

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

«На правах рукопису»

УДК 004.056

«До захисту допущено»

В.о. завідувача кафедри
_____ М.В.Грайворонський
“ ____ ” _____ 2019 р.

Магістерська дисертація
на здобуття ступеня магістра

зі спеціальності: 125 Кібербезпека

на тему: Класифікація пристроїв Інтернету речей з використанням методів машинного навчання

Виконав (-ла): студент (-ка) _____ курсу, групи _____
(шифр групи)

Мельник Вадим Васильович

_____ (прізвище, ім'я, по батькові) _____ (підпис)

Науковий керівник к.т.н., доц. Стьопочкина Ірина Валеріївна _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) _____ (підпис)

Рецензент _____
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали) _____ (підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.

Студент _____
(підпис)

Київ – 2019 року

РЕФЕРАТ

Робота обсягом 110 сторінок, містить 15 ілюстрацій, 31 таблиці, 15 формул та 38 літературних посилань.

Метою даної кваліфікаційної роботи є дослідження методів класифікації потоку мережевих пакетів на основі алгоритмів машинного навчання, що можуть бути застосовані для розпізнання класу пристроїв Інтернету речей в мережі, їх удосконалення та адаптація для використання в реальних умовах.

Об'єктом дослідження є процес поширення потоку пакетів пристроїв Інтернету речей в мережі.

Предметом дослідження є алгоритми класифікації потоку пакетів на відповідні класи за допомогою використання методів машинного навчання.

Методами дослідження є дослідження літературних джерел, алгоритми машинного навчання та методи математичної статистики, проведення експерименту із використанням програмних засобів.

Результати роботи викладені у вигляді таблиць, рисунків, а також програмного коду, що реалізує удосконалений алгоритм класифікації пристроїв Інтернету речей, та демонструють працездатність побудованого алгоритму класифікації, використовуючи різні набори реальних даних одночасно.

Результати роботи можуть бути використані на практиці, якщо адміністраторам «розумних» середовищ є необхідність у класифікації пристроїв Інтернету речей для задач з кібербезпеки.

Результати роботи доповідалися на XVII Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених «Теоретичні та прикладні проблеми фізики, математики та інформатики» та прийнято до друку в журналі «Theoretical and applied cybersecurity».

Інтернет речі, локальна мережа, поведінковий профіль пристрою, мережевий пакет, машинне навчання

ABSTRACT

The work includes 110 pages, 15 figures, 31 tables, 15 formulas, 38 literary references.

The aim of this qualification work is to study methods for classifying the network packets flow based on machine learning algorithms that can be used to recognize the class of Internet of things devices in the network, their improvement and adaptation for use in real conditions.

The object of research is the process of disseminating the packets flow of Internet of things devices in the network.

The subject of the research is the algorithms for classifying the packets flow into corresponding classes using machine learning methods.

The research methods are research of literary sources, machine learning algorithms, methods of mathematical statistics, an experiment using software solution.

The results of the work are presented in the form of the tables, figures, program code that implements an advanced classification algorithm for Internet of Things devices and demonstrate the operability of the constructed classification algorithm using different real datasets at the same time.

The results of the work can be used in practice if the administrators of “smart” environments need to classify Internet of things devices for cybersecurity tasks.

The results of the work were presented at the XVII All-Ukrainian Conference of students, graduate students and young scientists “Theoretical and applied problems of physics, mathematics and computer science” and report accepted for publication in the journal “Theoretical and applied cybersecurity”.

Internet of Things, local network, behavior device profile, network packet, machine learning

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів.....	8
Вступ.....	11
1 Аналіз поточного стану Інтернету речей	14
1.1 Поняття Інтернету речей та їх вплив на сьогодення.....	14
1.2 Основні проблеми безпечного використання Інтернету речей	21
1.3 Існуючі підходи до класифікації пристроїв Інтернету речей	29
Висновки до розділу 1.....	38
2 Аналіз ключових компонентів алгоритму класифікації пристроїв Інтернету речей. 39	
2.1 Визначення основних понять	39
2.2 Аналіз характеристик потоку пакетів пристроїв Інтернету речей	42
2.3 Побудова моделі бінарної класифікації та аналіз алгоритмів машинного навчання	52
2.4 Побудова моделі нейронних мереж мультигрупової класифікації	56
2.5 Використання результатів класифікації з точки зору кібербезпеки	61
Висновки до розділу 2.....	63
3 Класифікація пристроїв Інтернету речей	64
3.1 Загальна архітектура	64
3.2 Реалізація алгоритму	66
3.3 Результати дослідження.....	70
Висновки до розділу 3	80
4 Розроблення стартап проекту.....	81
4.1 Опис ідеї проекту.....	81
4.2 Технологічний аудит ідеї проекту	84
4.3. Аналіз ринкових можливостей запуску стартап-проекту	86
4.4. Розроблення ринкової стратегії проекту	95

4.5. Розроблення маркетингової програми стартап-проекту	99
Висновки до розділу 4.....	103
Висновки.....	104
Перелік джерел посилання	106

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

5G (англ. fifth generation) - 5-те покоління мобільних мереж.

ACL (англ. Access Control List) – список контролю доступу до ресурсів комп'ютерної системи.

CNN (англ. Convolutional neural network) – клас згорткових нейронних мереж.

D2D (англ. Device-to-Device Communications) – тип комунікації, в якій пристрої спілкуються між собою безпосереднім образом в мережі.

D2I (англ. Device-to-Infrastructure Communications) - тип комунікації, в якій пристрій ділиться інформацією з численною кількістю пристроїв в мережі.

DBSCAN (англ. density-based spatial clustering of applications with noise) – алгоритм машинного навчання кластеризації даних.

DNS (англ. Domain Name System) – розподілена система перетворення IP-адрес в ім'я хосту.

DoS/DDoS (англ. (Distributed) Denial-of-service attack) – атака, що спрямована на блокування ресурсів комп'ютерної системи.

EC (англ. Edge Computing) – розподілена парадигма обчислення; перенесення всіх центральних обчислень до «краю» мережі, тобто в місці, де відбувається генерація потоків даних.

HTTP / HTTPS (англ. Hypertext Transfer Protocol / Hypertext Transfer Protocol Secure) – мережеві протоколи передачі даних.

IETF (англ. Internet Engineering Task Force) – міжнародне дослідницьке товариство з питань Інтернету.

IoT (англ. Internet of Things, укр. Інтернет речі) – концепція побудови мереж, що складаються з фізичних взаємопов'язаних речей, що мають здатність до генерування, обробки та передачі даних.

ITU (англ. International Telecommunication Union) – Міжнародний союз електрозв'язку.

kNN (англ. k-nearest neighbor method) – метод машинного навчання, що характеризується пошуком k-найближчих сусідів.

LOF (англ. Local Outlier Factor) – метод машинного навчання, що характеризується пошуком аномалій.

LSTM (англ. long short-term memory) – архітектура нейронних мереж довгої короткочасної пам'яті.

MAC (англ. Media Access Control) – унікальний ідентифікатор мережевого обладнання.

MitM (англ. Man in the middle) – атака «людина посередині».

ML (англ. Machine Learning) – алгоритми машинного навчання.

NB (англ. Naive Bayes) – наївний баєсів алгоритм машинного навчання.

NTP (англ. Network Time Protocol) – мережевий протокол синхронізації часу.

OSI (англ. Open Systems Interconnection model) – абстрактна мережева модель для комунікацій і розробки протоколів.

Random Forest – алгоритм машинного навчання, що характеризується побудовою численних випадкових дерев рішень.

RNN (англ. Recurrent neural network) – клас рекурентних нейронних мереж.

Smart Home (укр. «розумний» будинок) – система ефективного управління будівлею, що складається з пристроїв Інтернету речей.

Smart City (укр. «розумне» місто) – система ефективного управління містом, що складається з пристроїв Інтернету речей.

SVM (англ. support vector machine) – метод машинного навчання опорних векторів.

TCP / UDP (англ. Transmission Control Protocol / User Datagram Protocol) – протоколи управління передачі даних.

VLAN (англ. Virtual Local Area Network) – віртуальна локальна комп'ютерна мережа.

Поведінковий профіль пристрою – це вектор характеристик потоку пакетів мережевого пристрою.

Сила зв'язності пристроїв – це показник, що описує наскільки сильно взаємопов'язані пристрої в мережі.

ВСТУП

На сьогоднішній день проблеми пов'язані з розповсюдженням пристроїв Інтернету речей набувають все більшої і більшої актуальності в області кібербезпеки. Кожного дня ми можемо спостерігати за стрімким розвитком таких типів пристроїв, все більше і більше їх долучають до створення розумних середовищ, наприклад, «розумний будинок» або «розумне місто». З впровадженням нових технологій, включаючи 5G мережі, ці пристрої стануть ключовими компонентами як ефективними інструментами в будь-якій сфері людської діяльності.

Однак, оператори розумних середовищ стискаються з великою проблемою розпізнання підключених пристроїв, через причину їх масовості. Таким чином, через неправильне налаштування мережі Інтернет речі стають вразливим місцем. Цей факт підсилюється також недоліками в проектуванні самих пристроїв та реалізації, включаючи стандартні заводські параметри доступу, що можуть бути легко скомпроментовані. Як результат, дані пристрої стають легкими цілями в мережі, за допомогою яких є можливість долучити їх до ботнет мережі, також провести, наприклад, DoS та DDoS атаки, сканування тощо.

На сьогоднішній день, існує низка підходів до захисту пристроїв Інтернету речей, що включають створення політик безпеки, використання типових засобів захисту, таких як: IDS, IPS, міжмережеві екрани. Однак дані підходи стають все менш ефективними в силу масовості пристроїв Інтернету речей та складністю контролювати кожен з них окремо.

Тому, **актуальність** даної роботи зумовлюється необхідністю ефективного алгоритму розпізнання та класифікації пристроїв на відповідні групи для подальшого застосування заходів щодо захисту виокремлених класів Інтернету речей в цілому, що включають: створення на базі маршрутизаторів віртуальних LAN, для забезпечення базової ізоляції класів пристроїв Інтернету речей та необхідних політик безпеки.

Метою даної роботи є дослідження методів класифікації потоку мережових пакетів на основі алгоритмів машинного навчання, що можуть бути застосовані для розпізнання класу пристроїв Інтернету речей, їх удосконалення та адаптація для використання в реальних умовах.

Для досягнення поставленої мети необхідно виконати наступні **завдання**:

- проаналізувати існуючі на сьогоднішній день пристрої Інтернету речей, можливі види атак на інфраструктуру, що складається з даних пристроїв та актуальні напрями забезпечення захисту Інтернету речей для запобігання компрометації від недобросовісних користувачів;
- визначити існуючі підходи на основі машинного навчання до класифікації потоку мережових пакетів пристроїв Інтернету речей, проаналізувати працезданість в реальних умовах;
- визначити існуючі на сьогоднішній день алгоритми машинного навчання та проаналізувати їх ефективність;
- виконати необхідний аналіз потоку мережових пакетів кожного пристрою та, зробивши висновки на базі цього аналізу, побудувати поведінкові профілі для застосування в якості вхідних даних класифікатора;
- розробити необхідний програмний код для проведення експерименту та провести аналіз виконання класифікації на реальних наборах даних.

Об'єктом дослідження є процес поширення потоку пакетів пристроїв Інтернету речей в мережі.

Предметом дослідження є алгоритми класифікації потоку пакетів на відповідні класи за допомогою використання методів машинного навчання.

Наукова новизна: запропоновано удосконалений алгоритм класифікації пристроїв Інтернету речей, який заснований на алгоритмах машинного навчання Random Forest (для бінарної класифікації) і нейронної мережі типу RNN-CNN

(для мультикласової класифікації), що відрізняється від існуючих вищою точністю за рахунок виділення набору найбільш важливих мережевих ознак, що дозволяє відокремити пристрої як з фіксованою так і нефіксованою природою підключень. Також, уперше запропоновано множину універсальних класів пристроїв Інтернету речей, яка є достатньою для їх мережного розмежування та керування ними із застосуванням відповідних політик безпеки.

Практичне значення: результати роботи можуть бути використані адміністраторами «розумних» мереж для класифікації відповідних пристроїв Інтернету речей на окремі групи для забезпечення ефективної безпеки шляхом створення окремих VLAN та необхідних політик безпеки для кожної групи в цілому.

Апробація. Результати роботи доповідалися на XVII Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених «Теоретичні та прикладні проблеми фізики, математики та інформатики» та прийнято до друку в журналі «Theoretical and applied cybersecurity».

1 АНАЛІЗ ПОТОЧНОГО СТАНУ ІНТЕРНЕТУ РЕЧЕЙ

Інтернет речей (IoT) - це нова парадигма, яка швидко набирає силу в сценарії сучасних бездротових телекомунікацій. Основною ідеєю цієї концепції є постійна присутність навколо нас різних речей або об'єктів, таких як датчики, виконавчі механізми, мобільні телефони тощо, які за допомогою унікальних схем адресації здатні взаємодіяти один з одним і співпрацювати в навколишньому середовищі для досягнення загальних цілей.

Безсумнівно, провідна ідея Інтернету речей - це її великий вплив на деякі аспекти повсякденного життя і поведінки потенційних користувачів. З точки зору приватного користувача, найбільш очевидні переваги впровадження Інтернету речей буде відображено як в робочій, так і в домашній сфері. У цьому контексті впровадження розумних будинків, електронної системи охорони здоров'я - це лише кілька прикладів можливих сценаріїв застосування, в яких нова парадигма буде відігравати провідну роль в найближчому майбутньому. Аналогічним чином, з точки зору бізнес-користувачів, найбільш очевидні переваги будуть в рівній мірі помітні в таких областях, як автоматизація і промислове виробництво, логістика, управління бізнесом / процесами, інтелектуальне транспортування людей і товарів.

Однак, насправді, багато складних питань ще належить вирішити. Перш за все, щоб ідея Інтернету речей була широко прийнята, як технологічні, так і соціальні аспекти повинні бути проаналізовані та вирішені.

1.1 Поняття Інтернету речей та їх вплив на сьогодення

Численні визначення поняття «Інтернету речей», що простежуються в дослідницьких співтовариствах, свідчать про сильний інтерес до проблеми Інтернету речей і жвавості дискусій по ній. Переглядаючи літературу, зацікавлений читач може зіткнутися з реальними труднощами в розумінні того,

що насправді означає «Інтернет річ», які основні ідеї стоять за цією концепцією і які соціальні, економічні і технічні наслідки матиме повне розгортання та розповсюдження Інтернету речей. Основною причиною неясності цього терміна є сама назва «Інтернет речей», яка синтаксично складається з двох термінів [1]. Перший орієнтований на мережеве бачення Інтернету речей, а другий - на загальні «об'єкти», які необхідно інтегрувати в загальну структуру. Істотні відмінності у видіннях поняття Інтернету речей виникають через те, що зацікавлені сторони, бізнес-альянси, дослідницькі товариства і органи стандартизації починають підходити до проблеми з точки зору «орієнтованості на Інтернет» або «орієнтованості на річ», в залежності від їх конкретних інтересів. У будь-якому випадку, не слід забувати, що слова «Інтернет» і «Річ», взяті разом, приймають значення, що вводить руйнівний рівень інновацій в сучасний світ інформаційних і комунікаційних технологій.

