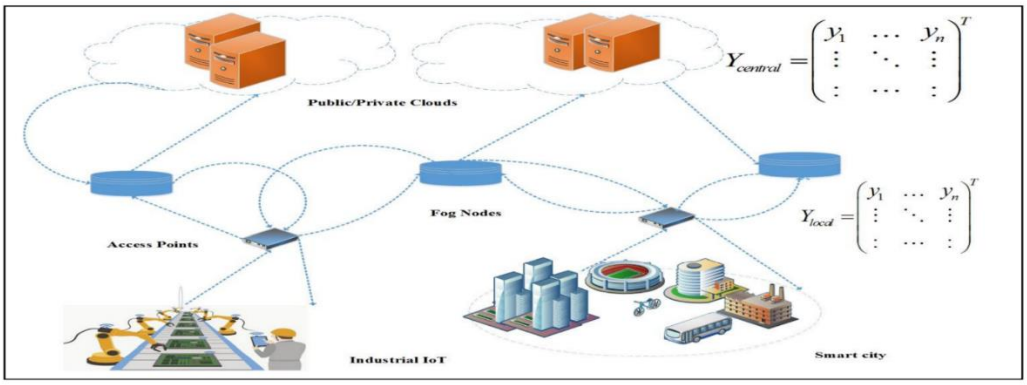



Рисунок 3.1 - Сценарій масової агрегації даних IoT

Схему вибору голови кластера серед інших членів кластера та схему вибору головної частини кластера для міжкластерного зв'язку між пристроями можна знайти у розділі 1.

### 3.4 Невизначеність у сенсорних даних датчиків IoT

Невизначеність - це широке поняття, і досі немає єдиної моделі, яка була б вірна для всіх сценаріїв. Тому невизначеність даних вивчається на основі конкретних сценаріїв. У існуючій літературі представлено п'ять різних видів невизначеності, саме:

Таблиця 3.1 - Опис основних символів у розділі 3

| Символ | Опис |
|---|---|
| N | Загальна кількість вузлів IoT |
| H(X) | Ентропія Шеннонса випадкової величини X |
| $P(x_i)$ | Розподіл імовірності $x_i$ |
| U | Базис ортогонального підпростору |
| $\beta$ | Значення зміщення датчиків |
| Z | Вектор дійсних показань датчика |
| $\alpha$ | Вектор посилення датчика |
| Y | Матриця трафіку даних датчиків IoT |
| $\vec{Y}$ | $\vec{Y} = diag(Y)$ |
| $\bar{Y}$ | Матриця трафіку даних датчиків IoT із нульовим середнім центром |
| $\widehat{Y}$ | Матриця патріархату на основі лінійного оператора $\Phi(\bar{Y}) = \widehat{Y}$ |
| L | Апроксимація низького рангу $\bar{Y}$ |
| $h_\Delta$ | Гладка апроксимація неопуклої штрафної функції h |
| $\theta$ | Доповнення сигнального простору $I - \delta$ |
| $\omega$ | Загальна кількість ітерацій |
| $\Delta$ | Параметр згладжування |

Ентропія Шеннона, ентропія класифікації, нечіткість, неспецифічність та грубий ступінь, невизначеності що обговорюються в цьому розділі, обумовлені наявністю шуму, викидів, відсутніх показань та надмірності. У цьому розділі невизначеність даних датчиків IoT сприймається на основі моделі ентропії Шеннона. Ентропія Шеннона $\mathrm{H}(X)$: У розділі розглядається випадкова змінна $X = \{x_1, ....., x_n\}$. Розподіл ймовірності випадкової величини може бути представлений в вигляді $\mathrm{as}\, P = \{p_1, ... p_n\}$.

$$\mathrm{H}(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i) \tag{3.1}$$

Інтуїтивно ентропія Шеннона випадкової змінної – це кількість інформації (тобто. істинний підпростір у цьому розділі), що міститься в змінній. Це не просто загальна кількість різних значень для випадкової змінної (тобто необроблених даних датчика у разі). Наприклад, інформація в електронному листі - це не лише кількість можливих слів або різних варіантів використання слів, виходить навпаки, інформація електронного листа пропорційна кількості здивування, яке викликає його прочитання. Виходячи з вищезгаданого пояснення, в розділі можна з упевненістю зробити висновок, що інформація - це не просто вихідні дані, а низькорозмірний патерн всередині вихідних даних, який містить у собі більшість властивостей вихідних даних. У розділі 3.8 використовується ця ідея розробки схеми агрегації даних.

### 3.5 Зв'язок D2D

3GPP відіграла ключову роль у створенні та оптимізації машинного типу зв'язку (MTC) у релізах 11, 12 та 13, яка надалі переросла в D2D зв'язок. Зв'язок D2D може бути класифікований на стільниковий D2D (внутрішньосмуговий) або

безліцензійний D2D (несмуговий). Внутрішньосмуговий зв'язок D2D може бути класифікований на підсмуговий зв'язок та внутрішньосмуговий D2D, де пристрої D2D ділять стільниковий спектр з іншими пристроями і одночасно накладати внутрішньосмуговий D2D, де пристрої отримують виділений стільниковий спектр. Однак цей розділ присвячений позасмуговому зв'язку D2D на основі неліцензованого спектру з використанням WiFi. У цьому розділі розраховується енергоефективність ( η ) для Сценарію D2D з n кількістю IoT-пристроїв:

$$\eta = \sum_{i=1}^{n} \frac{d_i}{E_i \cdot r_i \cdot TTI} \tag{3.2}$$

Де $d_i$ загальний обсяг завантажених даних, $E_i$ це середня енергія споживана для доставки одного пакета $r_i$ це загальна кількість пакетів даних, які мають бути передані всіма вузлами, а TTI=1 - інтервал часу передачі, який постійний для всіх пакетів.

У пропозиції використовуються два різні набори даних, перший набір даних – це реальний набір даних датчиків, а другий – синтетичний набір даних. Обидва набори даних призначені для експериментальної перевірки двох різних аспектів запропонованого рішення, з одного боку, експерименти з реальним набором даних датчиків підтримують твердження реального застосування, а з іншого боку, експерименти з синтетичними даними датчиків підтримують твердження про ідеальний сценарій.

Набір даних реального світу: отриманий внаслідок розгортання датчиків IoT у будинку, що знаходиться у Стамбрюге, Бельгія. Набір даних складається із 4,5 місяців даних моніторингу температури та вологості, у середньому за 10 хвилин під час передачі із використанням бездротових сенсорних мереж. Бездротовий датчик складається з типового датчика DHT 22 для вимірювання температури та вологості разом із Zigbee для передачі даних по радіоканалу та мікроконтролеру ATmega328P, вбудованого в один модуль. Розміщення датчиків у будинку можна побачити на рис.

3.2. Більше того, було вибрано 400 незалежних потоків даних, які мають 400 показань датчиків у кожному потоці, і було розглянуто кожен потік як незалежний датчик, крім того, було сформульовано матрицю 400 X 400 з тестових даних, де було розглянуто 400 показань температури від 400 різних датчиків.

Враховуючи тестові дані з нульовим середнім значенням, аналіз основних компонентів (PCA) обчислює осі з максимальною дисперсією та мінімальною надмірністю. Ця максимальна дисперсія також відбиває динаміку необроблених даних датчиків IoT. Застосування PCA до данної реальних даних датчиків показує, що основна дисперсія відображається лише у невеликій частині домінуючих підпросторів (рис. 3.3). Це спостереження дає нам надійну основу для припущення, що необроблені дані датчиків IoT мають внутрішній підпростір більш низької розмірності.


(a) (b)

Рисунок 3.2 - Розташування датчиків температури та вологості у приміщенні: (a) перший поверх; (b) другий поверх

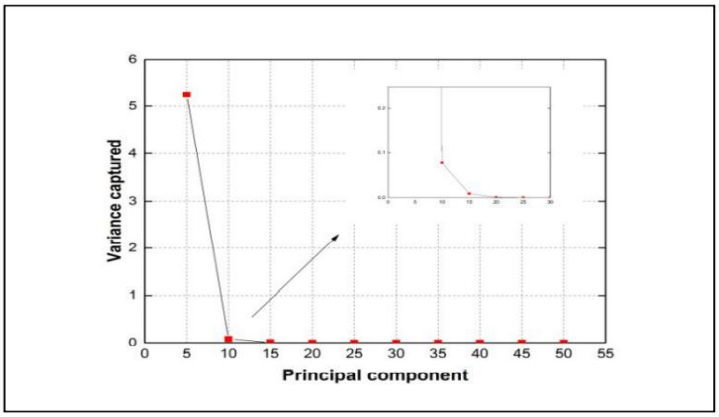



Рисунок 3.3 - Частка загальної дисперсії трафіку каналу, захоплена домінуючим підпростором

Синтетичний набір даних: Створення матриці трафіку тестових даних 400 x 400, щоб емулювати трафік даних трафіку 400 вузлів IoT, кожен із яких генерує 400 зразків протягом заданого часу. Тестова матриця трафіку даних є сумою матриці фіксованого рангу k і розрідженої матриці.

Посилення та зміщення: Посилення та зміщення генеруються за допомогою рівномірних розподілів [0.5, 1.5] та [-0.5, 0.5], відповідно.

Шум: Тут було згенеровано матрицю 400 x 400 псевдовипадкового середнього нульового шуму Гауса. У данних експериментах для візуального представлення простої залежності між дисперсією шуму та обуренням підпростору, було вирішено безпосередньо обурити підпростір шумом, а не додавати шум до даних датчика, що зрештою призводить до обурення підпростору.

Викиди: було побудовано розріджену матрицю викидів розміром 400 x 400 із щільністю 0,2. величина викиду має рівномірний розподіл [-10, 10], і ця величина досить велика, щоб досить велика, щоб спотворити обчислений підпростір. Записи розрідженої матриці викидів мають наступний розподіл Бернуллі.

Пропущені значення: було навмисно зроблено 30% записів трафіку IoT відсутнім значенням. Перед додаванням обурень до даних за допомогою шуму, викидів та пропущених значень, було складено низькорозміру матрицю рангу k, яка

вилучається з реальних та синтетичних даних за допомогою розкладання за сингулярними значеннями, після чого всі сингулярні значення, що перевищують k, стають рівними нулю. $(\Sigma_{i\geq k} = 0),$ потім перемножуються коефіцієнти $(U\Sigma V^T).$ Ця матриця низької розмірності служить справжніми даними датчика і використовується для порівняння данних результатів у розділі 3.10.

### 3.6 Формулювання справжніх сенсорних данних

Для збереження простоти у всьому розділі передбачається однрвимірний сенсор, але цей приклад можна поширити і на сценарій із багатовимірними датчиками, де група датчиків реєструє різні явища. Тут на чолі робиться важливе припущення, що існує лінійний підпростір нижчої розмірності, який несе в собі справжні характеристики вихідних даних, тобто справжній підпростір n-мірного евклідового простору датчиків. Розглядаючи датчики як системи (рис. 3.4), інтуїтивно кожен датчик показань може бути представлено як

$$y_{(i)} = \frac{z_{(i)} - \beta_{(i)}}{\alpha_{(i)}} \tag{3.3}$$

Де $y$ вектор необроблених даних датчиків IoT, $\beta$ вектор, що містить значення зміщення датчика значення, $\alpha$ вектор, що містить коефіцієнти посилення датчиків та $z$ це вектор, що містить істинні значення показань датчиків. Рівняння 3.4 може бути переформульовано для отримання рівняння, яке відображає істинні дані датчиків із матриці трафіку необроблених даних IoT.

$$z = \vec{Y}\alpha + \beta \tag{3.4}$$

Де $\vec{Y} = diag(Y)$ та кожна ненульова компонента $\vec{Y}$ середнє значення $y_{\{1\ldots n\}}$ в на даний момент часу. Датчики реагують на зміну фізичних умов, тобто, стимулів, шляхом чергування електричних властивостей, таких як питомий опір, напруга та

струм. Зміщення зазвичай використовується для балансування зміщення і явно вказується у технічному паспорті датчика. Тому практично припустити, що кожен вузол IoT знає про своє зміщення значення.

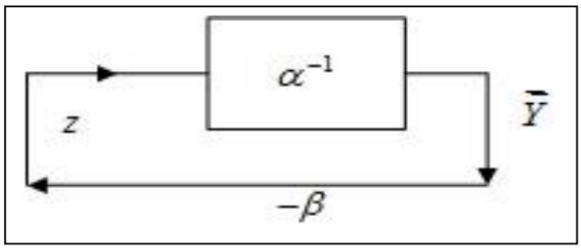



Рис 3.4 - Частка загальної дисперсії трафіку каналу, захоплена домінуючим підпростором

### 3.7 Основний підхід, надійна оцінка підпростору

Передбачається, що необроблені дані датчиків IoT мають внутрішній істинний підпростір, який несе в собі більшу частину динаміки необроблених даних, а в розділі 3.4 було визначено інформаційний зміст, тобто. справжні дані, як більш надійні дані з мінімальними чи відсутніми невизначеностями, також придатними для аналізу даних. Для вирішення проблеми достовірності даних IoT відбувається пошук підходу, який знаходить надмірний лінійний внутрішній підпростір навіть у умовах високої невизначеності, потім цей підпростір використовується для генерації більш надійних даних, які емулюють справжню динаміку необроблених даних IoT без або з мінімальними невизначеностями.

Надійна оцінка підпростору фокусується на двох кроках, перший – реконструкція підпростору використовуючи дані вибірки, і другий – ітеративно відстежуючи внутрішній домінуючий підпростір. $\bar{Y}$ нульове середнє центроване $m \times n$. Матриця трафіку даних датчиків IoT (Рис. 3.1). домінуючий підпростір (DS)

може бути визначений як k некорельовані лінійні компоненти $\bar{Y}$ як згідно з рівнянням 3. 5.

$$DS(\bar{Y}) = U_i^T \bar{Y} \tag{3.5}$$

Де k ≤ m і $\{U_i\}_{i=1}^{k}$ є k ортогональними власними векторами від $\Sigma_{\bar{Y}}=\mathrm{E}\left[\bar{Y}\times\bar{Y}^T\right]$. Більш того $\{U_i\}_{i=1}^{k}$ також є першим *k* стовпці лівого сингулярного вектора *U* у сингулярному розкладанні за значеннями $\bar{Y} = U\sum V^T$. У цьому розділі передбачається, що матрична модель даних трафіку модель як $\bar{Y} = L + S$, де *L* є апроксимацією низького рангу та S це розріджений компонент $\bar{Y}$. Оскільки ранг L менше ніж kL матриця може бути записана як $L = U\bar{Y}$. Спочатку для такого підходу вибирається часткова матриця $\hat{Y}$ as $\hat{Y} \subseteq \bar{Y}$ використовуючи лінійний оператор $\hat{Y} = \Phi(\bar{Y})$. Ця часткова матриця відновлюється ітеративно. Як підсумок, ціль полягає в тому, щоб визначити оптимальний базис для внутрішнього підпростору, використовуючи часткові записи (реконструйовану матрицю) матриці даних IoT, тобто розв'язування рівняння 3. 6.

$$\min_{rk(L)\le k} \left\| \hat{Y} - \Phi(L) \right\|_0 \tag{3.6}$$

Рівняння 3.6 є складним з обчислювальної точки зору, а звичайні методи оптимізації дуже повільні. для практичної реалізації, тому в даному підході вводиться неопукла розряджена штрафна функція $h$ Екв. 3. 7.

$$\min_{rk(L)\le k} h\left(\hat{Y} - \Phi(L)\right) \tag{3.7}$$

Цей підхід враховує $h_\Delta$ як гладку апроксимацію h яка може бути сформульована як $l_p$ норма,

$$h_\Delta(\widehat{Y}) \approx \sum_{j=1}^{n}\sum_{i=1}^{m}(\widehat{y}_{ij}^2+\Delta)^{p/2}, 0<p<1 \tag{3.8}$$

Параметр Δ може бути налаштований для видалення викидів, а також для створення розрідженості в викиди. Ця штрафна функція, що згладжує, має такі переваги, як $l_0$ регуляризація з більш швидкою збіжністю. На основі емпіричної інформації помічено, що велика дельта призводить до швидшої збіжності, а відносно мала дельта - до розрідженого результату. Як підсумок, цей підхід шукає ортонормальний базис для домінуючого підпростору як показано в наступному рівнянні.

$$U^{(i+1)} = \underset{U^TU=I}{\arg\min}\, h_\Delta(\widehat{Y} - \Phi(UU^T L^{(i)})) \tag{3.9}$$

Наведена вищее оптимізаційна задача (рівняння 3.9) вирішується ітераційно з використанням методу змінного направлення методу множників. Для досягнення плавного U у всьому, Δ зменшується ітеративно з кожним кроком. Більш того, результуюча U не є однозначно визначененою. Оскільки підхід працює $\widehat{Y}$ (рівняння 3.9), яке є підмножиною трафіку вихідних даних, а не домінуючий підпростір стійкий до різних невизначеностей, який вбудований у необроблені дані датчиків IoT.

### 3.8 Оцінка дійсних даних датчика

На підставі рівнянь 3.3 та 3.4 передбачається, що існує лінійний підпростір усередині неструктурованих масивних необроблених даних IoT. У цьому розділі далі обговорюється підхід до вилучення справжніх показань датчиків із масивних необроблених даних IoT. Цей підхід складається з двох кроків, спочатку підхід проектує необроблені дані датчиків IoT в домінуючий підпростір і обчислює вектор

посилення датчиків, по-друге, підхід оцінює справжню матрицю датчиків IoT на основі вектора посилення.

Підхід розглядає простір сигналів як $\delta = UU^T$ and $\theta$ доповнення до сигнального підпростору. Тоді з рівняння 3.4 можна з впевненістю припустити, що кожна точка в $z$ також відноситься до $\bar{Y}$ i.e. $z \subseteq \bar{Y}$. В рамках даного підходу спостерігається зв'язок між домінуючим підпростором та істинним власним підпростором, як показано (рівняння 3.10),

$$\theta z = \theta(\vec{Y}\alpha + \beta) = 0 \tag{3.10}$$

Теоретично це співвідношення виконується, оскільки всі вектори-стовпці з β є константою. Тому підхід зацікавлений лише у відновленні вектора посилення α із рівняння 3.10. Тепер у цьому розділі буде виведено рівняння для середнього центрування трафіку з рівняння 3.10, що залежить тільки від α.

$$\theta\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y\right)\alpha + \beta\right) = 0 \tag{3.11}$$

Переформулюємо рівняння 3.11 таким способом:

$$\theta\left(\frac{1}{n}\sum_{i=1}^{n} Y\right)\alpha = -\theta\beta \tag{3.12}$$

Підстановка рівняння 3.12 до рівняння 3.10

$$\theta\left(\vec{Y} - \left(\frac{1}{n}\sum_{i=1}^{n} Y\right)\right)\alpha = 0 \tag{3.13}$$

$\bar{Y}$ є матрицею трафіку даних IoT із нульовим середнім центром. Тому рівняння 3.13 може бути записане наступним чином:

$$\theta\bar{Y}\alpha = 0 \tag{3.14}$$

Більше того, за наявності навіть незначних невизначеностей неможливо розв'язати рівняння 3.14 Для α у закритій формі. Отже, цей підхід шукає надійне

оптимальне рішення a не справжнє рішення для рівняння 3.14, отже, цей підхід моделює рівняння 3.14 як оптимізаційне завдання.

$$\arg\min_{\alpha} \left\| \theta \bar{Y} \alpha \right\|_2^2 \tag{3.15}$$

Розумним еквівалентом для оптимізаційного завдання, поданого в рівнянні 3.15, є знаходження правих сингулярних векторів $\theta \bar{Y}$ , які пов'язані з одиничними значеннями порядку зростання. Як тільки підхід знаходить оптимальний набір α, використовуючи рівняння 3.4, можна знайти справжній внутрішній підпростір присутній у необроблених великих даних IoT. Кроки для агрегування даних узагальнені у наступних розділах алгоритму 1. Цей алгоритм починається з SVD довільної частини матриці трафіку датчиків IoT $\overline{Y_0}$ та генерує приблизну оцінку L. Спочатку Δ відносно великому, що змушує швидко відновлювати підпростір, доки не буде досягнута хороша достовірна оцінка підпростору, Δ ітеративно зменшується для подальшої оптимізації підпростору. Отриманий домінуючий підпростір використовується на етапі 2 та обчислюється оптимальний вектор посилення. Як Підсумок, на кроці 3 з допомогою рівняння 3.4 генеруються більш надійні дані датчика.