З точки зору технічної стандартизації, згідно з визначення International Telecommunication Union [2], «Інтернет речей – це глобальна інфраструктура для інформаційного суспільства, що забезпечує розширені послуги за рахунок об'єднання (фізичних і віртуальних) речей на основі існуючих та тих, що розвиваються, інтероперабельних інформаційних і комунікаційних технологій». Поняття речей ідентифікується як «об'єкт фізичного світу (фізичні річі) або інформаційного світу (віртуальні річі), який може бути ідентифікований та інтегрований в комунікаційні мережі» [2]. Фізичні річі існують у фізичному світі, їх є здатність виміряти, вони можуть приводитися в дію і з'єднуватися. Прикладами фізичних речей є навколишнє середовище, промислові роботи, товари та електрообладнання. Віртуальні річі існують в інформаційному світі і можуть бути збережені, оброблені та доступними. Прикладами віртуальних речей можуть бути мультимедійний контент, прикладне програмне забезпечення тощо. Фізична річ може бути представлена в інформаційному світі через одну або кілька віртуальних речей (відображення), віртуальна річ може існувати без будь-якої пов'язаної фізичної річі.

Таким чином, «Інтернет речей» семантично означає «всесвітню мережу взаємопов'язаних об'єктів, однозначно адресованих на основі стандартних протоколів зв'язку» [3]. Фактично, мається на увазі величезна кількість (різномірних) об'єктів, залучених до процесу.

Фундаментальними характеристиками Інтернету речей є [2]:

- Взаємопов'язаність: все може бути взаємопов'язано з глобальною інформаційною та комунікаційною інфраструктурою.
- Неоднорідність: пристрої в IoT є гетерогенними, оскільки засновані на різних апаратних платформах і мережах. Вони можуть взаємодіяти з іншими пристроями або сервісними платформами через різні мережі.
- Динамічні зміни: стан пристроїв змінюється динамічно, наприклад, сон і пробудження, підключення і / або відключення, а також місце розташування і швидкість. Більш того, кількість пристроїв може змінюватися динамічно.
- Величезний масштаб: кількість пристроїв, якими необхідно управляти і які обмінюються даними один з одним, буде як мінімум на порядок більше, ніж пристроїв, що підключені до поточного Інтернету. Співвідношення взаємозв'язку, що встановлюються пристроями, в порівнянні з взаєзв'язком, що встановлюються людьми, помітно зміститься в бік взаємозв'язку, що встановлюються пристроями. Ще більш важливим буде управління згенерованими даними і їх інтерпретація для прикладних цілей. Це відноситься до семантики даних, а також до ефективної обробки даних.

Для побудови мережі Інтернету речей, необхідно щоб кожен пристрій містив наступні технології [4]:

- Технологія унікальної ідентифікації пристрою.
- Для відслідковування змін в середовищі, кожен пристрій повинен містити пристрій аналізу навколишнього середовища в реальному часі, тобто сенсори, датчики тощо.

- Реалізована технологія накопичення та обробки отриманих з сенсорів та інших фізичних пристроїв даних.
- Реалізована технологія передачі даних як по дротовій так і по бездротовій мережах (Wi-Fi, Bluetooth, ZigBee тощо).

Мета Інтернету речей може бути зведена до двох важливих аспектів. Перша мета полягає в тому, щоб перетворити звичайні нам об'єкти в об'єкти з здатністю об'єднуватися один з одним або інтелектуальні об'єкти, надаючи їм деякий інтелект і засоби, що дозволяють їм обмінюватися власною інформацією. Для цього необхідно інтегрувати апаратне забезпечення, яке дозволить об'єкту взаємодіяти з будь-якою платформою. Завдяки мініатюризації апаратного забезпечення, в даний час є можливість розміщувати обладнання будь-куди, що дозволяє включати його в усі види об'єктів. Ця мініатюризація дозволяє Інтернету речей охоплювати все більше і більше практичних застосунків, де це було неможливо в минулому [5].

Інша важлива мета - створити необхідні умови для взаємодії цих об'єктів (всередині однієї і тієї ж мережі) і глобально (Інтернет), дозволяючи підвищити цінність одержуваної інформації. Спілкування є однією з основних задач Інтернету речей в даний час, з огляду на різноманітність об'єктів, платформ (Iotivity, HomeKit, AllJoyn) і методів зв'язку (WiFi, Bluetooth) [5]. Зростання числа існуючих об'єктів збільшує обсяг знеенерованих даних, і основна мета буде полягати в тому, щоб ретельно вивчити цю інформацію і докласти всіх зусиль для отримання різних переваг у світі Інтернету речей.

Термін «Інтернет речей» вперше зв'явилося в 1999 році [6] і з тих пір все більше і більше уваги приділяється даним пристроям. На рисунку 1 можна побачити динаміку росту пристроїв Інтернету речей від самого початку до сьогодення [7].

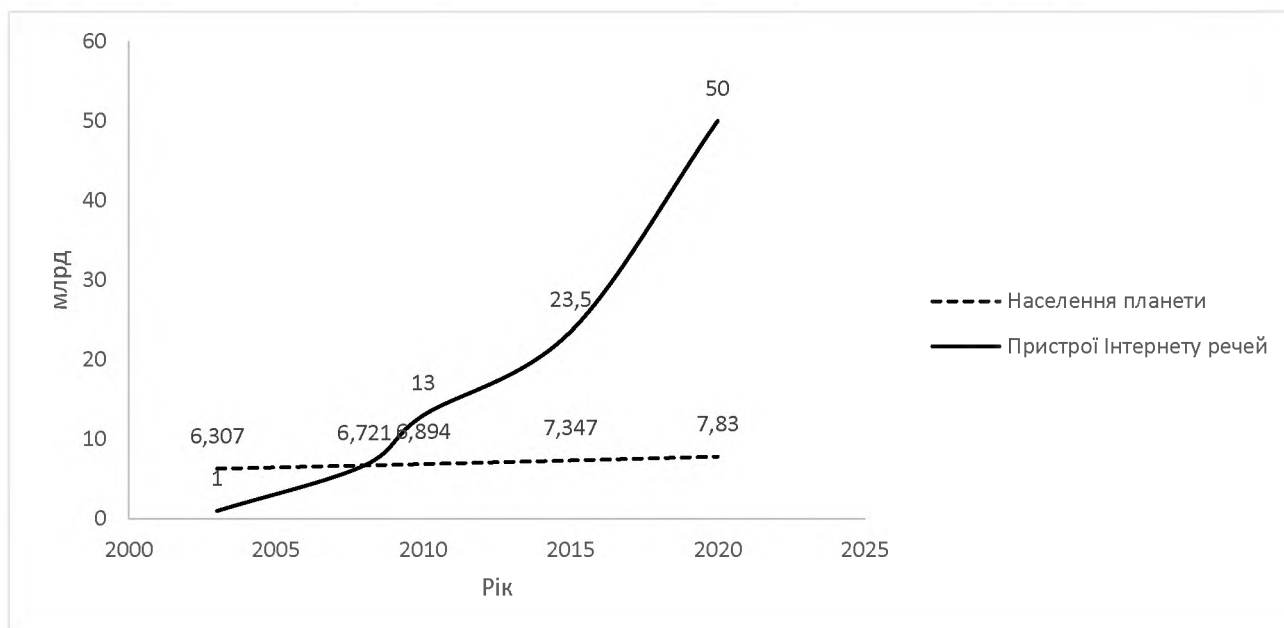


Рисунок 1.1 – Розвиток пристроїв Інтернету речей

Як можна помітити, число існуючих пристроїв Інтернету речей ніколи не припиняло рости і навіть перевищило число існуючих людей на планеті в 2008 році. В останні роки кількість нових пристроїв зростала в геометричній прогресії і якщо сьогодні у нас налічується 15-17 мільярдів пристроїв, то в 2020 році очікується, що в усьому світі буде 50 мільярдів пристроїв, що дорівнює близько 6 пристроїв на людину [7]. Це зростання в основному пов'язане з розвитком технологій, що дозволили мінімізувати розміри обладнання і знизити матеріальні витрати. Іншою причиною є ажіотаж, який ми можемо спостерігати за допомогою зростаючого маркетингу навколо цих продуктів, використовуючи слово «розумний». Це дійсно неймовірні цифри, які змушують нас задуматися, чи дійсно людство готове до цього зростання.

Можливості, що пропонують Інтернет речі, дозволяють розробляти численні практичні застосунки, з яких в даний час розгорнуто лише декілька. У майбутньому з'являться інтелектуальні застосунки для розумних будинків і офісів, розумних транспортних систем, розумних лікарень, розумних підприємств і заводів. Наведемо короткий огляд основних практичних застосувань Інтернету речей.

Охорона здоров'я. За допомогою Інтернету речей є можливість поліпшити якість людського здоров'я та життя шляхом автоматизації деяких основних завдань, що виконуються людьми. У цьому сенсі моніторинг і прийняття рішень можна перенести з людського боку на машинну. Одне з основних застосувань Інтернету речей в охороні здоров'я проявляється в допоміжних життєвих сценаріях. Датчики можуть бути розміщені на обладнанні для моніторингу стану здоров'я, що використовуються пацієнтами. Інформація, зібрана цими датчиками, надається лікарям, членам сім'ї та іншим зацікавленим сторонам в Інтернеті для поліпшення лікування та підвищення оперативності [8]. Крім того, пристрої IoT можна використовувати для моніторингу поточних ліків пацієнта і оцінки ризику нових ліків з точки зору алергічних реакцій і несприятливих взаємодій [9]. Автор в [10] запропонував систему для пацієнтів, які перенесли інсульт. Запропонована система використовує машинні обчислення для виявлення, прогнозування та запобігання падінню пацієнтів, які перенесли інсульт. Вони використовували пристрої Інтернету речей для алгоритмічного навчання моделі виявлення падіння.

Інфраструктура. Інтеграція інтелектуальних об'єктів в фізичну інфраструктуру може підвищити гнучкість, надійність і ефективність роботи інфраструктури. Ці переваги можуть знизити витрати і вимоги до робочої сили, а також підвищити безпеку. Інтелектуальні мережі використовують технологію Інтернету речей для збору даних про споживання енергії та забезпечення доступності даних в Інтернеті. Дані зазвичай формуються в звіти, що показують схеми використання, і включають рекомендації щодо зниження енергоспоживання і витрат [11]. Технології Інтернету речей також використовуються в будинках і офісах. Будинки та будівлі оснащені датчиками і виконавчими механізмами, які відстежують споживання комунальних послуг, контролюють ресурси будівлі, такі як освітлення і системи опалення, вентиляції та кондиціонування повітря, і проводять спостереження для задоволення потреб в безпеці [11-12]. У більш широкому масштабі технології Інтернету речей можуть використовуватися для підвищення ефективності міст. Метою розумних

міст є використання Інтернету речей для поліпшення життя громадян за рахунок поліпшення контролю над трафіком, моніторингу доступності паркувальних місць, оцінки якості повітря і навіть повідомлення про заповнення сміттєвих контейнерів [13-14].

Соціальні аспекти. З огляду на те, що пристрої Інтернету речей можуть взаємодіяти з багатьма об'єктами і навіть з самими людьми, вивчення потенційних соціальних і особистих застосувань Інтернету речей можуть значно поліпшити соціальний аспект людства. Пристрої Інтернету речей надають ряд функціональних можливостей, які можуть сприяти соціальній взаємодії та задоволенні особистих потреб. Одним з можливих застосувань Інтернету речей в соціальному аспекті є взаємодія пристроїв з існуючими соціальними мережами, такими як Facebook або Twitter [15]. Використання пристроїв Інтернету речей для надання інформації про дії та місцезнаходження окремих користувачів може заощадити час користувача. Крім того, ці застосунки, що автоматично збирають і інтегрують інформацію, можуть інформувати людей про соціальні події або іншими важливими фактами [16]. Крім того, мобільні телефони можуть підключатися безпосередньо до інших мобільних телефонів та обмінюватися контактною або службовою інформацією [17].

Однак, збільшення кількості пристроїв Інтернету речей мережі обумовлює величезну кількість даних для обчислення та аналізу в центрах обробки даних, що призводить до перевантаження пропускної здатності мережі. Хоча й витрачається немало сил на вдосконалення мережевих технологій, центри обробки даних не можуть гарантувати необхідні швидкості передачі і час відгуку, що може бути критично для багатьох застосунків. Більш того, периферійні пристрої постійно споживають дані, що надходять з хмари, таким чином змушуючи компанії створювати нові підходи надання контенту для децентралізованого надання даних і послуг, використовуючи фізичну близькість до кінцевого користувача. Тому, одним з таких бачень є нова парадигма Edge Computing [18] (далі EC), основна мета якої полягає в тому, щоб перемістити обчислення від центрів обробки даних до «краю» мережі, використовуючи

інтелектуальні об'єкти, мобільні телефони або мережеві шлюзи для виконання завдань і надання послуг від імені хмари. Переміщаючи сервіси на периферію, можна забезпечити кращий час відгуку і швидкість передачі. У той же час розподіл логіки в різних мережевих вузлах породжує нові проблеми і проблеми.

Наведемо три основні переваги Edge Computing [18]. По-перше, ЕС дозволяє фільтрувати великі обсяги даних, що генеруються пристроями Інтернету речей на «краю» мережі, заощаджуючи час і витрати на передачу даних в центри обробки даних. По-друге, ЕС полегшує швидке прийняття рішень на основі локально оброблених даних, зменшуючи наскрізну затримку (E2E). Це дуже важливо в критично важливих застосуваннях пристроїв Інтернету речей (наприклад, віддалені операції, інтелектуальні мережі, автономні транспортні засоби та відеоконференції), які вимагають надвисокі вимоги до надійності, доступності та низької затримки. По-третє, ЕС поліпшить масштабованість в міру збільшення числа пристроїв Інтернету речей і зменшить витрату заряду батареї через менший час передачі даних між пристроєм і сервером застосунків.

1.2 Основні проблеми безпечного використання Інтернету речей

Проте, існування такої великої мережі взаємопов'язаних об'єктів, безумовно, створює нові загрози безпеки, конфіденційності та довіри, тим самим завдаючи шкоду потенційним користувачам. Інтернет є основою і ядром, що підтримують Інтернет речі, тому майже всі загрози безпеки, які розповсюджуються в Інтернеті, поширюються і на Інтернет речі. Крім того, швидка розробка та широке використання механізмів Інтернету речей в нашому житті означає термінову необхідність усунення цих загроз безпеки перед практичним розгортанням. Хоча, багато компаній заявляють, що їх технології захищені, вони як і раніше залишаються схильними до різних типів атак. Оскільки взаємопов'язані пристрої безпосередньо впливають на життя

користувачів, існує необхідність в чітко визначеній класифікації загроз безпеки і належної інфраструктури безпеки з новими системами, які можуть пом'якшити проблеми безпеки, що стосуються конфіденційності, цілісності даних і доступності Інтернету речей.