**Algorithm 1: Data Aggregation Scheme**

**Initialization:** $\Phi(\overline{Y_0}) = \hat{Y}$, perform $SVD(\overline{Y_0})$ to obtain $U^{(0)}$ and $L^{(0)} = U^{(0)} U^{(0)T} \overline{Y_0}$

**Set:** $\Delta^{(0)}$, $\Delta^{(\omega)}$ and $\Omega = \left( \Delta^{(\omega)} / \Delta^{(0)} \right)^{1/(\omega-1)}$

1: **for** i=1: $\omega$ ; **do**
2: find optimal $U^{(i+1)}$ {Step 1}
3 $U^{(i+1)} = \arg\min_{U^T U = I} h_{\Omega}(\hat{Y} - \Phi(UU^T L^{(i)}))$
4: $L^{(i+1)} = U^{(i+1)} U^{(i+1)T} \overline{Y}^{(i+1)}$
5: $\Delta^{(i+1)} = \Omega \Delta^{(i)}$
6: **end for**
7: retain only k rows of U
8: find optimal $\alpha$ {Step 2}
9: $\alpha^* = \arg\min_{\alpha} \left\| \theta \bar{Y} \alpha \right\|_2^2$
10: $z = \bar{Y}\alpha + \beta$ {Step 3}

### 3.9 Методологія та оцінка продуктивності

Оцінка продуктивності спрямована на отримання відповідей на такі важливі питання: які характеристики підпростору? наскільки точно запропонований підхід оцінює справжні дані датчиків IoT в умовах високої невизначеності? чи масштабована запропонована схема для IoT додатків? для вирішення першого питання було представлено графічну залежність між дисперсією шуму та помилкою підпростору. Для відповіді на друге питання було представлено оцінку справжніх даних датчиків IoT дані у присутності високого гаусівського шуму з викидами та пропущеними значеннями. Тут було також зіставлено данний підхід до оцінки істинних даних із використанням надійного оцінювача підпростору з базовим алгоритмом. Переходимо до відповіді на третє питання про масштабованість, обчислюючи енергоефективності за зміни кількості пристроїв на основі архітектури мережі D2D.

Підхід націлений на покращення якості масивних даних датчиків IoT, тому вважається, що ефективність підходу може бути доведена навіть при невеликому обсязі даних.

Базовий рівень: Аналіз основних компонентів (PCA) – це класичний інструмент для низькорозмірної апроксимації лінійного підпростору, що називається основними компонентами. Ефективність розкладання по сингулярним значенням (SVD) грає велику роль в популярності PCA. Однак PCA на основі SVD нестійкий до високої невизначеності і може генерувати довільні підпростори. У цьому розділі PCA обраний як базовий рівень, тому що PCA дуже популярний і має багато варіантів, тому може бути хорошим базовим алгоритмом. Крім того, важливо зазначити, що базовий алгоритм використовує PCA тільки для обчислення домінуючого підпростору, тобто, крок 1 для алгоритму 1, а домінуючий підпростір використовується для генерації істинних сенсорних даних, як обговорюється у кроці 2 та 3 алгоритму 1.

Щоб продемонструвати ефективність данного підходу, було проведено чотири експерименти, спочатку був побудуваний графік помилки реконструкції підпростору (рівняння 3.16) як функцію дисперсії шуму, щоб проілюструвати взаємозв'язок між помилкою реконструкції підпростору та дисперсією шуму (Рис. 3.5). Для всіх експериментів розглядаємо ранг низькорозмірного підпростору k=10. На рис. 3.5 дисперсія шуму та помилка реконструкції підпростору мають приблизно лінійну залежність, тобто. дисперсія шуму прямо пропорційна помилці реконструкції підпростору. Зазвичай, ця залежність залежить від природи шуму. Більш того, було припущено, що додавання шуму збільшить помилку (помилка підпростору в данному сценарії). у данному сценарії). Оскільки тепер відомо про взаємозв'язок між помилкою підпростору та дисперсія шуму, було використаноо помилку реконструкції підпростору як еталон характеристики для подання різних рівнів обурень підпростору, залежно від дисперсії шуму.

Фактично, дисперсія шуму варіюється в межах [-0,01, 0,06], щоб обурені дані залишалися корельованими. і якщо збільшити дисперсію шуму набагато більше, ніж у цьому діапазоні, це буде розглядатися як викид, що вже було показано в експерименті 4. Було обчислено помилку реконструкції підпростору так:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\left\|U^{T}\tilde{Y}-U^{T}\bar{Y}_{i}\right\|_{2}}{\left\|U^{T}\bar{Y}_{i}\right\|_{2}} \tag{3.16}$$

Де $\tilde{Y}$ розрахункова середня матриця даних датчика та $\bar{Y}$ це трафік даних датчиків IoT матриці. Розрахування помилки оцінки істинних даних датчиків відбувається наступним чином:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\left\|\bar{z}_{i}-z_{i}\right\|_{2}}{\left\|z_{i}\right\|_{2}} \tag{3.17}$$

Де $\bar{z}$ вектор оціночних даних датчика IoT вузла *i* і z це вектор істинних даних датчика IoT вузла IoT і. Для наступного експерименту було доповнено в підпростір Гаусівський шум. В цьому експерименті просто для того, щоб забезпечити візуально

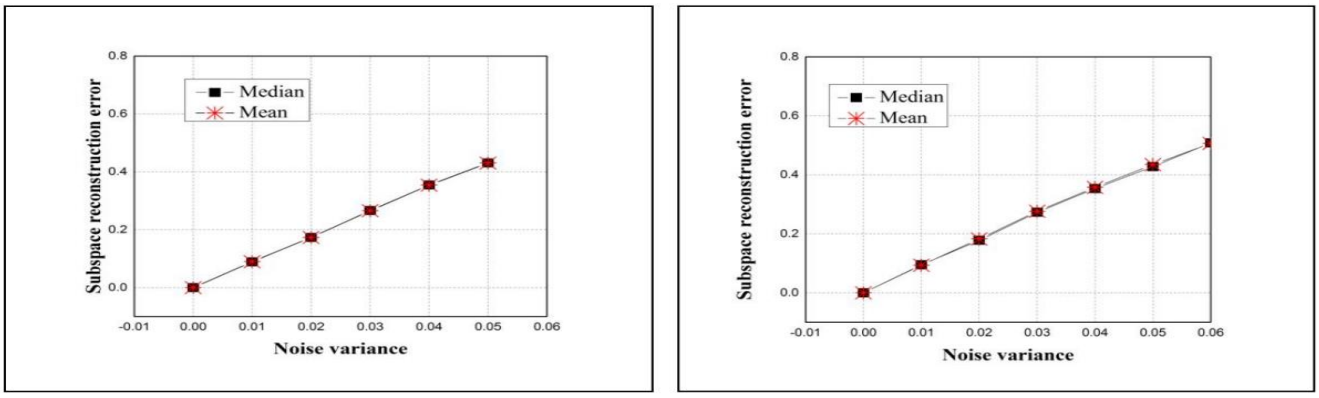
простий зв'язок між дисперсією шуму та обуренням підпростору було вирішено безпосередньо обурити підпростір шумом, а не додавати шум до даних датчика, що зрештою призводить до обурення підпростору, як показано на рис. 3.6, було успішно оцінено справжні дані датчиків у розумній степені для обох наборів даних.

Для третього експерименту був додатково введений високий Гаусівський шум разом із значною кількістю викидів у дані набори даних та оцінили справжні значення датчиків. На рис. 3.7 показано. що данний алгоритм оцінює справжні дані датчиків IoT в розумній мірі навіть за при великих викидів. Більше того, було також зіставлено запропонований підхід з базовим підходом, де було обчислено підпростір за допомогою класичного PCA, а потім оцінюємо справжні дані датчиків, як це обговорювалося в алгоритмі 1. Видно, що базовий підхід генерує дуже високу помилку оцінки істинних даних датчика навіть за низької помилки реконструкції (низька дисперсія шуму), а також генерує випадкові результати без будь-якої закономірності. Більш того, з емпіричних досліджень добре відомо, що SVD на основі PCA нестійка до високих невизначеностей та генерує випадкові основи підпростору.

У четвертому експерименті  було доповнено високий гаусівський шум разом із синтетичними відсутніми значень набір даних. На рис. 3.8 показано, що данний алгоритм може оцінити справжні дані датчика навіть за наявності відсутніх значень. Знову ж таки, PCA на основі SVD генерує високу помилку оцінки поряд із випадковими результатами. Загалом можна помітити, що точність данного підходу на наборі даних датчиків реального світу нижче, ніж на синтетичному наборі даних, при тому, що набір даних датчика вже зіпсований шумом, відмінним від синтетичного адитивного Гаусівського шуму.