Наведемо основні проблемні місця пристроїв Інтернету речей [5], [19]:

- **Безпека:** той факт, що кожен пристрій може відправляти і отримувати інформацію, створює нові проблеми безпеки. Одна з головних причин цих побоювань полягає в тому, що самі пристрої часто знаходяться у вразливих місцях. Безпека повинна бути інтегрована як частина інфраструктури Інтернету речей, оскільки мережі, в яких пристрої взаємодіють, не завжди можуть бути безпечними, і в цьому випадку кожен пристрій стає основною точкою інтересу для зловмисних атак. Тому, проектуючи мережі такого типу, можлива вразливість безпечного використання повинна бути перевірена, а заходи безпеки повинні розвиватися з тією ж швидкістю, що й швидкість зростання кількості пристроїв Інтернету речей. Також, пристрої Інтернету речей, як правило, бездротові і можуть бути розташовані в громадських місцях. Бездротовий зв'язок в сучасному Інтернеті, як правило, стає більш безпечним завдяки шифруванню. Шифрування також розглядається як ключ до забезпечення інформаційної безпеки Інтернету речей. Однак, багато пристроїв Інтернету речей в даний час недостатньо потужні, щоб підтримувати надійне шифрування. А за статистикою, 70% всіх пристроїв Інтернету речей взагалі не використовують алгоритми шифрування. Тому, щоб забезпечити шифрування Інтернету речей, необхідно зробити алгоритми більш ефективні і менш енерговитратні, а також необхідні ефективні схеми розподілу ключів. Додаючи до шифрування, управління ідентифікацією є важливим компонентом будь-якої моделі безпеки, а унікальні ідентифікатори необхідні для пристроїв Інтернету речей. Ці

ідентифікатори можуть використовуватися для встановлення особи в фінансових установах, виявлення незаконної діяльності та інших функцій. Таким чином, гарантуючи, що розумні об'єкти - ті, за кого вони себе видають – необхідна складова для успіху Інтернету речей.

- Конфіденційність: в даний час Інтернет вже стикається з багатьма проблемами конфіденційності. Через особливості пристроїв Інтернету речей і їх використання практично в будь-якій сфері діяльності людства, віддалено, на серверах, зберігається багато приватної інформації, що змушує замислитися кожного, а саме: «чи гарантована моя конфіденційність?», «чи можуть компанії та будь-хто сторонній отримати доступ до моїх особистих даних?». Розпізнавання голосу використовується на пристроях, таких як Amazon Echo, Google Home, Apple HomePod, які відправляють записи нашого голосу на хмарні сервери. Однак, існують пристрої, що здатні виконувати дану операцію і локально, безпосередньо в пристрої, що призводить до потенційної вразливості прослуховування внаслідок безпосередньої атаки на цей пристрій. Пристрої такого типу можуть безперервно аналізувати розмови, маючи справу з дуже чутливими даними. Збір цієї інформації виявляє правові та нормативні проблеми, що стосуються законів про захист даних і конфіденційності, які необхідно удосконалити. Таким чином, права власності на дані, зібрані зі смарт-об'єктів, повинні бути чітко встановлені. Власники даних повинні бути впевнені, що дані не будуть використані без його згоди, особливо в момент передачі на обробку. Політика конфіденційності може бути одним з підходів до забезпечення конфіденційності інформації. Смарт-об'єкти та зчитуючі пристрої Інтернету речей можуть бути оснащені політиками конфіденційності. Таким чином, коли об'єкт і користувач вступають у контакт, кожен може перевірити політику

конфіденційності, що стосуються спільної сумісності, перш ніж ініціювати взаємодію.

- Функціональна сумісність: це одна з основних проблем в світі Інтернету речей, що стосується автоматичної інтеграції різних пристроїв Інтернету речей. Сьогодні, існує безліч протоколів зв'язку, що використовуються пристроями і різними платформами для забезпечення необхідної інфраструктури зв'язку між пристроями. Однак, відсутність унікального стандарту дозволяє кожному пристрою використовувати свою власну реалізацію, і, отже, сумісність Інтернету речей продовжує залишатися великою проблемою.

Описані вище проблеми стають ще більш масштабнішими в силу розвитку децентралізованого розподілення хмарних обчислень, тобто Edge Computing. Відповідно до цього, управління забезпечення безпеки кожної створеної мережі пристроїв Інтернету речей покладається не на централізоване керівництво, а на локальне управління.

Кожна мережа Інтернету речей зазвичай містять суміш пристроїв різноманітного походження, основне призначення яких може суттєво відрізнятися одне від одного. Як правило, ці мережі налаштовуються з використанням єдиного шлюзу, щоб забезпечити доступ в Інтернет для всіх підключених пристроїв. Шлюз пропонує базові функції безпеки, такі як фільтрація MAC / IP. Однак ці функції зазвичай не настраюються користувачами, а в деяких випадках і адміністраторами мереж [20]. Якщо в мережі встановлена будь-яка система виявлення вторгнень або брандмауер, вона фільтрує тільки вхідний і вихідний трафік і розглядає всі підключені пристрої в мережі як безпечні і довірені.

Розглядаючи випадок пристроїв Інтернету речей описане вище припущення про надійність пристроїв не виконується, оскільки досить легко використовувати вразливі пристрої Інтернету речей і, таким чином, отримати доступ до всієї локальної мережі. Через, в більшості випадків, базові налаштування пристроїв Інтернету речей, зловмиснику не складає зусиль

увірватися в мережу (ним може виявитися будь-хто зацікавлений), і, як наслідок, отримуючи невинуватий доступ для виконання будь-якого типу атаки на інші пристрої в тій же мережі. В загальному, ці атаки можуть бути класифіковані як [21]:

1. Сканування мережі: ці атаки використовуються для розпізнавання будь-яких служб TCP і UDP, які працюють на цільових машинах, і для визначення того, який тип фільтрації трафіку виконується в мережі. Мережеве сканування також можна використовувати для визначення прошивки, виконуваної на цільовому пристрої. Ці атаки зазвичай використовуються для сканування цільових вузлів перед запуском спеціальних атак на відскановані цілі. Зазвичай використовуються варіанти атак мережевого сканування, що включають в себе сканування адреси і сканування порту.
2. Підвищення привілеїв. Після того як ціль ідентифікована і відсканована, зловмисник намагається отримати до неї привілейований доступ для розгортання шкідливого коду. Багато пристроїв Інтернету речей використовують полегшену Linux-подібну систему в якості прошивки, і тому зловмисник може спробувати викликати командний рядок для отримання root-доступу. Зловмисник також може використовувати стандартні заводські налаштування або спеціальні пристрої-експлойти для отримання привілейованого доступу. У разі успіху зловмисник може завантажити шкідливий код і виконати необхідні зміни стану, щоб скомпрометувати цільовий вузол.
3. Людина по середині (MitM). Зловмисник, підключений до користувальницької мережі, може переглядати весь трафік в мережі. Він може використовувати схеми зв'язку законних пристроїв для проведення атак повторного відтворення. Наприклад, зловмисник може відтворити трафік, перехоплений від зв'язку між смартфоном і датчиком двері будинку, щоб пізніше відкрити двері будинку без

відома користувача. Атаки типу «людина по середині» мають серйозні наслідки для безпеки та конфіденційності, тому що вони можуть використовуватися для крадіжки даних користувача і злому потенційно важливих пристроїв.

4. Крадіжка даних. Інтелектуальні пристрої Інтернету речей збирають багато даних про користувачів, що можуть містити багато конфіденційної інформації про них. Як правило, користувачі не мають прямого контролю над тим, як ці дані збираються і передаються. Зловмисник може скомпрометувати призначені для користувача пристрої, щоб вкрасти ці дані і використовувати їх для цільових атак.
5. Ботнети. Ботнети зазвичай складаються із заражених пристроїв, встановлених всередині периферійних мереж. Ці пристрої підтримують нормальний стан своїх операцій до тих пір, поки сервер управління і команд не проінструктує їх почати атаку на певні цілі. Атаки з розподіленою відмовою в обслуговуванні (DDoS) - це типовий приклад того, як здавалося б безпечні для користувача пристрої використовуються для запуску атак в безпрецедентних масштабах.

Неминуче швидке зростання кількості пристроїв Інтернету речей створює операційну проблему – адміністратор стикається з проблемою ідентифікації Інтернет речей, а саме, які пристрої Інтернету речей підключені і чи працюють вони нормально. Це може статися через те, що різні підприємства або органи влади залучені в управління активами. Наприклад, датчики освітлення можуть бути встановлені місцевою радою, датчики каналізації та сміття - відділом санітарії, а камери - місцевим відділенням поліції. Управління ресурсами Інтернету речей різних відділів може бути обтяжливою роботою, а також схильною до помилок. І тому це ускладнює визначення того, які пристрої Інтернету речей працюють в мережі в будь-який момент часу. Відсутність «видимості» пристроїв Інтернету речей може дуже ускладнити адміністратору

пошук і усунення проблем в їх інфраструктурі «розумний будинок / місто» і може стати особливо катастрофічною, коли кібератаки порушили критично важливу інфраструктуру. Ще одним важливим фактом є те, що більшість користувачів пристроїв Інтернету речей, будуючи свої розумні середовища, не переймаються питанням захисту підключених пристроїв. Оскільки в таких середовищах всі пристрої під'єднані до однієї точки доступу, виникає критична проблема стороннього вторгнення, оскільки, знову ж таки, користувачі не замислюються над зміною стандартних налаштувань облікового запису кожного пристрою через кількісний чинник даних пристроїв. Таким чином, це є вхідною точкою для зловмисників.

Майбутні користувачі на «краю» стають все більш вразливими, так як все більше і більше пристроїв Інтернету речей використовуватимуться для передачі інформації. Підвищена доступність даних для користувача в ЕС і Інтернету речей може створити безліч можливих способів порушення конфіденційних даних. З ростом інтелекту інтелектуальних пристроїв може з'явитися вірогідність того, що один пристрій Інтернету речей зрадить неявну довіру іншого пристрою, бо як правило, пристрої Інтернету речей запрограмовані на автоматичну довіру іншому підключеному пристрою і обмін даними без процесу перевірки. Якщо всі пристрої спочатку довіряють один одному і обмінюються даними, тоді важко визначити, чи існує підозрілий пристрій. Це може створити велику проблему, особливо через відсутність безпечних периметрів в мережі. Також у системах ЕС складно ідентифікувати, аутентифікувати і авторизувати пристрій і дані, що він генерує, через наднизьку затримку в кілька мілісекунд [22].

На сьогоднішній день існують численні випадки атак на пристрої Інтернету речей. Так, наприклад, в 2017 році за допомогою торгових автоматів, що знаходилися на території університету, таку атаку зазнав університетський кампус [23]. Як результат, було пошкоджено 5000 пристроїв Інтернету речей. Однією з найбільших атак на пристрої Інтернету речей була мережа ботнету Mirai [24], створена в 2016 році. Основною особливістю такого масштабного ботнету було використання найтиповішої вразливості більшості пристроїв

Інтернету речей. Ця вразливість полягала в використанні стандартного, встановленого виробником пристрою, пароля доступу до облікового запису пристрою Інтернету речей. В подальшому, на базі даного ботнету було створено чимало інших, що вдало використовувалися проти різних організацій всього світу.

Дана робота спрямована на створення ефективного інструменту ідентифікації та класифікації пристроїв Інтернету речей для забезпечення необхідного рівня безпеки кожного з них. Досягається це шляхом характеристики трафіку Інтернету речей на мережевому рівні і використання отриманої характеристики для ідентифікації та класифікації пристроїв Інтернету речей в реальному часі, а також для виявлення аномальної поведінки. Своєчасне отримання «видимості» пристроїв Інтернету речей має першорядне значення для оператора, якому доручено забезпечити захист активів у відповідних сегментах безпеки мережі, а також для забезпечування необхідної якості обслуговування і швидкого реагування на будь-яке порушення цілісності, як наприклад, переміщення в карантин при порушенні роботоздатності. Для забезпечення безпеки мережі адміністратори можуть заборонити використання певних типів пристроїв (наприклад, камер в захищеному об'єкті). Для гарантії якості обслуговування, трафік від різних типів пристроїв може бути розподілений за пріоритетами, наприклад, мережеві потоки пакетів пристроїв охорони здоров'я можуть мати пріоритет над потоками пакетів розважальних пристроїв в періоди великого навантаження на мережу. Особливу необхідність в правильному розподіленні пристроїв Інтернету речей будуть потребувати мережі на «краю», тобто Edge Computing. Адже кількість пристроїв в кожній окремій обчислювальній мережі може динамічно змінюватися, як наприклад, система автоматизованого керування транспорту в розумних містах, в якій автомобілі динамічно переходять з мережі в мережу. А швидка та точна ідентифікація допоможе забезпечити безпеку дорожнього руху, а саме від стороннього вторгнення в управління.

1.3 Існуючі підходи до класифікації пристроїв Інтернету речей

На сьогоднішній день велику увагу приділено до підходів для класифікації пристроїв Інтернету речей. Основні методи класифікації пропонують різні техніки ідентифікації, більшість з них ґрунтуються на аналізі потоків пакетів по кожному з них.

Так в [25] запропоновано метод виявлення пристроїв IoT за допомогою аналізу інтернет-трафіку в поєднанні зі знанням доменних імен серверів, що використовуються пристроями Інтернету речей для комунікації. Ці доменні імена включають виробників Інтернету речей, з якими пристрої зв'язуються. За основною ідеєю, даний метод виявляє пристрої Інтернету речей як з загальнодоступними IP-адресами, так і ті, що знаходяться в NAT, зберігаючи при цьому конфіденційність користувачів пристроїв Інтернету речей шляхом вилучення мінімальної інформації (анонімних IP-адрес) з трафіку.

Основним припущенням даного підходу є те, більшість пристроїв Інтернету речей регулярно обмінюються даними з серверами, що належать їх виробникам. Якщо нам відомі ці сервери, то є можливість ідентифікувати пристрої Інтернету речей спостерігаючи за трафіком обміну даних з сервером. Оскільки імена серверів в більшості випадків є унікальними для кожного пристрою, то можна встановити й тип конкретного пристрою.

Для збору інформації про імена серверів, що використовуються в комунікації, вилучаються всі DNS пакети з мережевого потоку і фільтруються за запитом типу A. Далі серед отриманих імен доменів відсіюються ті, сервера яких належать третій стороні та людсько-подібним доменам. До третіх сторін відносяться ті, що пропонують будь-які допоміжні сервіси, як наприклад, синхронізації часу тощо. До людсько-подібних належать ті, в комунікації яких використовуються протоколи зв'язку, що легко можуть представлятися для людини, тобто HTTP та HTTPS. Необхідно зазначити, що для фільтрації доменних імен третіх сторін використовується основне доменне ім'я (тобто

останні два слова, розділених останньою крапкою). Таким чином, якщо повне доменне ім'я не є піддоменом основного доменного імені виробника, то воно вважається таким, що належить третій стороні.

MAC також може бути корисним для визначення виробника пристрою Інтернету речей. Так, автори в [26] зробили спробу зв'язати діапазони виробників MAC з моделями, але їх методика часто була неефективною через відсутність регуляризації в цій області.

Основним і найголовнішим недоліком описаних вище підходів є те, що один виробник може представляти продукцію, що належить до різних класів пристроїв. Наприклад, виробник Samsung може представляти різного роду продукцію, включаючи від телефонів до мережевого обладнання. Таким чином, один виробник може надавати один і той же сервіс для різного обладнання, як наприклад (Samsung SmartTV та Samsung Galaxy).

Методика, що базується на використанні інформації поля user-agent користувача, була запропонована в [27] для класифікації пристроїв Інтернету речей. Значення користувацького поля user-agent відправляється під час HTTP-запитів і містить короткий опис особливостей запитувача пристрою. Для пристроїв, що використовуються безпосередньо людиною, цей параметр зазвичай має велику довжину, оскільки він описує різні властивості пристрою, такі як розмір екрану, використовувану мову, операційну систему тощо.