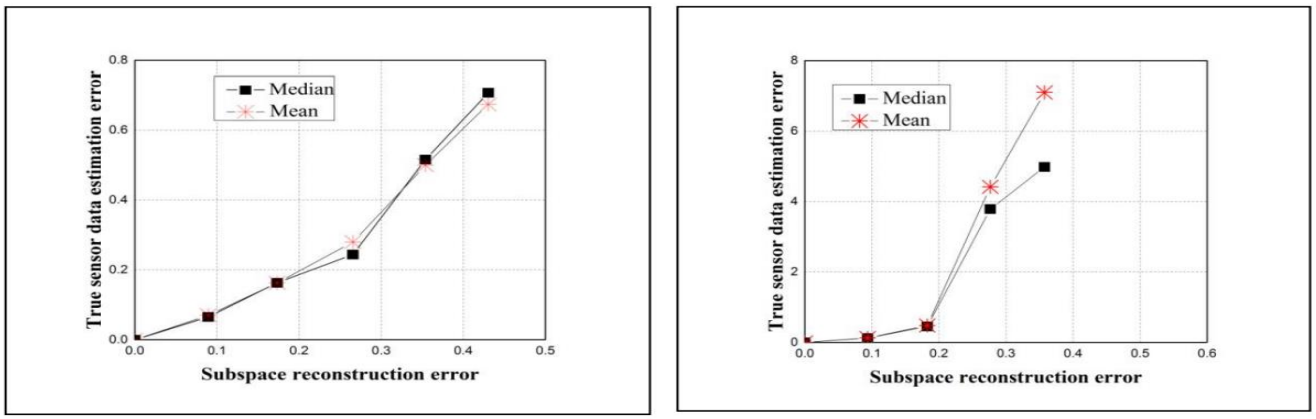
Крім вищезгаданих чотирьох експериментів, спрямованих на перевірку основного підходу до агрегації даних. Було також перевірено масштабованість D2D зв'язку для її практичної реалізації (рис. 3.9). Тут було змодельовано енергоефективність (рівняння 3.2) у разі зміни кількості IoT-пристроїв на основі

моделі системи у розділі 3.3. Було проведено порівняння представлені кластерні D2D комунікації для доставки даних IoT з нещодавньою схемою D2D-EE.



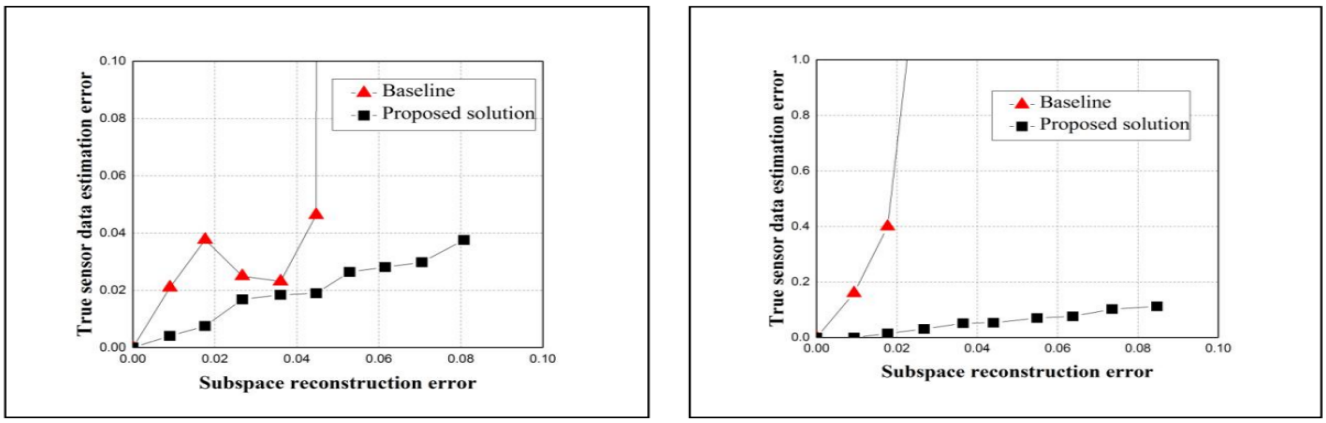

(a) (b)

Рисунок 3.5 - Помилка реконструкції підпростору як функція дисперсії Гаусівського шуму. (a) Синтетичний набір даних. (b) Сенсорний набір даних



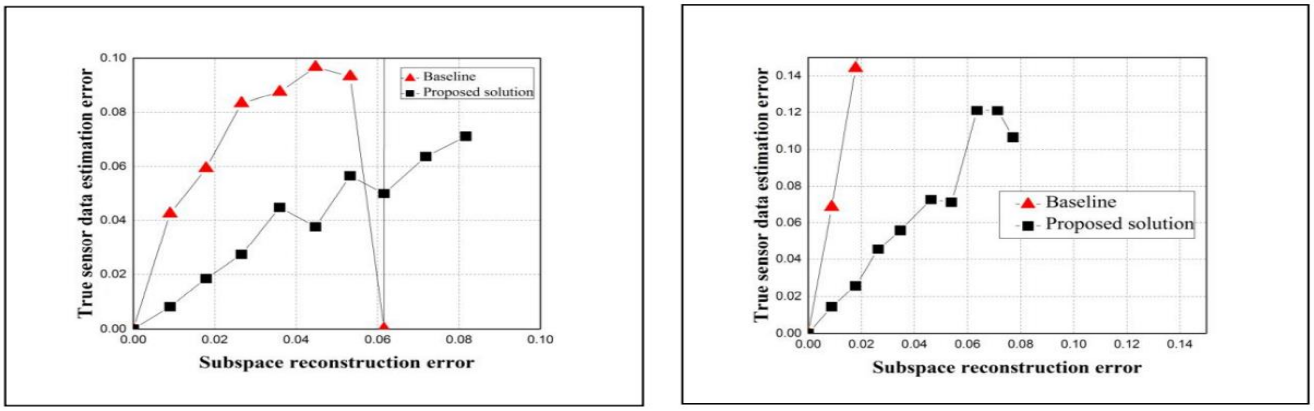

(a) (b)

Рисунок 3.6 - Помилка оцінки справжніх сенсорних даних як функція помилки реконструкції підпростору. Усі значення є середніми за 10 вибірками, де n=400 та k=10. (a) Синтетичний набір даних. (b) Набір даних датчика




(a) (b)

Рисунок 3.7 - Помилка оцінки справжніх сенсорних даних як функція помилки реконструкції підпростору у присутності викидів. Розмір викидів між [-10, 10] із щільністю 0.2, всі значення є середніми по 10 вибірках, де n=400 та k=10. (a) Синтетичний набір даних (b) Сенсорний набір даних




(a) (b)

Рисунок 3.8 - Помилка оцінки справжніх сенсорних даних як функція помилки реконструкції підпростору за наявності пропущеного значення. 30% відсутніх значень і всі значення є середніми за 10 вибірками, де n=400 та k=10. (a) Синтетичний набір даних (b) Дані датчика

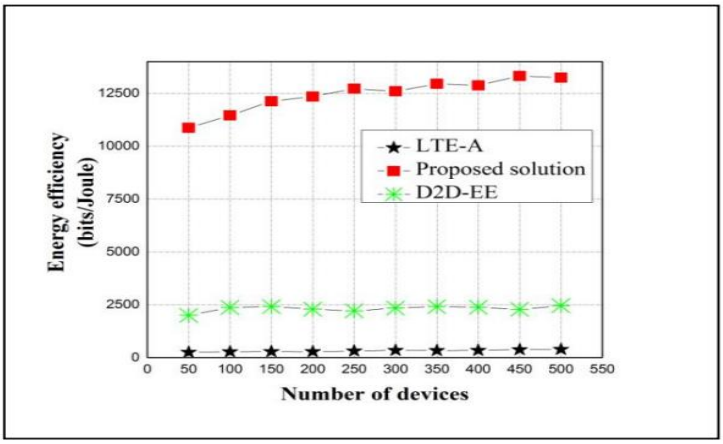



Рисунок 3.9 - Масштабованість, розмір пакета - 10 байт

Базовий підхід нестійкий до високих невизначеностей через те, що підхід розкладання по одиничних значеннях безпосередньо обробляє дані датчиків IoT. декомпозиції безпосередньо обробляє необроблені дані датчиків IoT для обчислення оптимальної базису підпростору. Емпіричні спостереження показують, що він скоріше генерує випадковий базис підпростору у присутності високих невизначеностей, таких як великі викиди та відсутні значення. Тому через вибір випадкового базису підпростору замість оптимального базовий алгоритм не може оцінити справжні дані датчика та генерує випадкові результати. Запропонований підхід не обчислює оптимальний підпростір безпосередньо за необробленими даними IoT дані датчиків, натомість на початку він грубо реконструює підпростір за допомогою часткової матриці, а потім ітеративно відстежує оптимальне домінуюче підпростір. Більше того, використовуючи оптимального підпростору запропоноване рішення ефективно оцінює справжні дані датчиків IoT навіть за при високій невизначеності. Таким чином, запропонований підхід стійкий до високих невизначеності та працює краще, ніж базовий підхід.

Обмеження: Пропозиція робить загальне, але суворе припущення, що кожен базис ортогональний попередньому, більше того, якщо дані розташовані в неортогональному базисі, підхід оцінки домінуючого підпростору (розділ 3.8)

генерує випадкові результати. Пропозиція також передбачає, що викиди більші за розміром, ніж вкраплення, більше, якщо і викиди, і вкраплення можна порівняти за розміром, то запропонована схема може не впоратися із поставленим завданням.

# РОЗДІЛ 4 ФЕДЕРАТИВНА ФІЛЬТРАЦІЯ ТА АГРЕГАЦІЯ IoTMT

## 4.1 Огляд розділу

Всесвітня організація охорони здоровья (ВОЗ) нещодавно повідомила про глобальну нестачу медичних працівників, нестачу 12,9 мільйона людей протягом наступного десятиліття. Ця очікувана нестача, а також різнимі іншими факторами, сприяли поштовхом до повільної, але неухильної зміни парадигми від традиційної охорони здоров'я на "розумну" охорону здоров'я. Інтелектуальна охорона здоров'я дозволяє пацієнтам здійснювати цілодобовий моніторинг та зворотний зв'язок, а також очікується автоматизація критичних операцій усередині відділення інтенсивної терапії. Інтернет речей (IoT) широко визнаний як найважливіший рушійний фактор для парадигми підключеного охорони здоров'я.

Типовий пристрій IoMT складається з крихітного акумулятора, який в більшості випадків не заряджається, що призводить до утилізації обладнання, як воно розряджається. А значна причина швидшої утилізації пристроїв IoMT пов'язана з домінуючою парадигмою хмарних обчислень, коли всі зібрані дані передаються на віддалені хмарні сервери аналітики та прийняття рішень. Це явище спричиняє значні втрати енергії через високих комунікаційних накладних видатків. Більш того, це також впливає на агреговані конфіденційні медичні дані, які наражаються на ризик безпеки. У цьому розділі розглядається проблема високої втрати потужності, конфіденційності, медичних даних та високої затримки у хмарній аналітиці охорони здоров'я. Це цікава проблема, оскільки має соціальні наслідки; крім того, уряди та промисловості вкладають багато грошей та ресурсів у розвиток майбутньої інфраструктури охорони здоров'я, інфраструктури.

У цьому розділі представлено алгоритмічну структуру, а саме Federated Filtering Framework (FFF) (рис. 4.1) для IoMT, яка підкріплена теоретичним аналізом. Запропонована структура представляє альтернативне вирішення проблем

енергоефективності, затримки та конфіденційності для пристроїв IoMT з обмеженими ресурсами. Коротко, кожен пристрій IoMT обчислює локальну модель даних і ділиться цією моделлю із туманним сервером. Роль туманного серверу є потрійною. По-перше він прогнозує матрицю даних (агреговану матрицю даних), використовуючи середнє значення агрегованої моделі (Розділ 4.10); по-друге, він обчислює та передає параметри фільтру для всіх пристроїв IoMT і, як підсумок, приймає рішення, використовуючи агреговану матрицю даних. Для керування обурення власних значень матриці даних, що ставить під загрозу точність рішення, у цьому розділі виводиться теоретичний зв'язок між параметром локальної фільтрації та глобальним допустимим власним збуренням за допомогою теорії збурення матриці (MPT).

## 4.2 IoT у охороні здоров'я

Домінуюча парадигма для аналітичних систем охорони здоров'я на основі IoT може бути класифікована як моніторинг здоров'я на основі хмарних, мобільних обчислень. Обидва вищезгадані сценарії дуже часто передають дані на сервер (хмарний сервер/мобільний пристрій) для прийняття рішень. Цей розділ рішуче виступає проти безперервної передачі даних та представляє схему агрегації даних на основі прогнозування з межами помилок задля забезпечення точності прийняття рішень.

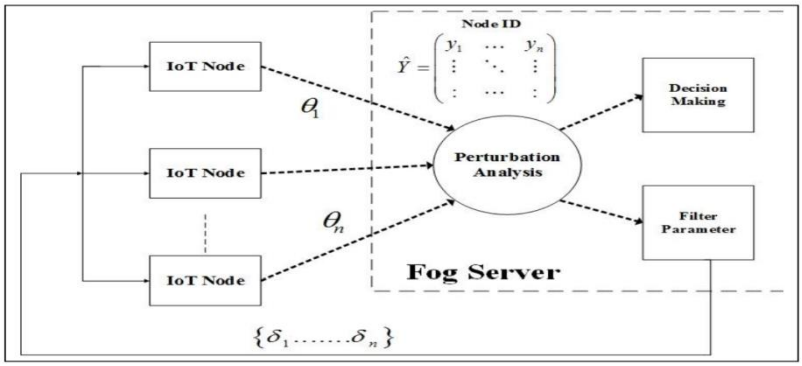



Рисунок 4.1 - Федеративна структура фільтрації

Деякі недавні приклади використання аналітики в охороні здоров'я на основі IoT, також виступають за централізоване прийняття рішень, однак в обох випадках відсутнє теоретичне формулювання для забезпечення точності прийняття рішень.

### 4.3 Системи IoT на основі прогнозування

У літературі описано декілька підходів на основі прогнозування зниження комунікаційних накладних витрат у сенсорних мережах. Підходи на основі прогнозування поділяються на підходи з одним прогнозом і підходи з двома прогнозами. В разі підходів з одним пророцтвом система виконує передбачення тільки в одному місці у той час, як у разі підходів з подвійним пророцтвом система виконує передбачення на локальному вузлі разом із центральним сервером. Деякі відомі схеми передбачення, які застосовуються для обох вищезгаданих категорій, являються схемою адаптивної фільтрації, використовується авторегресійний інтегрований фільтр ковзаючого середнього (ARIMA), фільтрація Калмана та методи машинного навчання. Хоча деякі з попередніх підходів можуть забезпечити кращу точність для створення моделі на пристрої IoT, однак, враховуючи серйозні обчислювальні обмеження пристроїв IoMT, ці підходи недоцільно використовують локальний аналіз. обробки. Більше того, жоден із попередніх підходів не показує жодного взаємозв'язку між локальною та глобальною обробкою з використанням теоретичних верхніх кордонів.

### 4.4 Федеративне навчання у мережах

Оригінальна схема федеративного навчання є перспективним кандидатом для вирішення проблем затримки та конфіденційності. Однак, оскільки вона спочатку була розроблена для мобільних пристроїв, схема припускає, що розподілені пристрої клієнта будуть працювати в мережах, схема передбачає, що розподілені клієнтські пристрої мають достатню кількість ресурсів. Отже, парадигма відкладеного

навчання виступає за реалізацію алгоритмів глибокого навчання усередині розподілених клієнтських пристроїв для локальної генерації моделі. Другим суттєвим недоліком парадигми федеративного навчання полягає в тому, що обчислення локальної моделі та поточної глобальної моделі повністю ізольовані, тому дуже складно централізовано регулювати формування бажаної локальної моделі.

Ефективність алгоритму федеративного усереднення для розподіленого навчання, запропонованого Мкмаханом. забезпечує сильну мотивацію для розробки федеративної фільтрації для пристроїв IoMT. Більше того, існують інші помітні розподілені підходи до оптимізації, що підвищують ефективність зв'язку. Усі розподілені та об'єднані підходи в літературі дуже складні для запуску в крихітному пристрої IoMT. Крім того, вони націлені на прийняття рішень на пристрої. Запропонована Федеративно фільтраційна структура, з іншого боку, пропонує дуже легку підпрограму для сильно обмеженого в ресурсах пристрою IoMT і також націлена на прийняття рішень на сервері з використанням загальної локальної моделі. Запропонована структура також пропонує теоретичну основу для регулювання генерації локальної моделі на основі поточної глобальної моделі. Двома словами, робота в розділі 4 успадкувала всі позитивні сторони парадигми федеративного навчання і налаштувала її. парадигми навчання та підлаштувала парадигму під обмежені ресурси IoMT. сценарію.

### 4.5 Модель системи

Модель системи розглядає сценарій масивного IoMT, де n кількість пристроїв IoMT сумарно працюють над визначенням конкретного явища. Всі пристрої IoMT підключені до туманного сервера за допомогою Wi-Fi. Кожен пристрій IoMT $\{N_1, \ldots, N_n\} \in N_i$ генерує поток данних тимчасового ряду. У цій голові передбачається, що централізовано а агрегрегаційна матриця $Y_i$ також відома як

глобальна матриця (речова) розміру $m \times n$ де кожен стовпець $(Y_i)$ представляє конкретний пристрій IoMT, і в кожному рядку показання датчика фіксуються кожні 30 секунд. Ця генерація глобальної матриці $Y$ вимагає безперервної передачі даних на туманному сервері. Однак цей розділ не є прихильником безперервної передачі даних і тому пропонується система, що базується на пророкуваннях. Сервер туману генерує агреговану матриця даних $(\hat{Y})$; тобто. матриця передбачених даних з обуренням, і як раніше $\hat{Y}_i$ являє собою вектор-стовпчик матриці даних.

Таблиця 4.1 - Опис основних символів у розділі 4

| Символ | Опис |
|---|---|
| $N_i$ | $i^{th}$ Пристрій IoTMT |
| Y | Глобальна матриця |
| $\hat{Y}_i(t)$ | Стовпець $i^{th}$ глобальної матриці |
| $\hat{\cdot}$ | Потривожена версія оригінального символу |
| $\delta_i$ | Параметр $i^{th}$ фільтра |
| $\theta_i$ | Модель прогнозування даних $i^{th}$ IoTMT |
| $e$ | Функція середньої квадратичної помилки |
| $\alpha$ | Швидкість навчання/розмір кроку |

Продовження Таблиці 4.1 - Опис основних символів у розділі 4

| Символ | Опис |
|---|---|
| $\lambda$ | Власне значення матриці |
| $\Delta$ | Помилка обурення |

Обурення в глобальній матриці даних пов'язане з помилками, спричиненими фільтрацією та прогнозуванням. Формування матриці агрегованих даних обговорюється у розділі 4.10. Роль туманного серверу складається із двох частин. По перше, він оцінює/передбачає матрицю обурених даних $(\hat{Y})$, і по-друге, обчислює та доставляє параметр фільтра $(\delta_i)$ для всіх пристроїв IoMT, та як підсумок, він приймає рішення, використовуючи матрицю обурених даних. У таблиці 4.1 наведено деякі корисні позначення. Спочатку всі вузли IoMT навчають модель прогнозування шляхом запуску деяких примірників фільтра найменшого середнього квадрата (LMS) (розділ 4.6). Як локальний пристрій IoMT і туманний сервер використовує таку ж саму схему передбачення. Локальний пристрій IoMT запускає локальну підпрограму обробки, як описано в алгоритмі 1, а туманний сервер виконує алгоритм 2.

### 4.6 Адаптивна фільтрація у пристроях IoMT

Теоретичний аналіз означає дослідження методів розв'язання проблеми та особливостей опису проблеми та впливу вихідних даних на отримані результати. Мета існування полягає в тому, щоб встановити прецедентів та визначити зовнішні зв'язки між ними, зробити докладне пояснення, що спричинило їх виникнення, якою є сутність їх існування.