Однак параметр user-agent можна спостерігати тільки в незашифрованому трафіку. Внаслідок різних спостережень, було встановлено, що тільки 69,5% пристроїв передають цей параметр у вигляді простого тексту, і аналогічний результат був показаний в [28]. Однак, якщо поглянути з іншої сторони, тільки 30% пристроїв Інтернету речей використовують методику шифрування дани. Таким чином, інші 70% пристроїв ставлять під загрозу конфіденційність даних користувачів.

Повертаючись до огляду запропонованої методити, затримка класифікації висока - ймовірність знайти значення поля user-agent користувача, відправлене пристроєм Інтернету речей протягом 20 хвилин, становить близько 25%.

Ще одним підходом до класифікації пристроїв Інтернету речей є побудова так званого відбитку пристрою за допомогою потоку пакетів.

Так, в [29] запропоновано систему DIoT - автономна самонавчальна розподілена система для виявлення скомпрометованих пристроїв Інтернету речей. Система DIoT ефективно будується за допомогою профілей зв'язку, характерних для кожного конкретного пристрою в незалежності втручання людського фактору, а також з технікою помічення даних, які згодом використовуються для виявлення аномальних відхилень у поведінці пристроїв під впливом зловмисних діянь. DIOT використовує федеративний підхід до навчання для ефективного об'єднання профілів поведінки. Таким чином, DIOT може впоратися з новими виникаючими і невідомими атаками. В результаті було систематично і всебічно оцінено понад 30 пристроїв Інтернету речей протягом тривалого часу і зазначено, що DIOT є високоефективним (95,6% виявлення) і швидким (≈ 257 мс) при виявленні пристроїв, що були скомпрометовані, наприклад, від відомої Mirai атаки.

Даний підхід ідентифікації пристроїв базується на основі періодичного фоновому мережевого трафіку кожного пристрою. На відміну від інших підходів, цей метод може ідентифікувати тип пристрою Інтернету речей в будь-якому стані роботи пристрою, включаючи режим очікування, з постійним часом 30 хвилин. Методика ідентифікації складається з трьох етапів, заснованих на пасивному моніторингу мережевого трафіку через мережевий шлюз:

- Аналіз періодичності потоку пакетів кожного пристрою, встановлення періоду і стабільність цього періоду.
- Побудова вектору особливостей потоку пакетів кожного пристрою (відбитка пристрою), що характеризує тип пристрою, на основі його періодичних потоків.
- Використання побудованого вектора в системі класифікації, яка ідентифікує типи пристроїв.

Для визначення періодичності в потоці пакетів кожного пристрою застосовується перетворення Фур'є і автокореляційна функція. Також, необхідно помітити, що періодичність встановлюється по кожному окремому потоку одного типу пакетів, фіксуючи при цьому лише службові протоколи, такі як APR, DNS, NTP, RSTP тощо, тим самим відокремлюючи від користувацьких протоколів HTTP, HTTPS тощо. Саме службовим протоколам притаманна періодичність. Потік пакетів кожного пристрою ідентифікується відповідно по MAC цього пристрою. Кожен мережевий потік пакетів дискретизується одним значенням за часом, а саме, фіксація наявності пакетів за кожну секунду.

До отриманого часового ряду далі відповідно застосовується перетворення Фур'є для отримання періодів та автокореляційна функція.

З отриманих даних будується наступний вектор особливостей пристрою:

- кількість періодичних потоків;
- кількість періодичних потоків пакетів протоколу конкретного рівня OSI;
- середнє значення періоду потоку;
- стандартне розподілення періодів за потік;
- кількість потоків, що мають один/декілька періодів;
- потоки, що мають незмінний порт джерела відправлення;
- середня частота зміну портів;
- ряд вказаних вище показників для кожного окремого підпотoku всього потоку пакетів.

За отриманими векторами було застосовано класифікацію за допомогою алгоритму машинного навчання k-Nearest Neighbors.

Однак, ретельно проаналізувавши даний підхід, вдалося встановити значні недоліки. А саме, застосувавши алгоритм класифікації в реальному житті, не всі пристрої Інтернету речей мають періодичність в потоці пакетів службових протоколів. На рисунку 1.2 можна побачити реальну картину періодичності для різних пристроїв. Необхідно зазначити, що аналіз трафіку проводився в інтервалі

90 хвилин, так як збільшивши час спостереження, час на обчислення необхідних параметрів вектору значно збільшується, тим самим збільшуючи час ідентифікації пристрою.

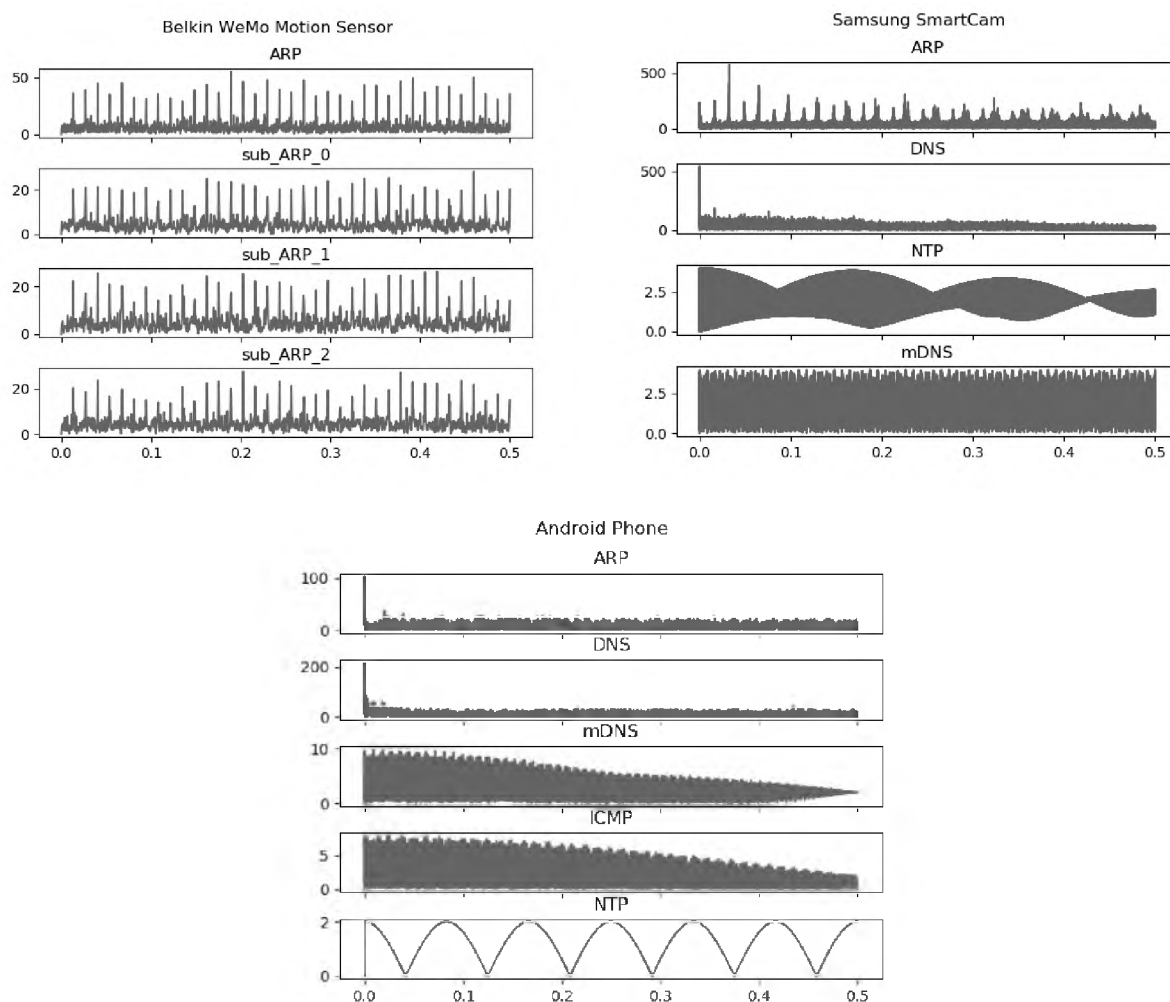


Рисунок 1.2 - Періодичність потоку пакетів для різних пристроїв

Також варто зазначити, що даний алгоритм виявився непрацездатним для нових підключених пристроїв, точність встановлення типу цих пристроїв значно знизилася (до 20%).

Існують також підходи статистичного аналізу потоку пакетів кожного пристрою. Так, в [30] представлено механізм побудови поведінкових профілей кожного пристрою Інтернету речей з точки зору характеру їх діяльності (швидкості трафіку, активного часу, часу простою тощо). І як наслідок,

використовуючи технології машинного навчання, провести класифікацію пристроїв Інтернету речей.

Серед основних параметрів поведінкового профілю, автори виділяють наступні показники:

- час простою;
- час активної роботи;
- середнє значення розміру пакетів;
- середня швидкість потоку пакетів;
- пікове значення швидкості потоку пакетів;
- кількість типів протоколів, що використовується пристроєм для комунікації;
- інтервал використання DNS пакетів;
- інтервал використання NTP пакетів;
- кількість комунікацій (середнє значення унікальних IP-адрес).

Для проведення експериментів, було зібрано набір даних, що включав мережевий потік пакетів від 20 пристроїв Інтернету речей тривалістю 3 тижні в університетському кампусі. В результаті класифікації авторами було досягнуто високої точності класифікації пристроїв, що сягала 95%.

Однак, запропонований метод при практичному застосуванні проявив ряд недоліків. Даний підхід виявився непрацездатний для різних практично змодельованих мереж пристроїв Інтернету речей через не універсальність виділених класів пристроїв Інтернету речей. А саме при застосуванні даного методу для нових, раніше не підключених пристроїв до мережі, точність класифікації значно знизилася. Ще одним недоліком даного методу полягає в тривалості спостереження за пристроями Інтернету речей. Так, за даними авторів, тривалість спостереження для класифікації пристроїв сягало 3 тижнів, що практично не відповідає вимогам ефективної класифікації пристроїв в реальному житті. Дана вразливість методу полягає в зборі показників, що залежать від часу, як наприклад, час простою, та час активної роботи пристроїв.

Таким чином, виникає необхідність в побудові таких показників поведінкового профілю пристроїв Інтернету речей, які не залежали б від часових параметрів роботи пристрою, орієнтуючись тільки на характеристики самих пакетів. Тому, дана робота буде зосереджена на зборі статистичної інформації з найменш необхідної кількості пакетів кожного пристрою за допомогою якої можна ефективно класифікувати пристрої Інтернету речей.

Також, існують підходи до класифікації пристроїв Інтернету речей використовуючи технології нейронних мереж. Так, в [31] запропоновано підхід до автоматичної класифікації пристроїв Інтернету речей для виявлення нових і раніше не підключених до мережі пристроїв. Метод використовує статистичну інформацію, отриману за допомогою аналізу мережеских потоків пакетів пристроїв Інтернету речей, для характеристики атрибутів різних пристроїв. Відповідно до отриманого набору характерних ознак від потоку пакетів застосовується каскадна модель LSTM-CNN для автоматичної ідентифікації семантичного типу пристрою. Особливістю даного підходу є те, що для збору відповідних показників потоку пакетів використовуються пакети заданого інтервалу часу, таким чином отримуючи вектори характеристик за заданий період.

Основними недоліками даного підходу є ресурсоемна запропонована модель нейронної мережі та бідність характеристик потоку пакетів.

Також важливо зазначити й інший напрям забезпечення безпечного використання пристроїв Інтернету речей, що започаткувала міжнародна спільнота IETF (Internet Engineering Task Force) [32]. Відповідний напрям підтримала й компанія Cisco.

Сутність даної пропозиції полягає в тому, щоб виробники пристроїв Інтернету речей публікувати поведінковий профіль свого пристрою. Саме виробники найкраще розуміють, які ресурси будуть потрібні пристроїв після їх підключення до мережі. Наприклад, IP-камера повинна використовувати тільки DNS і DHCP в локальній мережі і обмінюватися даними з NTP-сервером і власним хмарним контролером, але не більше того. Ці вимоги відрізняються в

залежності від пристроїв Інтернету речей різних виробників. Знання вимог кожного з них дозволить нав'язати вузький набір списків контролю доступу (ACL). Таким чином, MUD стає авторитетним ідентифікатором і засобом реалізації політики для пристроїв в мережі.

Виробники повинні як можна краще підходити до задачі створення мережесхем профілів, які будуть потрібні для їх пристроїв. Таким чином, пропозиція IETF MUD надає полегшену модель досягнення дуже ефективної базової безпеки для пристроїв Інтернету речей, просто дозволяючи мережі автоматично налаштовувати необхідний доступ до мережі для пристроїв Інтернету речей, щоб вони могли виконувати свої прямі функції, не надаючи їм при цьому необмежені мережеві привілеї.

На рисунку 1.3 відображено базовий алгоритм роботи запропонованого MUD підходу.

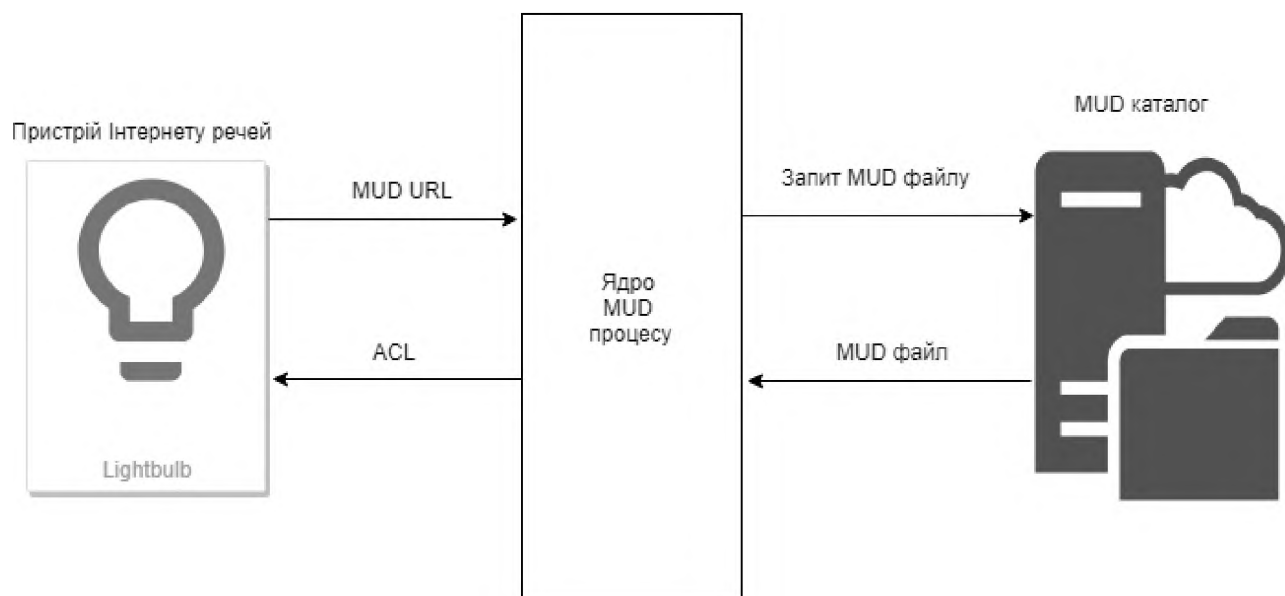


Рисунок 1.3 - Архітектура MUD системи

Як видно з рисунку 1.3, рішення MUD складається з трьох компонентів:

- URL-адреси, який пристрій Інтернету речей генерує при підключенні до мережі.