Адаптивні фільтри зазвичай використовуються для сигналів з нестаціонарною статистикою і в тих випадках. коли відсутня попередня інформація. Типовий адаптивний фільтр зображено на рис. 4.2. Серед різних адаптивних фільтрів у цьому розділі вибрано фільтр найменшого середнього квадрата (LMS) для локальної обробки всередині вузла IoMT, оскільки він має дуже низькі обчислювальні накладні витрати. Нехай для пристрою IoMT $N_i$ в момент часу t прогнозований вектор датчиків IoMT $\hat{Y}_i(t)$ буде лінійною апроксимацією реального вектора датчика

Адаптивний фільтр LMS вбудований у пристрої IoMT, спрямований на мінімізацію функції помилки $e(t)$, Котрий є найменшим середнє квадратичнм наближення між передбачуваним вектором датчика та реальним вектором датчика.

$$e_i(t) = \frac{1}{2}\sum_{i=1}^{n}\left(\hat{Y}_i(t) - Y_i(t)\right)^2 \quad (4.1)$$

Зв'язок між прогнозованим вектором датчика $\hat{Y}_i(t)$ (вихід LMS фільтра) та вектор реального датчика $Y_i(t)$ полягає в наступному.

$$\hat{Y}_i(t) = \Theta_i^T Y_i(t) \quad (4.2)$$

Фільтр LMS заснований на стохастичній оптимізації градієнтного спуску (SGD). підхід потребує ітераційних кроків $(\alpha_i)$ у напрямку найбільш крутого зменшення функції помилки $e_i(t)$. Рівняння 4.3 показує правило оновлення LMS, також відоме зазвичай навчання Уідроу-Хоффа.

$$\Theta_i(t) = \Theta_i(t-1) + \alpha_i(t) \cdot e_i(t) \cdot Y_i(t) \quad (4.3)$$

На підставі емпіричного спостереження для забезпечення збіжності розмір кроку $\alpha_i(t)$ повинен задовольняти такі вимоги.

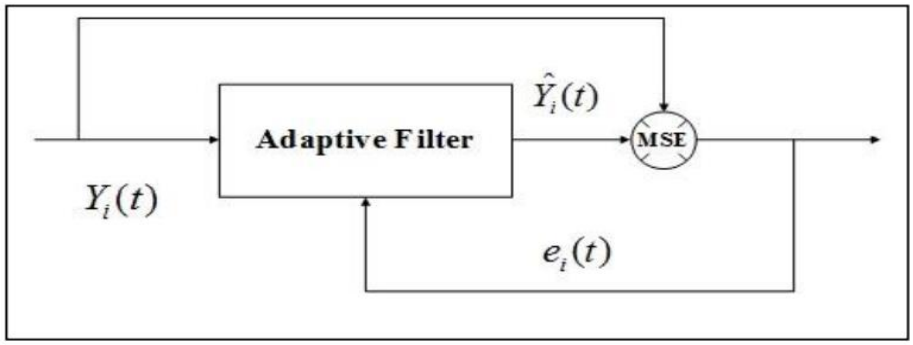



Рисунок 4.2 - Блок-схема адаптивного фільтра

$$0 \le \alpha_i(t) \le \frac{1}{P_Y} \tag{4.4}$$

Де $P_Y = \frac{1}{M}\sum_{j=1}^{M}|Y_i(j)|^2$ та M – кількість ітерацій, необхідних для навчання LMS фільтра.

### 4.7 Аналіз обурень на туманному сервері

Параметри фільтра відіграють ключову роль у балансуванні компромісу між бажаною втратою точності рішення (допускаючи обурення до $\tilde{Y}$) та низькими комунікаційними витратами. У цьому розділі використовується теорія матричних обурень для обмеження помилки обурення (Δ)матр иці обурених даних, що, своєю чергою, впливає точність рішення. Туманний сервер генерує матрицю обурених даних $\hat{Y} = Y + W$ , де W помилка обурення/фільтрації та елементи стовпців W, $W_i \in [-\delta_i, \delta_i]$. Нехай $\lambda_i$ та $\hat{\lambda}_i$ та позначають власні значення речовинної коваріаційної матриці $A = \frac{1}{m}Y^T Y$ та обурена коваріаційна матриця $\hat{A} = \frac{1}{m}\hat{Y}^T\hat{Y}$

відповідно. Норма матриці помилок обурення Δ = A - $\hat{A}$ може бути сформульована з використанням властивості нерівності трикутника , зображується наступним чином.

$$\begin{aligned}\|\Delta\| &= \left\|Y^T W + W^T Y + W^T W\right\| \\ &\le \left\|Y^T W\right\| + \left\|W^T Y\right\| + \left\|W^T W\right\|\end{aligned} \tag{4.5}$$

Завдання полягає в тому, щоб визначити верхню межу для очікування RHS у наведеній вище нерівності нерівності.

У цьому розділі передбачається, що всі вектори-стовпці з W незалежні і всі елементи стовпців є випадковими величинами з нульовим середнім значенням $(\mu = 0)$ та дисперсія $\sigma_i^2 \approx \sigma_i^2(\delta_i)$ разом із четвертим моментом у вигляді $\mu_i^4 = \mu_i^4(\delta_i)$.

Використовуючи нерівність Єнсена $E(x) \le \sqrt{E(x^2)}$.

$$\begin{aligned}E(\|\Delta\|_F) &\le 2E\left(\left\|Y^T W\right\|_F\right) + E\left(\left\|W^T W\right\|_F\right) \\ &\le 2\sqrt{E\left(\left\|Y^T W\right\|_F^2\right)} + \sqrt{E\left(\left\|W^T W\right\|_F^2\right)}\end{aligned} \tag{4.6}$$

Засноване на теоремі Мирського.

$$E\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\lambda}_i - \lambda_i\right)^2}\right) \le E\left(\frac{\|\Delta\|_F}{n}\right) \le Tol_F \tag{4.7}$$

$$E(\|\Delta\|_F) \le 2\sqrt{\frac{1}{m^2 n} Tr\left(Y^T Y\right) \cdot \sum_{i=1}^{n} \sigma_i^2} + \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \cdot \sum_{i=1}^{n} \sigma_i^4} \tag{4.8}$$

$$E(\|\Delta\|_F) \le Tol_F \tag{4.9}$$

Рівняння 4.9 представляє верхню межу ($Tol_F$) про помилку обурення, спричинену локальною фільтрацією на пристроях IoMT та оцінка матриці обурених даних з використанням застарілої загальної Моделі.

$$Tol_F = 2\sqrt{\frac{1}{m^2 n} Tr\left(Y^T Y\right) \cdot \sum_{i=1}^{n} \sigma_i^2} + \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \cdot \sum_{i=1}^{n} \sigma_i^4} \quad (4.10)$$

Аналогічні верхні межі можуть бути отримані за допомогою спектральної норми $\|\cdot\|_2$ крім того, у цьому розділі обирається норма Фробеніуса $\|\cdot\|_F$ без особливих причин.

### 4.8 Рівномірний вибір параметрів фільтру

У цьому розділі передбачається незалежний та рівномірний розподіл параметра фільтра IoT в інтервалі $[-\delta_i, \delta_i]$ Більш того, також маємо на увазі однорідний

розподіл параметрів фільтра між усіма пристроями IoMT, тому $\delta_i = \delta$ і $\sigma_i = \frac{\delta^2}{3}$ Вирішуючи рівняння 4.10 Для δ.

$$\delta = \frac{\sqrt{\frac{3Tr(Y^T Y)}{m} + 3 \cdot Tol_F \cdot \sqrt{nm + m^2}} - \sqrt{\frac{3 \cdot Tr(Y^T Y)}{m}}}{\sqrt{m+n}} \quad (4.11)$$

Рівняння 4.11 забезпечує зв'язок між локальною фільтрацією та глобальною помилкою обурення, що відіграє вирішальну роль у балансуванні компромісу між локальною фільтрацією в IoMT та глобальною помилкою власного обурення.

Федеративна структура фільтрації (FFF) заснована на вільній федерації пристроїв, що беруть участь у ній (IoMT-пристроїв), що координуються центральним сервером туману. FFF складається з із двох найважливіших протоколів, перший - протокол обробки локальних даних, а другий - глобальний протокол обробки даних та координації.

Враховуючи серйозні обмеження ресурсів обчислень для пристроїв IoMT, у цьому розділі пропонується дуже легкий протокол фільтрації локальної обробки.

Локальна фільтрація заснована на адаптивному фільтрі LMS (розділ 4.6). Пристрої IoMT обчислюють локальну модель передбачення $\theta_i$ (Рівняння 4.3) на основі зібраних даних і передають цю модель на туманний сервер. Тепер припустимо, що $\theta_i$ як поточна модель прогнозування та $\delta_i$ як останній параметр фільтрації параметр для $N_i$ Пристрій IoMT. $N_i$ у будь-який момент часу t відстежує відхилення прогнозованого вектор датчика $\hat{Y}_i(t)$ з вектора реального датчика $Y_i$ (t) $Y_i(t)$ використовуючи $W_i(t) = Y_i(t) - \hat{Y}_i(t)$. Коли $|W_i(t)| > \delta_i$ пристрій IoMT оновлює модель прогнозування $\theta_i(t)$ і скидає $W_i$ (t) нанівець. Оновлена модель прогнозування разом із невеликою кількістю даних вибірки передається на туманний сервер. Однак LMS-фільтр несе незначні обчислювальні витрати, що дозволяє пристрою IoMT запускати деяких екземплярів фільтрації для підвищення точності. Згадані деталі для локальної обробки на пристроях IoMT узагальнені в алгоритмі 1.

**Algorithm 1: Local Processing Protocol**

**Input:** current $\theta_i(t), \delta_i(t), SVD(\overline{Y_0}), Y_i(t)$ and $\alpha_i(t)$
**Output:** $\theta_i^*(t)$

```
1: for (true) do
2:     t= current time
3:     compute: W_i(t) = Y_i(t) − Ŷ_i(t)
4:     if |W_i(t)| > δ_i then
5:         [θ_i*(t)] := LMS(Y_i(t), α_i(t))
6:         N_i sends (i, θ_i*(t), Y_i(t)) to fog server
7:         Set W_i(t) ← 0
8:         Set θ_i(t) ← θ_i*(t)
9:     end if
10: end for
```

### 4.9 Федеративна оробка на туманному сервері

На початку кожного раунду туманний сервер оновлює поточні моделі прогнозування за допомогою нових загальних моделей. Туманний сервер вибирає

випадкову частку K з n що беруть участь у програмі пристроїв IoMT. У цьому розділі вибирається випадкова частка пристроїв IoMT, оскільки точність рішення погіршується після певного числа. Розмір кроку $\alpha_i$ (t) підтримується постійно на основі емпіричного результату (розділ 4.6). Туманний сервер агрегує модель використовуючи рівняння 4.12.

$$\eta_i(t) = \sum_{i=1}^{K} \frac{n_k}{n} \theta_i(t-1) \tag{4.12}$$

Після цього туманний сервер прогнозує матрицю обурених даних, використовуючи наступне рівняння.

$$\hat{Y}(t) = \eta_i^T Y(t) \tag{4.13}$$

Матриця обурених даних $\hat{Y}(t)$ використовується до прийняття рішень. Вплив власної помилки обурення на точність прийняття рішення можна вивчити у.

Туманний сервер безперервно відстежує $E(\|\Delta\|_F) > Tol_F$, як тільки помилка обурення матриці даних перевищить поріг припустимої помилки обурення, туманний сервер передає оновлений параметр фільтра та закликає всі пристрої IoMT поділитися своєю оновленою моделлю прогнозування. Вищезгадана схема коротко представлена в алгоритмі 2.

Переваги: Запропонована схема мінімізує комунікаційні накладні витрати (розділ 4.11) обмежуючи кількість передач на центральний сервер. Алгоритм 2, тобто, усереднення моделі робить практично неможливим вилучення індивідуальної моделі з усередненої моделі; це забезпечує конфіденційність медичних даних. Крім того, туманний сервер, на відміну від сервера хмар, розташований ближче до джерела, що зменшує затримку.

### 4.10 Оцінка ефективності

У цьому розділі було представлено деякі експериментальні результати, що базуються на реальних медичних даних IoT.

**Algorithm 2: Filter Model Averaging**

1: **for** (true) **do**
2: t= current time
3: **if** $E(\|\Delta\|_F) \le Tol_F$ **then**
4: $\eta_i(t) \leftarrow \sum_{i=1}^{K} \frac{n_k}{n} \theta_i(t-1)$
5: $\hat{Y}(t) \leftarrow \eta_i^T Y(t)$
6: Perform decision making
7: **else**
8: Fog server shares $\delta_i$ with $N_i$
9: Fog server receives (i, $\theta_i^*(t)$, $Y_i(t)$)
9: **end if**
7: **end for**

Результати включають передбачення з використанням усереднення моделі фільтрації сервером туману, графік витрат на зв'язок при зміні параметра локальної фільтрації та загальну масштабованість пропозиції щодо енергоефективності. Експерименти проводяться з використанням набору даних про здоров'я IoT, відомого як дані MHEALTH. Набір даних включає в себе записи руху та життєвих показників десяти добровольців різного профілю при виконанні 12 видів фізичної активності. Для експериментів було розглянуто лише свідчення нагрудного показання датчика акселерометра в грудях, тобто, стовпці 1-3, і тимчасовий ряд гіроскопа на нижній правій руці, тобто, стовпець 1 дані часової низки гіроскопа правої нижньої руки, тобто, стовпці 18-20.

Мається на увазі однорідний параметр фільтра для всіх пристроїв IoMT. Спочатку розподіляємо дані порівну між 50 пристроями IoMT та обчислюємо нормовану припустиму помилку обурення, як показано на рівнянні 4.14.

$$\langle Tol_F \rangle = Tol_F \Big/ \sqrt{\frac{\sum \lambda_i^2}{n}} \tag{4.14}$$

Представляється взаємозв'язок між нормованою припустимою помилкою обурення та параметром рівномірного фільтра на рис. 4.3. На ньому зображено приблизно лінійну залежність, залежність між нормованою припустимою помилкою обурення та параметром локального фільтра. Це також інтуїтивно зрозуміло, оскільки при кожному збільшенні параметра $\langle Tol_F \rangle$ фільтр в IoMT пристроях пропускає більше даних.

Далі було представлено ефективність прогнозування схеми усереднення моделі фільтра (Алгоритм 2) туманного серверу. Через обмеженість простору було запропоновано результати прогнозування для двох різних пристроїв IoMT (рис. 4.4). І локальна, і глобальна фільтрація використовують той самий LMS-фільтр. Наявні складні методи, які забезпечують більш високу точність, не можуть бути використані на туманному сервері, оскільки ці методи повинні бути доступні для локальної обробки на пристроях IoMT, вони повинні застосовуватися для локальної обробки на пристроях IoMT. Враховуючи серйозні обмеження щодо ресурсів потужності та обчислень, складні методи не можуть бути використані пристроями IoMT для локальної обробки.

І тому був побудуваний графік комунікаційних витрат, як функція параметру фільтра. Бачимо, що на мал. 4.5 витрати на зв'язок можуть бути значно знижені навіть при допустимій помилці збурення. На основі оцінки продуктивності, було досягнуто до 95% зниження кількості передач, як показано на рис. 4.6 Це підтверджує данне твердження, що запропонована схема може забезпечити хороший компроміс між комунікаційною ефективністю зв'язку та помилкою власного обурення матриці даних.

Як підсумок, було досліджено масштабованість запропонованої схеми як малої, так великої кількості пристроїв. Енергетична ефективність (η) системи з $n$ кількістю IoMT може бути розрахована як:

$$\eta = \sum_{n} \frac{d_n}{E_n \cdot r_n \cdot TTI} \tag{4.15}$$

Де $d_n$ загальний обсяг завантажених даних, $E_n$ це середня енергія споживана для доставки одного пакета, $r_n$ загальна кількість пакетів даних, які мають бути завантажені у всі пристрої IoMT, а TTI = 1 – інтервал часу передачі, який є постійним для всіх пакетів. З графіка (рис. 4.6) видно, що схема FFF має високу масштабованість. зіставлена з іншими недавніми дослідженнями, такими як AM-DR та добре відомою ARIMA. Судячи з графіка, енергоефективність збільшується зі зростанням кількості пристроїв. Тому запропонована схема може бути поширена у сценаріях масового IoMT.

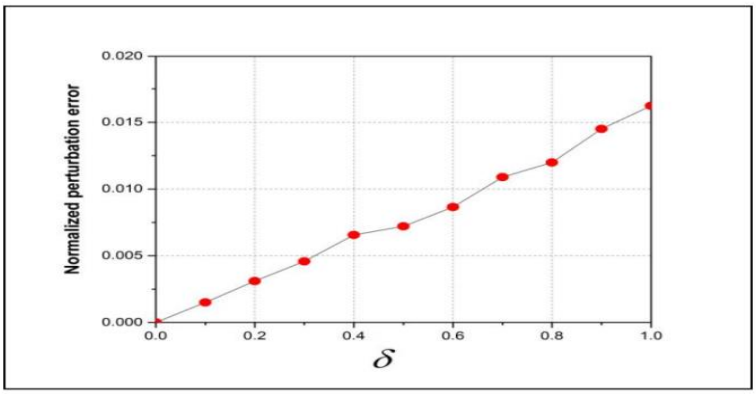



Рисунок 4.3 - Нормована припустима помилка обурення як функція від η

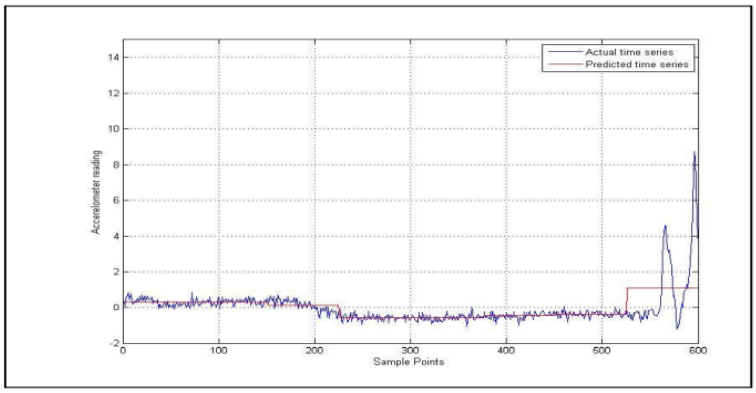
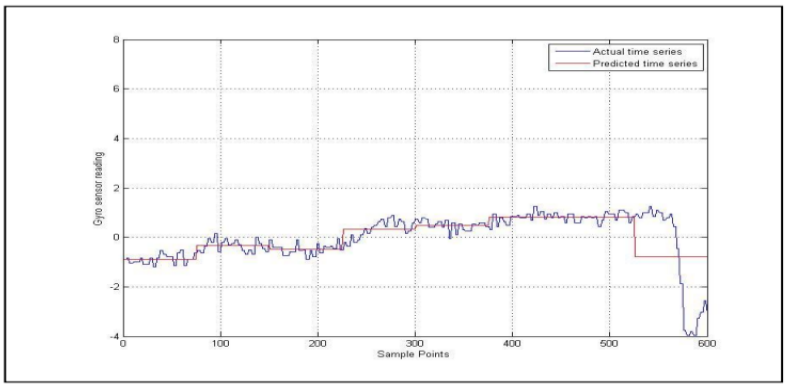