- Файл, розміщений у віддаленому MUD каталозі, на який вказує URL. Цей файл містить абстрактну політику, яка визначає рівень доступу до мережі, який необхідний пристрою Інтернету речей для нормальної роботи.
- Основний процес, який отримує URL-адресу від пристрою Інтернету речей, отримує файл з файлового сервера MUD і встановлює відповідні елементи управління доступом в мережі для цього пристрою Інтернету речей.

Підхід MUD відрізняється від існуючих підходів кібербезпеки, тому що це відкритий і заснований на стандартах підхід, який використовує предметний досвід виробників пристроїв і, що важливо, що масштабується, для мінімізації зусиль щодо забезпечення безпеки кінцевих пристроїв Інтернету речей.

MUD - це нова парадигма, яку необхідно ще розвивати, і тому сьогодні мало хто розуміє, як саме виробники повинні розробляти поведінкові профілі своїх пристроїв Інтернету речей і як організації повинні використовувати ці створені профілі для захисту своєї мережі. Таким чином, виробникам потрібна допомога у створенні та перевірці їх профілів MUD, а операторам мереж - перевірити сумісність профілів MUD з їх політиками організації. На сьогоднішній день, виробники ще не готові застосовувати даний підхід.

Таким чином, у [33] запропоновано методику вирішення поставлених проблем. Конкретний внесок полягає в наступному: спочатку розробляється інструмент, який приймає на аналіз потік пакетів пристроїв Інтернету речей в якості вхідних даних і автоматично генерує профіль MUD для нього. Потім застосовується формальна семантична структура, яка не тільки перевіряє узгодженість заданого профілю MUD, але і перевіряє його сумісність із заданою організаційною політикою. І нарешті, застосовується запропонована платформа для представницьких організацій і окремих пристроїв, щоб продемонструвати, як MUD може зменшити зусилля забезпечення політики безпеки, необхідної для практичного застосування пристроїв Інтернету речей.

Висновки до розділу 1

В даному розділі було розглянуто основні поняття та теперішній стан пристроїв Інтернету речей. А саме: поняття «Інтернету речей» та основні практичні застосування даних пристроїв.

Також, відповідно до неймовірного росту популярності пристроїв Інтернету речей, було виокремлено основні вразливості, що стають ключовими компонентами багатьох кібератак. Було представлено основні типи атак, що спрямовані на дані пристрої та наведено найбільші приклади в історії.

Найбільш популярним напрямком забезпечення безпеки пристроїв Інтернету речей в мережі є створення політик безпеки та застосування таких інструментів як IDS та Firewall. Однак, дані підходи стають складними в реалізації через кількісне завантаження фізичними об'єктами «розумного» середовища.

Таким чином, на сьогоднішній день актуальним залишається питання ефективної класифікації пристроїв Інтернету речей в мережі, для гнучкого ведення політики кібербезпеки.

Існують чимало досліджень, спрямованих на досягнення поставлених цілей, кожне з яких включає аналіз поведінкових профілей пристроїв Інтернету. Однак, провівши необхідний аналіз даних рішень, виявилось що вони є ефективними в межах одного набору даних, що ставить під сумнів їх реалізацію в реальних умовах.

2 АНАЛІЗ КЛЮЧОВИХ КОМПОНЕНТІВ АЛГОРИТМУ КЛАСИФІКАЦІЇ ПРИСТРОЇВ ІНТЕРНЕТУ РЕЧЕЙ

В даній роботі розподіл пристроїв Інтернету речей розглянуто з точки зору бінарної та мультигрупової класифікації. Обидва підходи включають в себе аналіз потоку пакетів кожного пристрою в мережі.

2.1 Визначення основних понять

Запропонований метод може бути застосований до будь-якої локальної мережі, в котрій всі пристрої мають підключення до мережі Інтернет. За умовчужанням, метод класифікації пристроїв може бути встановлений на мережевому шлюзі.

В процесі виконання своїх прямих обов'язків та функціоналу, пристрої Інтернету речей потребують мережевого з'єднання з іншими об'єктами мережі. В якості таких об'єктів можуть виступати віддалені сервери для аналізу надходжуваних даних так і локальні пристрої-об'єкти для передачі будь-якої функціональної або, навіть, важливої інформації. Таким чином, можна відокремити наступні категорії підключення фізичних об'єктів:

- Зв'язок між пристроями (D2D): ця група включає в себе прямий зв'язок фізичних об'єктів в локальній мережі. Даний тип зв'язку дозволяє зменшити часові затримки при обміні даних та надає можливість економити споживання енергії.
- Зв'язок між пристроями і інфраструктурою (D2I): ця група включає зв'язок між пристроями однієї мережі (локальна мережа) і пристроями або службами іншої мережі (віддалена мережа). Даний тип зв'язку використовується для передачі великої кількості даних для подальшого поглибленого аналізу.

Як було визначено в попередньому розділі, пристрої, що мають здатність до мережевого підключення та мають вбудований будь-якого роду передавач (сенсор), можна віднести до Інтернету речей. Таким чином, виникає необхідність в коректному розподілі пристроїв на два глобальні класи. Проаналізувавши моделі загроз на пристрої Інтернету речей, найбільш вразливими стають пристрої, що мають нескладну функціональність. Такі пристрої стають вразливими до девайсів, що мають більш складну функціональність, оскільки одне з призначень останніх – це користувацьке налаштування базового функціоналу вузькоцільових пристроїв. Зазвичай, таке налаштування виконується безпосередньо адміністраторами мереж або звичайними користувачами. Тому, виникає необхідність в виокремленні такого класу пристроїв, що безпосередньо надають інтерфейс комунікації з людиною напряду. Таким чином, можна виокремити наступні класи пристроїв:

- Вузькоцільові пристрої: до даної групи пристроїв відносяться ресурсо-обмежені пристрої, такі як передавачі, камери з обмеженою та/або нескладною функціональністю, розумні входні дзвінки або замки тощо. Зазвичай, пристрої даного типу не надають користувацький інтерфейс, тому будь-яке початкове налаштування можливе за допомогою програмного забезпечення пристроїв з більш складною функціональністю за допомогою дротового або бездротового з'єднання.
- Багатоцільові пристрої: до даної групи належать високотехнологічні пристрої з кращими апаратними ресурсами, до складу яких можна віднести смартфони, планшети, телевізори тощо. Такі пристрої надають користувачу необхідний інтерфейс комунікації (прикладні програмні забезпечення, операційні системи тощо).

Аналізуючи існуючі пристрої Інтернету речей, що присутні на сьогоднішній день на ринку, можна помітити чималу кількість їх типів. Існує чимало досліджень на тему розділення пристроїв на відповідні класи. Однак, більшість з них концентрується на локальному розділенні пристроїв, що підключенні до мережі. Тому, виникає необхідність в розділенні пристроїв на

універсальні класи, що відокремлюють пристрої не за функціональним застосуванням, а за принципом роботи.

Проаналізувавши відповідні пристрої, в даній роботі пропонується використовувати наступні класи пристроїв Інтернету речей та їх типовими представниками.

- **Контролери.** До даної групи пристроїв відносяться високотехнологічні маршрутизатори, контролери, що виступають в ролі посередника між пристроями Інтернету речей та виконує їх контроль для максимальної працездатності. Типовими представниками даної групи є: Samsung SmartThings Hub, Philips Hue Hub, Wink hub, Logitech Harmony Hub.
- **Камери.** До даної групи пристроїв належать ті, головним функціоналом яких є фото-відео зйомка та передача отриманих даних дротовим або бездротовим способом. Типовими представниками даної групи є: Nest Camera, Belkin Netcam, Netgear Arlo Camera, D-Link DSC-5009L Camera, Logitech Logi Circle, Nest Cam IQ.
- **Командні пристрої.** До даної групи пристроїв належать нескладні електронні пристрої, що оточують людство в повсякденному житті, та виконують основні людські команди відповідно до функціональних можливостей. Типовими представниками даної групи є: TP-link WiFi Plug, Belkin WeMo Switch, TP-Link Smart WiFi LED Bulb, WeMo Crockpot, Roomba.
- **Сенсори/датчики.** До даної групи пристроїв належать ті, основним функціоналом яких є збір даних про стан навколишнього середовища та реагування на нестандартні або аномальні явища. Типовими представниками даної групи є: Belkin WeMo motion Sensor, Nest Protect smoke alarm, Netatmo weather station.

- Гаджети. До даної групи належать пристрої з підвищеними функціональними можливостями та націлені на роботу безпосередньо з користувачами. Типовими представниками даної групи є: Android Tablet, iPhone, iPad, Samsung SmartTV.

2.2 Аналіз характеристик потоку пакетів пристроїв Інтернету речей

Після підключення до мережі, пристрої Інтернету речей генерують трафік (вхідний та вихідний) в залежності від певних функцій конфігурації та служб застосунків. Характер генерації трафіку кожного пристрою є унікальним. Таким чином, проаналізувавши істотні особливості по кожному пристрою, будуються поведінкові профілі, що описують характерні особливості. За допомогою цих векторів особливостей, можна ефективно проводити класифікацію за допомогою методів машинного навчання.

2.2.1 Структура потоку мережевих пакетів

Потік мережевих пакетів може бути відображено як наступну послідовність:

$$S = \{P^1, P^2, P^3, P^4, P^5, \dots, P^j, \dots\} \quad (2.1)$$

де P^j – відображає інформацію, отриману з j -того пакету. Кожен запис пакету P^j включає в себе інформацію, що відноситься до даного пакету та включає:

$$P^j = \{t^j, length^j, protocol^j, eth.src^j, eth.dst^j, others^j\} \quad (2.2)$$

де, t^j відображає час фіксації пакету; $length^j$ відображає довжину пакету; $protocol^j$ відображає тип протоколу; $eth.src^j, eth.dst^j$ відображають відповідно MAC адресу джерела пакету та кінцевого отримувача; $others^j$ – вся

інша отримана інформація, що включає IP адреси, встановлені прапорці тощо. Необхідно зазначити, що пакети впорядковані за часом $t^1 < t^2 < \dots < t^j < \dots$

Розглядаючи мережу, що складається з N пристроїв як $d_1, d_2, d_3, \dots, d_j, \dots (1 \leq n \leq N)$, потік трафіку відображається як:

$$S = \{P_{d_1}^1, P_{d_2}^1, P_{d_3}^1, P_{d_1}^2, P_{d_3}^2, \dots, P_{d_n}^i, \dots\} \quad (2.3)$$

де $P_{d_n}^i$ позначає i -й пакет пристрою d_n .

На етапі попередньої обробки даних нам потрібно дослідити і виокремити специфічні послідовності для кожного пристрою і відфільтрувати інформацію з P^j , яка є надлишковою. Кожен пристрій може бути однозначно ідентифікований на основі MAC-адрес. Таким чином, потоки пакетів для конкретного пристрою можуть бути розділені наступним чином:

$$S_{d_n} = \{P_{d_n}^1, P_{d_n}^2, P_{d_n}^3, \dots\} \quad (2.4)$$

2.2.2 Аналіз потоку пакетів

Відповідно до запропонованих раніше алгоритмів, автори виділяють наступні типові характеристиками потоку пакетів:

- Час сну.
- Середнє значення кількості трафіку за активний час роботи.
- Середній розмір пакету.
- Середній показник роз.
- Пікова / Середня швидкість потоку пакетів.
- Час активної роботи, тобто час активної генерації та приймання мережевих пакетів.
- Кількість IP-адрес, з якими встановлює пристрій з'єднання.
- Кількість типів протоколів, що використовується при комунікації.
- Кількість DNS-запитів.
- Інтервал DNS пакетів.

- Інтервал NTP пакетів.

Аналізуючи кожну з функцій, було встановлено, що не всі особливості в цьому списку здатні точно розділити та ідентифікувати пристрій на відповідний клас пристроїв. На рисунку 2.1 показано, що інтервали NTP і DNS кожного пристрою з різних класів відрізняються, і тому, ми не можемо точно встановити тип пристрою за допомогою цих показників. Таким чином, ці особливості потоку пакетів не зможуть надати достатньо інформації для класифікатора.

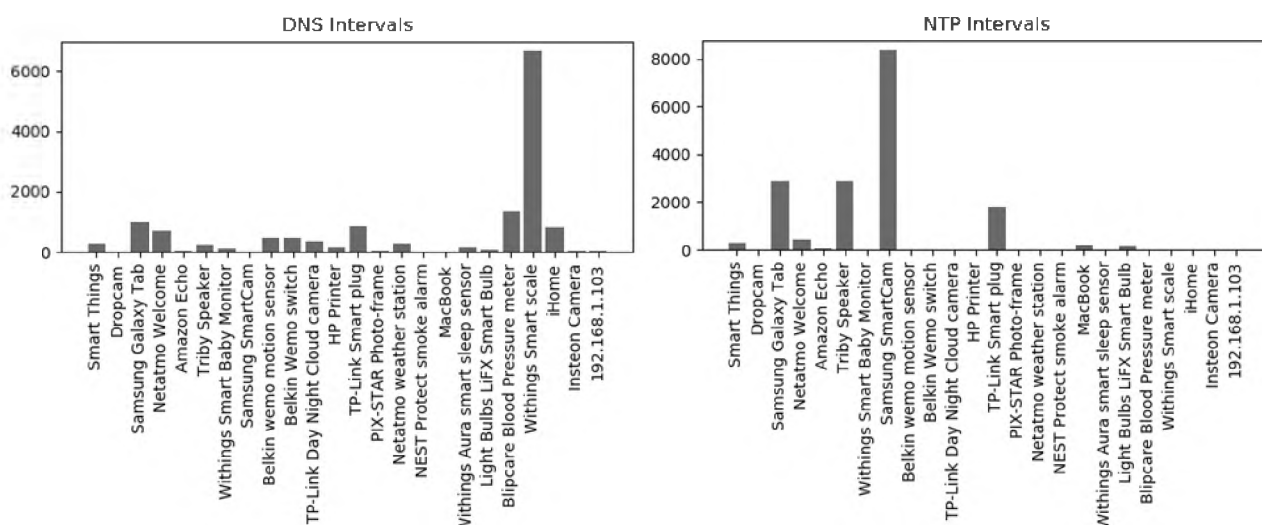


Рисунок 2.1 – DNS та NTP інтервали різних пристроїв в мережі

Проаналізувавши середній показник швидкості передачі пакетів, можна побачити наступний факт.

$$Rate = Active\ volume \div Active\ time, \quad (2.5)$$

де *Active volume* (значення кількості трафіку за активний час) - це загальний сума завантажених і переданих байтів; *Active time* (час активної роботи) - це час неперервної передачі між першим і останнім надісланим пакетом.

З (2.5) видно, що швидкість залежить від обсягу і часу. В результаті швидкість сильно корелює з цими двома показниками. Тому, даний показник має бути відхилений, як надлишковий. Ще одним фактором є фізичний аспект передачі, що залежить від конфігурації самої мережі.

Також, аналізуючи час сну кожного пристрою, ми можемо зробити наступне спостереження. Чимала кількість багатоцільових пристроїв, що підключені до мережі і не зазнають втручань користувача, можуть вести себе так само, як і вузькоцільові пристрої. Наприклад, сервіси операційних систем на мобільному телефоні або планшеті можуть періодично встановлювати з'єднання для таких службових цілей, як: отримувати оновлення, синхронізувати час тощо. Також, можна побачити неоднорідність в роботі пристроїв одного типу. Результат цього спостереження показаний на рисунку 2.2. Так, час сну Dropcam пристрою значно вищий ніж час сну Insteon Camera.

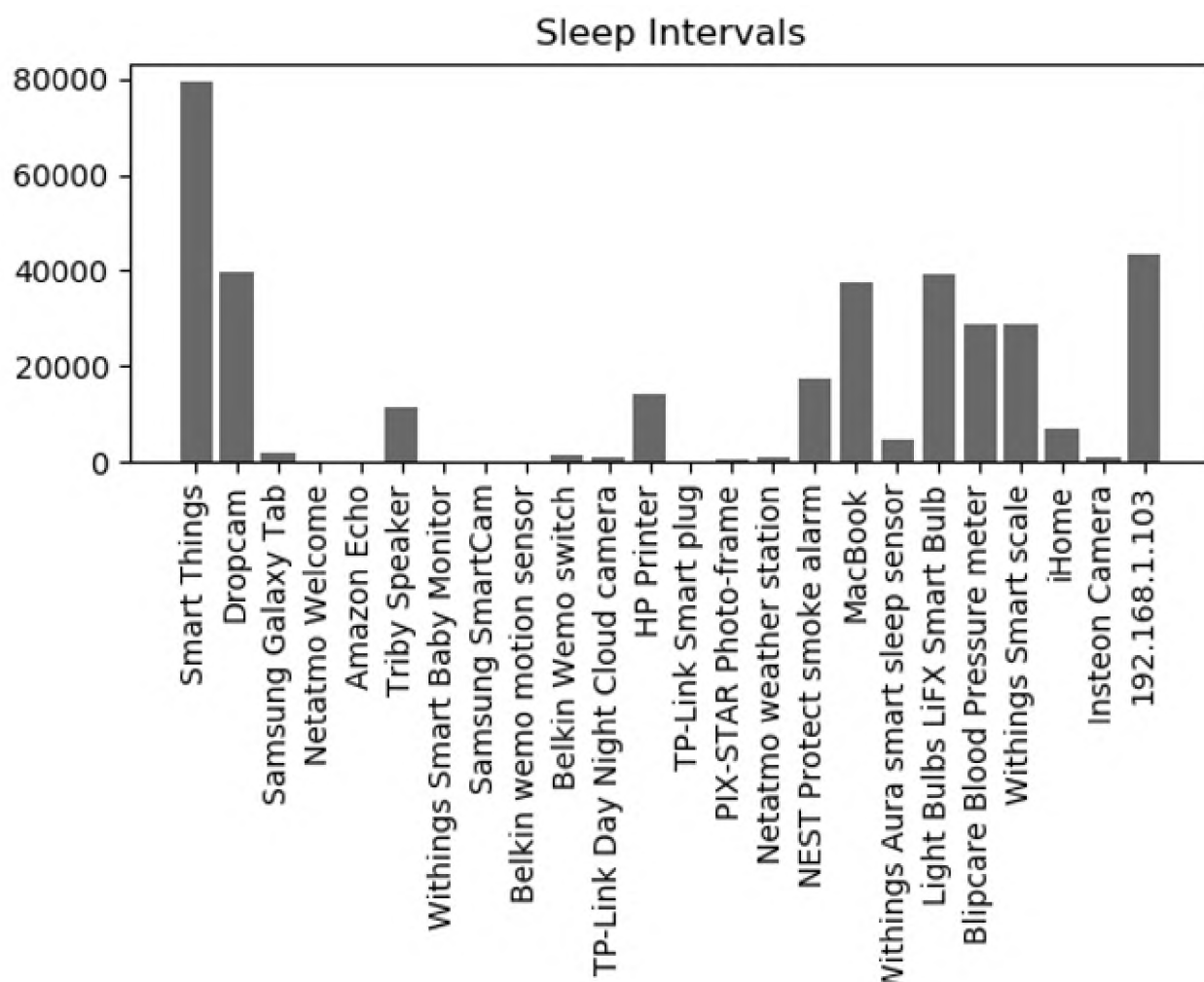


Рисунок 2.2 – Час сну різних пристроїв в мережі

Таким чином, показник часу сну пристроїв Інтернету речей, згідно зі спостереженнями не є ефективним для класифікації, через суперечливість даних. Опрацювавши всі можливі показники, було виявлено наступні факти, що здатні ефективно надати інформацію про пристрій.

Припущення 1. Кількість DNS-запитів від багатоцільових пристроїв Інтернету речей значно перевищує кількість DNS-запитів від вузькоцільових пристроїв Інтернету речей.

Domain Network System (DNS) – один з найбільш популярних протоколів серед Інтернету речей. Даний протокол допомагає пристроям розпізнати веб-сервіси використовуючи прості для людини імена замість числового типу IP-адреси.

Спрощений процес DNS працює наступним чином:

- Браузер, програмний застосунок або пристрій, що виступає в ролі DNS-клієнта, надсилає запит до DNS-серверу провайдера, надаючи явне ім'я хосту, наприклад «example.com».
- Запит приймається розпізнавачем DNS, який відповідає за пошук правильної IP-адреси для цього імені хоста. DNS-розпізнавач шукає в базі даних інформацію про сервер імен хостинг-провайдера, на яких присутній даний сайт, що вміщує коректну IP-адресу для імені хоста в запиті DNS.
- Розпізнавач, що запускається з кореневого DNS-сервера, переміщується вниз по ієрархії top-level domain (TLD) (в даному випадку «.com») до сервера імен, що відповідає заданому домену «example.com».
- Коли розпізнавач досягає потрібної інформації про сервер імен для «example.com», він отримує IP-адресу та інші відповідні відомості і повертає його клієнту DNS. Запит DNS вважається закінченим.
- Клієнтський пристрій DNS може встановлювати з сервером безпосереднє з'єднання, використовуючи коректну IP-адресу.

Найбільш поширені типи записів DNS, підтримувані протоколом DNS:

- Запис сервера імен (NS) - вказує на DNS сервер для вказаного домену.
- Запис зіставлення адрес IPv4 (A) – зв'язує ім'я хосту і його IPv4 адреси.
- Запис адрес IPv6 (AAAA) - зв'язує ім'я хосту і його IPv6 адреси.
- Записи канонічного імені (CNAME) - повертає псевдонім; використовується для перенаправлення на інше ім'я.
- Mail eXchanger record (MX) - вказує поштовий SMTP-сервер для домену.

Для отримання необхідної інформації, найбільш важливими для аналізу є записи типу A та CNAME. При цьому необхідно брати до уваги лише основне доменне ім'я запиту. Ця процедура необхідна для відсіювання всіх дочірніх доменів основного домену як унікальних. Також, по кожному домену збираються всі пов'язані IP-адреси, для встановлення точного числа сторін комунікації пристрою. В даному випадку, різними сторонами комунікації необхідно вважати ті, IP-адреси яких не належать одному домену.

Таким чином, шляхом аналізу потоків пакетів було встановлено, що кількість DNS запитів, за тривалий період спостереження, багатоцільових Інтернет речей значно перевищує кількість DNS запитів вузькоцільових Інтернет речей, причому більшість DNS запитів вузькоцільових Інтернет речей містять імена своїх виробників. Для багатоцільових пристроїв, як телефон, дана картина не є типовою. Для порівняння, результати аналізу відображені на рисунку 2.3.

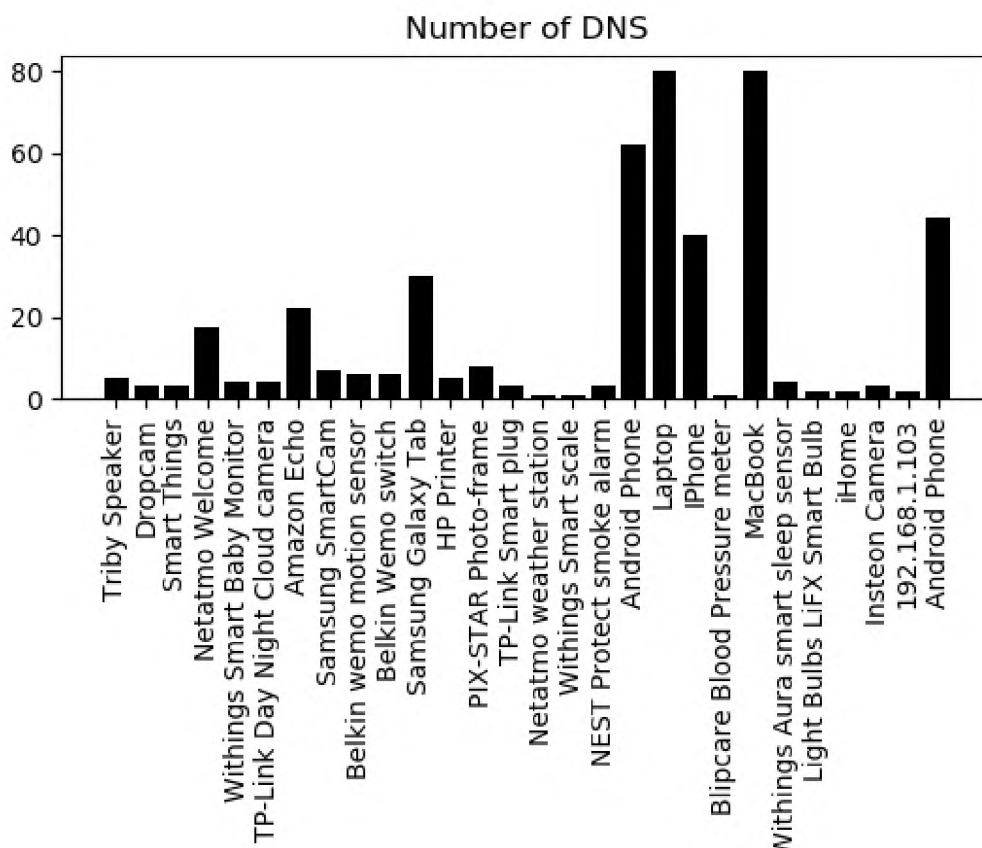


Рисунок 2.3 – Кількість DNS запитів пристроїв в мережі

Таким чином, кількість унікальних DNS запитів та кількість унікальних імен доменів, з якими встановлюють з'єднання Інтернет речі, є важливими показниками, що здатні класифікувати вузькоцільові та багатоцільові Інтернет речі.

Припущення 2. Сила зв'язності багатоцільових пристроїв перевищує силу зв'язності вузькоцільових через той факт, що багатоцільовий пристрій здатен встановлювати з'єднання з багатьма вузькоцільовими пристроями. Сила зв'язності, в даному випадку, – індекс, що вказує, як пов'язані пристрої між собою в мережі.

Як приклад: за допомогою мобільного телефона є можливість контролювати більшість сенсорів, камер, вузькоцільових акустичних систем тощо.

В даній роботі для обчислення сили зв'язності кожного пристрою використовується технологія Google's PageRank [34], що аналізує орієнтований граф зв'язності в мережі, кожне ребро якого має вагу, що дорівнює кількості пакетів згенерованих кожним пристроєм. Сила з'єднання обчислюється за наступною формулою (2.6):

$$PR(u) = (1 - d) + d \sum_{v \in M(u)} \frac{PR(v)}{L(v)}, \quad (2.6)$$

де v, u позначають пристрої, $PR(u)$ - показник PageRank, d - коефіцієнт згладження, який встановлюється зі значенням 0,85, $M(u)$ - це набір вузлів, які мають посилання на вузол u , а $L(v)$ - кількість вихідних посилань з вузла v .

Отриманий результат (таблиця 2.1) аналізу сили зв'язності доводить, що сила зв'язності багатоцільових пристроїв перевищує силу зв'язності вузькоцільових.

Таблиця 2.1 - Сила зв'язності пристроїв Інтернету речей

Пристрій	PageRank
1	2
Securify Almond	0.050131
Philips HUE Hub	0.041608
Samsung SmartThings Hub	0.030454
Android Tablet	0.026212
Apple HomePod	0.024202
Google Home	0.0221
Google Home mini	0.0221
Samsung SmartTV	0.020897
Sonos	0.018622
MiCasaVerde VeraLite	0.016152
Roku 4	0.016152
Roku TV	0.016152

Продовження таблиці 2.1

1	2
iPad	0.016091
Wink 2 Hub	0.014857
Amazon Fire TV	0.014113
Apple TV	0.013064
D-Link DCS-5000L Camera	0.013052
Amazon Echo	0.012345
Belkin Netcam	0.12345
Logitech Hurmony Hub	0.011895
Bose SoundTouch 10	0.01888
iPhone	0.011069
August doorbell cam	0.008807
Belkin WeMo Link	0.008807
Belkin WeMo Motion Sensor	0.008807
Belkin WeMo Switch	0.008807
Canary	0.008807
Caseta Wireless Hub	0.008807
Chamberlain myQ garage opener	0.008807
Harmon Kardon Invoke	0.008807
Insteon Hub	0.008807
Koogeek Lightbulb	0.008807
LIFT Virtual Bulb	0.008807
Logitech Logi Circle	0.008807
Nest Camera	0.008807
Nest Cam IQ	0.008807
Nest Quard	0.008807
Netgear Arlo Camera	0.008807

Припущення 3. Поле User-Agent, що передається в HTTP заголовку, притаманно в більшості випадків для багатоцільових пристроїв. Таким чином, проаналізувавши даний показник, можна встановити тип пристрою та зробити необхідний висновок.

Необхідно зауважити, що дана характеристика є категоріальною та сприяє збільшенню точності у визначенні класу пристрою, і може бути використаною тільки у випадку незашифрованого трафіку. Для класифікатора дана характеристика є бінарною. Для всіх пристроїв за умовчужанням цей показник рівний 0. Однак, якщо для пристрою вдалося розпізнати поле user-agent, цей показник стає рівним 1.

Серед усіх типів протоколів, що використовуються різними пристроями Інтернету речей у період комунікації, виникає необхідність у розділенні їх на протоколи службового типу та користувацько-подібні протоколи.

Службовий тип протоколів – це такий тип протоколів, що використовуються пристроями незалежно від впливу користувачів для службових цілей: синхронізації часу тощо. В даній роботі найбільш популярними службовими протоколами будуть вважатися наступні:

- ARP;
- DNS;
- ICMP;
- mDNS;
- NTP.

Користувацько-подібні протоколи – це такий тип протоколів, що генеруються пристроями внаслідок впливу користувачів. Даний тип використовується для представлення необхідної інформації внаслідок комунікації з людиною. Тому, даний тип протоколів є інтуїтивно зрозумілим. Типовими представниками даної групи є:

- HTTP;
- HTTPS.

Основними характеристиками виділених типів протоколів, як правило, є: середня та максимальна довжина кожного пакету, кількість кожного окремого типу протоколу в групі.