Рисунок 4.4 - Ефективність прогнозування показань датчика акселерометра




Рисуноук 4.5 - Ефективність прогнозування показань гіроскопу

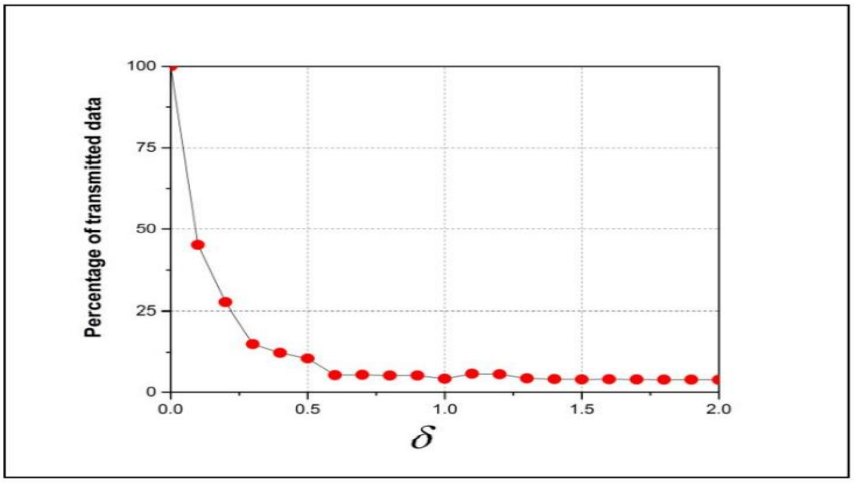



Рисунок 4.6 - Комунікаційні витрати як функція від δ

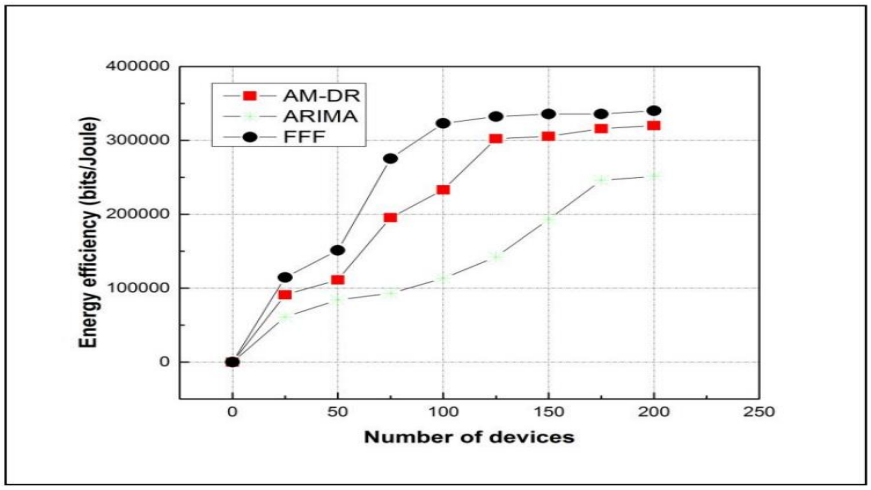



Рисунок. 4.7 - Енергоефективність в залежності від кількості пристроїв

У цьому розділі розглядаються відкриті проблеми, що стосуються енергоефективності, конфіденційності та затримки для інтелектуальної аналітики у охороні здоров'я. У цьому розділі виведено теоретичну верхню межу на обурення власного значення й надалі формулюється зв'язок між локальним квантуванням на пристроях IoMT із глобальною помилкою обурення на сервері туману. На основі теоретичної інфраструктури в цьому розділі пропонуються дві підпрограми, перша для локальної фільтрації на пристрої IoMT та друга для центрального туманного сервера. Запропонована система скорочує 95% комунікаційних накладних видатків. Використовуються матриці обурених даних (передбачених даних) замість використання реальної глобальної матриці для ухвалення рішення, забезпечує найкращу конфіденційність, а мала близькість туманного сервера забезпечує низьку затримку.

# ВИСНОВКИ

В данній магістерській роботі було проаналізовано дані, що генеруються масивними мережами IoT, розглядалися три ключові відкриті проблеми: обмеженість ресурсів IoT-пристроїв, проблема достовірності невизначених даних IoT, мережева затримка та проблеми конфіденційності для роботи IoTMT. В роботі запропоновано чотиришарову архітектуру аналітики даних. та, пропонуються повністю алгоритмічні та теоретичні основи для різних сценаріїв, пов'язаних з IoT, продемонстровані імітаційні дослідження.

Був запропонований підхід до кластеризації даних на основі мобільності. Також була розглянута D2D схема міжкластерної багатошляхової доставки даних. Були проаналізовані необроблені дані датчиків IoT. Була розглянута проблема невизначеності даних, в вигляді запропонованої схеми агрегації даних для невизначених необроблених даних датчиків IoT,

Був розглянутий сценарій IoTMT-аналітики з використанням масивних IoMT-пристроїв. Була представлена нова систему федеративної фільтрації для пристроїв IoMT.

Загальним внесоком даної магістерської роботи є інтеграція рівня агрегації даних IoT методами машинного навчання. У цій магістерській роботі представлені рамки для додатків, які поєднують туманні та хмарні обчислення.

Продемонстровано, що агрегація даних є критично важливим компонентом в архітектурі IoT аналітичної архітектури і може оптимізувати та контролювати продуктивність аналітичної системи.

# ПЕРЕЛІК ПОСИЛАНЬ


1. 1 Nokia Strategic White Paper, “An Internet of Things blueprint for a smarter world,” 2016.
2. С. Хайкін та Б. Уідроу, адаптивні фільтри за найменшим середнім квадратом. 2005 рік.D. Lund, C. Macgillivray, V. Turner, and M. Morales, “Worldwide and Regional Internet of Things (IoT) 2014-2020 Forecast: A Virtuous Circle of Proven Value and Demand,” 2014.
3. Gartner says 4.9 Billion Connected Things; Will be in use in 2015, Gartner 2014,
4. Jimmy Daly, “Algorithms, Big Data and the Importance of Smart Cities,” StateTech Magazine, 2013.
5. N. J. G. Falkner, Q. Z. Sheng, S. Dustdar, A. V. Vasilakos, Y. Qin, and H. Wang, “When things matter: A survey on data-centric internet of things,” J. Netw. Comput. Appl., vol. 64, pp. 137–153, 2016.
6. Parks Associates, “Smart Home Ecosystem: IoT and Consumers,” 2014.
7. Alfonso Velosa, James F. Hines and Hung LeHong, “Predicts 2015: The Internet of Things,” Gartner, 2014.
8. Sarah Silbert, “Intel and San Jose’s smart city will use real-time data to monitor air quality and more,” Engadget, 2014.
9. Deutsche Telekom, “Parking made easy: Smarter parking project in Pisa kicks off”, 2014.
10. S. Verma, Y. Kawamoto, Z. M. Fadlullah, H. Nishiyama, and N. Kato, “A Survey on Network Methodologies for Real-Time Analytics of Massive IoT Data and Open Research Issues,” IEEE Communication Surveys and Tutorials, vol. 19, no. 3, pp. 1457–1477, 2017. 12 Tracy Staedter, “100,000 IoT Sensors Monitor a 1,400-Kilometer Canal in China,” IEEE Spectrum, 2018.
11. Stacey Higginbotham, “The Internet of Trash: IoT Has a Looming E-Waste Problem”, IEEE Spectrum, 2018.

12. 3GPP TS 22.368, “Service Requirements for Machine-Type Communications (MTC),” vol. 0, no. 14.0.0, pp. 1–26, 2014.
13. 3GPP,”TS 22.803,”Feasibility Study for Proximity Services (ProSe),” 3GPP: Sophia Antipolis Cedex, Rel. 12, France, 2014.
14. A. Orsino, G. Araniti, L. Militano, J. Alonso-Zarate, A. Molinaro, and A. Iera, “Energy efficient IoT data collection in smart cities exploiting D2D communications,” Sensors (Switzerland), vol. 16, no. 6, pp. 1–19, 2016.
15. G. Rigazzi, F. Chiti, R. Fantacci, and C. Carlini, “Multi-hop D2D networking and resource management scheme for M2M communications over LTE-A systems”, in Proceedings of the 10th International Wireless Communications and Mobile Computing Conference, pp. 973–978, 2014.
16. K. Schubert and N. Bambos, “Data aggregation for low power wireless devices,” in Proceedings of IEEE Military Communications Conference MILCOM, pp. 1-6, 2016. Thesis for Master’s Degree of CQUPT References 75
17. A. O. Hero, R. L. Moses, N. Patwari, J. N. Ash, S. Kyperountas, and N. S. Correal, “Locating the nodes: cooperative localization in wireless sensor networks,” IEEE Signal Process. Mag., vol. 22, no. 4, pp. 54–69, 2005.
18. M. Azizian, S. Cherkaoui, and A. S. Hafid, “A distributed D-hop cluster formation for VANET,” in Proceedings of IEEE Wireless Communications Networking Conference WCNC, Doha, pp. 1–6, 2016.
19. L. Balzano and R. Nowak, “Blind Calibration of Sensor Networks,” in Proceedings of the 6th Int. Symp. Inf. Process. Sens. Networks, pp. 79–88, 2007.
20. T. Yu, X. Wang, and A. Shami, «Нові туманні обчислення які дозволили скоротити тимчасові дані схема в системах iot», у Proceedings of IEEE Global Communications Conference, стор. 1–5, 2018 рік.
21. Xu, C. Caramanis, and S. Sanghavi, «Надійний PCA через переслідування вибросів», IEEE Trans. Інф. Теорія, вип. 58, № 5, С. 3047–3064, 2012.

22. Ю. Черапанамірі, П. Джейн та П. Нетрапаллі., «На основі порогової стійкості ефективного викиду» PCA», 2017

23. E.J. Candes, X. Li and Y. Ma, «Надійний аналіз головних компонентів?», Journal of ACM, об. 58, № 3, С. 11:1-11:37, 2011.