2.3 Побудова моделі бінарної класифікації та аналіз алгоритмів машинного навчання

Таким чином, поведінковий профіль кожного пристрою пропонується будувати з наступних показників характеристики потоку мережевих пакетів:

- Кількість унікальних DNS запитів.
- Кількість типів протоколів, що використовуються в процесі комунікації.
- Тип пристрою за допомогою поля user-agent.
- Кількість унікальних з'єднань в локальній мережі.
- Сила зв'язності пристроїв в локальній мережі.
- Кількість доменних імен, з котрими пристрій має TCP з'єднання.
- Кількість невідомих з'єднань.
- Середній час TCP з'єднання.
- Середній трафік за одну TCP сесію.
- Кількість всіх з'єднань.

При обчисленні кожного з запропонованих показників необхідно враховувати пропускну здатність мережевого пристрою, в даному випадку мережевого шлюзу. Зважаючи на ресурсодоступність таких пристроїв, необхідно максимально ефективно виконувати необхідні обчислення. Таким чином, кожен показник доречно обчислювати інкрементальним способом. Сутність даного підходу полягає в зборі максимально ефективної та необхідної інформації без зайвої надлишковості, що в свою чергу позначається на об'ємі споживання пам'ятовуючого пристрою. Таким чином, кожен з запропонованих показників

слід оновлювати при надходженні нового пакету, шляхом підрахунку середнього значення.

Зазвичай вважається, що методи машинного навчання виправдовують свої застосування в ситуаціях, коли немає точної математичної моделі доступної системи, однак є досить великий обсяг навчальних даних, а також досліджувана система / модель є стаціонарною (тобто повільно змінюється) з плином часу, і чисельний аналіз прийнятний. Методи машинного навчання нещодавно привернули значну увагу до надання керованих рішень для різних складних проблем в системах мережевого зв'язку.

Розгортання реалізації методів машинного навчання в мережевих комунікаціях швидко набирає обертів. Зокрема, для створення самодостатніх і адаптивних мереж, здатних задовольнити вимоги динамічної реконфігурації майбутніх пристроїв і сервісів. Крім того, машинне навчання має значний потенціал для заміни традиційних алгоритмічних рішень на основі математичних обчислень, враховуючи наявність адекватних даних і обчислювальних потужностей.

На сьогоднішній день існує численна кількість методів машинного навчання, які глобально об'єднуються в дві групи: алгоритми контрольованого та неконтрольованого навчання [35].

Під час контрольованого навчання коефіцієнти проміжних ступенів вивчаються шляхом використання раніше доступного набору вхідних даних в парі з їх відповідними бажаними вихідними значеннями. Алгоритми можуть використовувати знання предметної області, а також приклади навчальних даних для вивчення необхідного поведінки і виконання необхідних операцій. Ідеальне застосування контрольованого машинного навчання може бути в сценаріях, в якому є справжній спільний розподіл вхідних і вихідних параметрів, що може бути досліджено з доступного набору знань предметної області. Прикладами є kNN, NB, SVM, Decision Tree.

Однак можуть бути сценарії, в яких математична модель або розподіл невідомі. При напівконтрольованому навчанні є невелика кількість розмічених

навчальних даних, в той час як в повністю неконтрольованому навчанні анотовані дані навчання відсутні. При неконтрольованому навчанні збір доступних зразків вихідних даних використовується для навчання системи, в той час як попередня інформація про бажану відповідь системи недоступна. Неконтрольоване навчання зазвичай використовується для кластеризації та класифікації природних проблем. Дані алгоритми можуть потенційно застосовуватися для виконання широкого кола завдань, пов'язаних з кластеризацією точок, визначенням головних ознак, класифікацією ознак, генерацією конкретних вибірок розподілу. Прикладами таких алгоритмів є K-Means, DBSCAN, LOF.

Для бінарної класифікації пристроїв Інтернету речей за допомогою отриманого поведінкового профілю, застосовуються наступний алгоритм контрольованого навчання.

Random Forest [36] - це гнучкий, простий у використанні алгоритм машинного навчання, який дає стабільно велику точність навіть без налаштування гіперпараметрів. Це також один з найбільш часто використовуваних алгоритмів через його простоту і різноманітність (його можна використовувати як для задач класифікації, так і для завдань регресії). Даний алгоритм працює за допомогою побудови численних дерев прийняття рішень.

Випадковий ліс має майже ті ж гіперпараметри, що і класичне дерево рішень. Однак, випадковий ліс додає моделі додаткову випадковість при вирощуванні дерев. Замість пошуку найбільш важливої функції при розбитті вузла, він шукає кращу ознаку серед випадкової підмножини ознак. Це призводить до великої різноманітності, яке зазвичай призводить до кращої моделі.

Отже, в випадковому лісі алгоритм розщеплення вузла враховує тільки випадкову підмножину ознак. Є можливість зробити дерева більш випадковими, додатково використовуючи випадкові пороги для кожної ознаки, замість того, щоб шукати найкращі можливі пороги (як це робить звичайне дерево рішень).

Ще одна відмінна якість алгоритму випадкового лісу полягає в тому, що дуже просто виміряти відносну важливість кожної ознаки в прогнозі. Алгоритм надає відмінний інструмент для цього, який вимірює важливість особливостей, спостерігаючи за тим, наскільки вузли дерева, які використовують цю особливість, зменшують забруднення всіх дерев в лісі. Він обчислює цей бал автоматично для кожної особливості після навчання і масштабує результати так, щоб сума всіх значень дорівнювала одиниці.

Розглядаючи важливість особливостей, можна вирішити, які особливості необхідно відкинути, оскільки вони не вносять достатнього (а іноді і зовсім ніякого) впливу в процес прогнозування. Це важливо, тому що загальне правило в машинному навчанні полягає в тому, що чим більше у вас особливостей, тим більша ймовірність того, що модель буде страждати від переоснащення, і навпаки.

Загальний алгоритм роботи алгоритму відображено на рисунку 2.4.

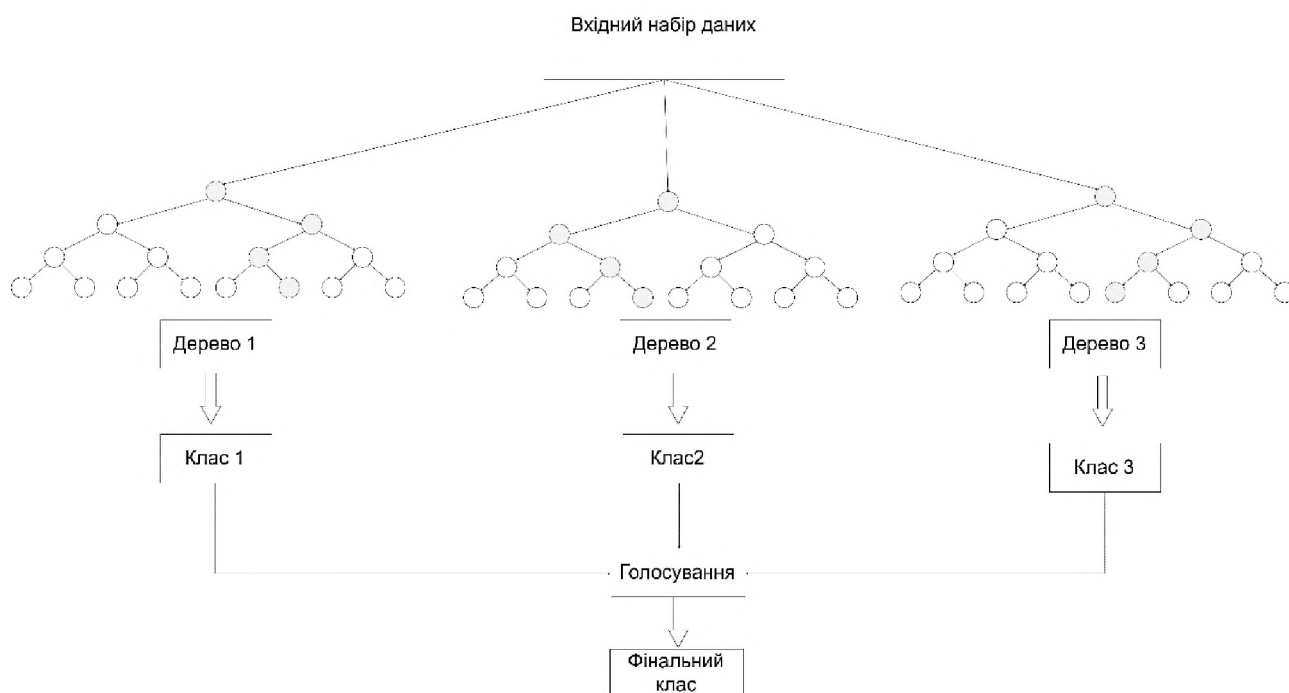


Рисунок 2.4 - Алгоритм роботи Random Forest

Припустимо, що наш набір даних складається з N векторів, розмірність яких задається кількістю ознак M , що входять до них. Також задається параметр

параметр m ($0 < m \leq M$) – кількість вибору випадкових ознак для побудови дерев рішень. Варто зазначити, що в задачах класифікації зазвичай $m = \sqrt{M}$. Також необхідно вибрати загальну кількість дерев, що необхідно побудувати. Усі дерева рішень будуються випадковим чином та незалежно один від одного за такою процедурою:

- Побудуємо випадковий піднабір даних, до складу якого входять вектори з m випадково вибраних особливостей.
- Згенеруємо дерево рішень, що класифікує вектори побудованого набору даних, причому в ході створення чергового вузла дерева вибирається ознака з m випадково вибраних.
- Дерево будується до повного вичерпання побудованого набору даних та не піддається процедурі відсікання.

Після побудови всіх дерев, відбувається голосування за найбільш вподобаний клас та генерується остаточний результат.

2.4 Побудова моделі нейронних мереж мультигрупової класифікації

Алгоритми нейронних мереж є біологічно натхненною структурою обробки даних, яка призначена для вивчення різних операцій на основі спостережуваних даних. Штучні нейронні мережі, як правило, використовуються для розпізнавання будь-яких профілів у вхідних даних шляхом передачі даних через різні рівні модельованих нейронних зв'язків. Штучна нейронна мережа складається з підключених вхідних, прихованих і вихідних шарів нейронів, де кожен вузол (нейрон) виконує операції об'єднання і / або обмеження, а кожне з'єднання виконує операції масштабування. Шари нейронів можуть бути повністю пов'язаними, частково пов'язаними, об'єднаними, з прямим зв'язком, рекурентними тощо. Із зростаючим застосуванням штучних нейронних мереж технології побудови топологій з'єднань між шарами нейронів швидко розвиваються, кілька помітних структур

можуть бути наступні: багаторівневий персептрон (MLP), згорткова нейронна мережа (CNN), рекурентна нейронна мережа (RNN), мережа Хопфілда (HN), GAN, мережа луна-станів (ESN), нейронна машина Тьюринга (NTM) тощо. Ці структури визначають потік даних в мережі; наприклад, в мережі з прямим зв'язком (FFN) кожен нейрон пов'язаний тільки з нейронами наступного рівня, в той час як RNN дозволяє з'єднанням від провідних рівнів бути пов'язаними зворотним зв'язком з попередніми рівнями. Навчання штучного інтелекту - це процес, в якому вивчаються ваги зв'язків між нейронами. Навчання штучних нейронних мереж зазвичай виконується в режимі контрольованого навчання, де раніше доступні дані, помічені з бажаним виходом, використовуються для обчислення помилки для коригування ваг. Помилка може бути визначена кількісно на основі різних метрик, де природним узагальненим квантифікатором є середня квадратична помилка (MSE). Помилка може бути ітеративно поширена назад від виходу до вхідного шару для квантування помилки на кожному шарі і подальшого оновлення ваг.

В якості моделі нейронної мережі для мультигрупової класифікації було обрано RNN-CNN конструкцію, зображену на рисунку 2.5.

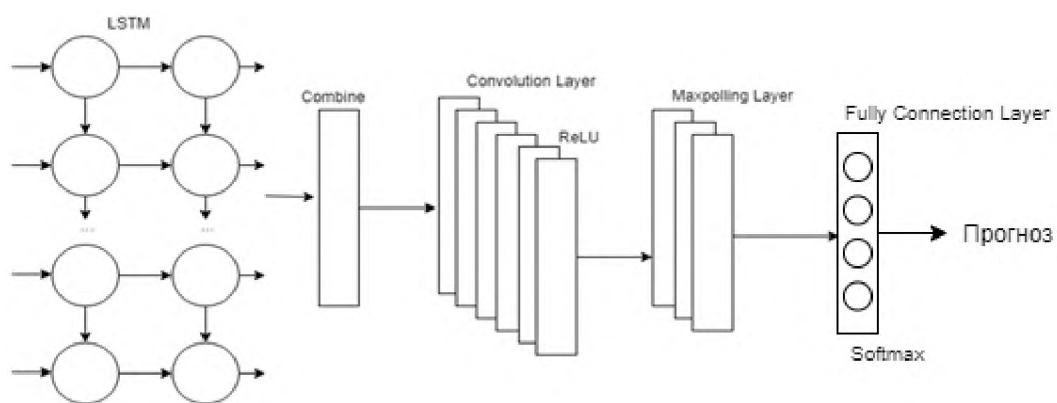


Рисунок 2.5 - Модель нейронної мережі

LSTM - це різновид рекурентних нейронних мереж (RNN), спеціально розроблена для обробки послідовних даних. Вона є дуже ефективною для

багатьох застосунків, починаючи від розпізнавання мови, розпізнавання рукописного введення і виявлення аномалій.

Для базової LSTM клітинки, вхід включає в себе дві частини x_k та h_{k-1} , де h_{k-1} – вихід останньої клітинки на тому ж рівні. Обчислення LSTM клітинки задається наступними рівняннями:

$$g^{(k)} = \tanh(W^{gx}x_k + W^{gh}h_{(k-1)} + b^g) \quad (2.7)$$

$$i^{(k)} = \sigma(W^{ix}x_k + W^{ih}h_{(k-1)} + b^i) \quad (2.8)$$

$$f^{(k)} = \sigma(W^{fx}x_k + W^{fh}h_{(k-1)} + b^f) \quad (2.9)$$

$$o^{(k)} = \sigma(W^{ox}x_k + W^{oh}h_{(k-1)} + b^o) \quad (2.10)$$

$$s^{(k)} = g^{(k)} \odot i^{(k)} + s^{(k-1)} \odot f^{(k)} \quad (2.11)$$

$$h^{(k)} = \tanh(s^{(k)}) \odot o^{(k)} \quad (2.12)$$

Convolution Layer. Виходом LSTM є t векторів. Вони об'єднуються в t стовпців, формуючи двовимірний вектор. Convolution Layer використовує кілька фільтрів, що проходять через двовимірний вектор. Потім вихідний результат цих фільтрів передаються в функції лінійної або нелінійної активації, такі як ReLU, для формування вихідних даних.

Maxpooling Layer. Вихідний сигнал потім безпосередньо подається в Maxpooling Layer. Maxpooling Layer зменшує розмірність входу, вибравши тільки максимальне значення з $n * n$ об'єктів, де $n * n$ - розмір фільтра.

Fully Connection Layer. Після Maxpooling Layer дані знову перетворюються в вектор і передаються в Fully Connection Layer з операцією відсіву перед подачею на останній рівень моделі. Операція відсіву допомагає запобігти перенавчання і домогтися кращого узагальнення. В останньому рівні функція softmax вибирається в якості активної функції для обчислення ймовірностей різних класів. Клас з найбільшою ймовірністю буде остаточним прогнозом для вхідних даних.

Як зазначено в пункті 2.2.1, з мережевого трафіку може бути записано кілька частин інформації, включаючи довжину пакета, часову мітку, тип

протоколу тощо. З цієї інформації може бути додатково вилучено багато додаткових корисних особливостей.

Оскільки кожен пристрій генерує великий обсяг трафіку, сегментація трафіку необхідна перед продовженням дослідження особливостей та навчання моделі. На рисунку 2.6 показано навантаження трафіку, що відноситься до різних пристроїв Інтернету речей.

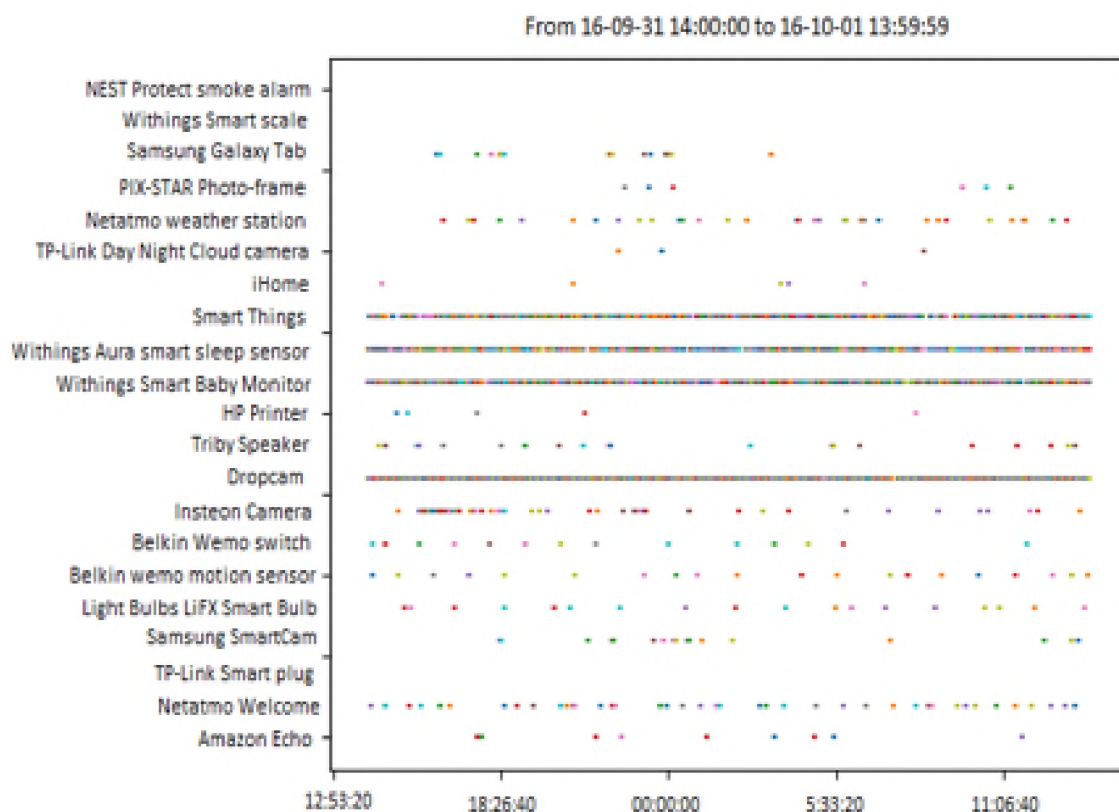


Рисунок 2.6 – Навантаження трафіку різних пристроїв Інтернету речей

Як можна бачити, інтенсивність трафіку варіюється від пристрою до пристрою, а також для різних часових масштабів. Хоча ці пакети містять корисну інформацію, вони можуть бути надмірними і зайвими. Хоча штучні нейронні мережі добре вміють знаходити шаблони, приховані в необроблених даних, обробка кожного пакету окремо зажадає великих витрат часу і обчислювальних витрат. Таким чином, виникає необхідність в сегментації необроблених потоків трафіку, перш ніж обчислювати показники особливостей.

Сегментація трафіку відбувається шляхом розбиття загального потоку пакетів пристроїв Інтернету речей на під потоки фіксованої довжини часу T . Для прикладу, потік пакетів пристрою d_1 може бути виражений як сукупність декількох сегментів тривалістю T (2.13):

$$S_{d_1} = \{sub_{d_1}^{0-T}, sub_{d_1}^{T-2T}, sub_{d_1}^{2T-3T}, \dots, sub_{d_1}^{iT-(i+1)T}, \dots\} \quad (2.13)$$

де $sub_{d_1}^{iT-(i+1)T}$ відображає всі пакети пристрою d_1 в період часу від iT до $(i+1)T$ і також може бути записаний як (2.14):

$$sub_{d_1}^{iT-(i+1)T} = \{p_{d_1}^l, p_{d_1}^{l+1}, p_{d_1}^{l+2}, \dots, p_{d_1}^{l+m}, \dots\} \quad (2.14)$$

Особливості кожного сегменту можуть включати наступні показники:

- Кількість користувацько-подібних пакетів.
- Середня довжина користувацько-подібних пакетів.
- Максимальне значення довжини користувацько-подібних пакетів.
- Кількість службових пакетів.
- Середня довжина службових пакетів.
- Максимальна довжина службових пакетів.
- Кількість унікальних MAC.
- Кількість DNS пакетів.
- Кількість NTP пакетів.
- Кількість унікальних IP адрес.
- Кількість ICMP пакетів.
- Кількість ARP пакетів.
- Кількість HTTP пакетів.
- Кількість HTTPS пакетів.
- Кількість отриманих пакетів.
- Кількість надісланих пакетів.
- Кількість mDNS пакетів.

Як результат, кожен сегмент $sub_{d_1}^{iT-(i+1)T}$ може бути представлений як вектор особливостей $x_{d_1}^i$, і потік мережевих пакетів пристрою d_1 представляється як:

$$S_{d_1} = \{x_{d_1}^0, x_{d_1}^1, x_{d_1}^2, \dots, x_{d_1}^3, \dots\} \quad (2.15)$$

2.5 Використання результатів класифікації з точки зору кібербезпеки

В той час як величезна кількість інтелектуальних пристроїв можуть збагатити наше життя, створені мережі стають вразливими для будь-яких атак, що спрямовані на ці пристрої. Часто це відбувається через відсутність належної підтримки оновлень для виправлення відомих вразливостей безпеки. Немає можливості змусити всіх виробників створювати абсолютно безпечні пристрої, але є деякі речі, які є можливістю зробити, щоб знизити ризики за рахунок скорочення доступу до мережі для пристроїв, які не потребують цього.

Найпростішим варіантом забезпечення базової безпеки пристроїв є налаштування правил мережевого екрану. Коли мережа Інтернет була вперше створена, не було особливої вимоги до безпеки - інтернетом користувалися тільки університети і військові бази, але незабаром знадобився хоч якийсь захист від зовнішнього світу. Таким чином, в середині 90-х були розроблені міжмережеві екрани для обмеження доступу до внутрішньої мережі з Інтернету. Брандмауери гарні для запобігання загрозам у мережі, однак вони не є дуже ефективними, якщо загроза вже знаходиться всередині мережі, що може ховатися в будь-якому пристрої Інтернету речей.

Виникає необхідність виходити з припущення, що всі пристрої вразливі і що всі пристрої потенційно можуть бути скомпрометовані. Згідно з отриманим припущенням, якщо пристрої Інтернету речей представляють серйозну загрозу для інших пристроїв в мережі, то найкращий захист в даному випадку - відокремити їх одне від одного.

Одним із способів поділу мережі пристроїв Інтернету речей - це створення повністю окремої мережі з використанням окремих мережевих маршрутизаторів, модемів, тощо. Але даний спосіб має досить вузьке практичне застосування в домашніх умовах та більшості бізнес-середовищ, оскільки дана система є досить коштовною.

Однак, виходом з даної ситуації є використання віртуалізації мережі, широко використовуваної в корпоративному світі. Віртуальні мережі або VLAN дозволяють створення декількох мереж на одному обладнанні. Один набір фізичних мережевих комутаторів і точок доступу може працювати з декількома віртуальними мережами, кожна з яких відділена один від одного брандмауером.

Наприклад, у корпоративних середовищах VLAN можуть використовуватися для поділу відділів і підтримки контролю і управління потоком даних. Таким чином, дана технологія може використовуватися в розумних середовищах, щоб відокремити пристрої Інтернету речей один від одного. Таким чином, якщо пристрій Інтернету речей скомпрометовано, його доступ до конфіденційних даних контролюється або повністю блокується брандмауером.

Кількість необхідних VLAN задається режимом класифікації пристроїв Інтернету речей. Наприклад, при бінарній класифікації необхідно створити 2 віртуальні мережі, та розподілити прокласифіковані пристрої на створені окремі мережі. Кожна VLAN матиме свою власну підмережу, окремий IP-пул адрес, що дозволяє використовувати в мережі набагато більшу кількість пристроїв. Це також спрощує поділ трафіку, оскільки є можливість застосовувати правила брандмауера до конкретних підмереж. Зазвичай VLAN можуть бути пронумеровані від 1 до 4094. Віртуальні локальні мережі можуть працювати на різних рівнях. Так, у випадку створення VLAN на основі портів, один фізичний комутатор просто ділиться на кілька логічних комутаторів. VLAN вищого рівня працюють шляхом додавання тега (ідентифікатора VLAN) на початок пакетів, що передаються по мережі.

Висновки до розділу 2

Проаналізувавши представлені на сьогоднішній день пристрої Інтернету речей, в даній роботі було виокремлено два основні напрямки класифікації даних пристроїв, а саме:

- a. класифікація пристроїв Інтернету речей на вузькоцільові та багатоцільові;
- b. класифікація пристроїв на універсальні групи.

Виокремлення саме таких запропонованих класів викликано тим, що більшість досліджень спрямовано на виокремлення пристроїв за цільовим застосуванням (тобто пристрої, що обслуговують медичну сферу діяльності тощо), не беручи до уваги базове функціональне призначення (камери, сенсори тощо), що значно погіршує якість класифікації.

Відповідно до запропонованих груп, було проаналізовано ключові характеристики поведінки пристроїв Інтернету речей в мережі та виокремлені ті, що здатні з великою ймовірністю розподілити їх до відповідних класів. Таким чином, запропоновано структуру поведінкових профілів як для бінарної, так і для мультигрупової класифікації. До кожного типу класифікації, було виокремлено найбільш підходящі алгоритми машинного навчання, що добре себе зарекомендували в дослідницькій спільноті. Тому, було запропоновано використовувати наступні технології:

- класичний алгоритм контрольованого навчання Random Forest;
- побудова штучної нейронної мережі.

Отримані результати класифікації є рушійною силою для ефективного управління політикою кібербезпеки. Як приклад, використовуючи технологію створення віртуалізованих LAN, що підтримують сучасні мережеві маршрутизатори, є можливість створення базової безпеки пристроїв шляхом ізоляції кожного класу.

3 КЛАСИФІКАЦІЯ ПРИСТРОЇВ ІНТЕРНЕТУ РЕЧЕЙ

3.1 Загальна архітектура

Запропонований метод класифікації, що заснований на поведінковому профілі пристроїв Інтернету речей, зображений на рисунку 3.1.

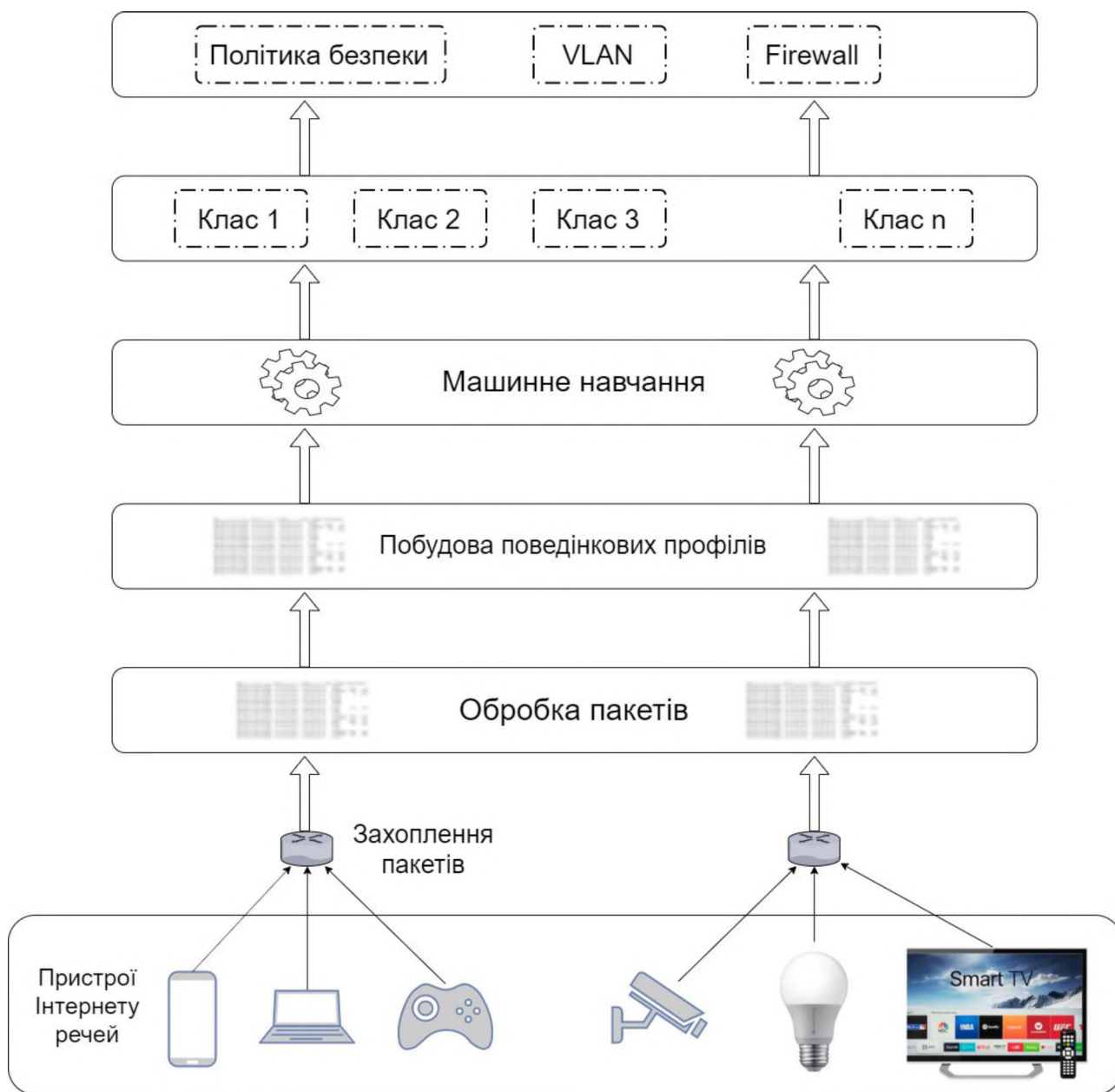


Рисунок 3.1 – Алгоритм класифікації пристроїв Інтернету речей

Першим кроком цього алгоритму є збір даних в локальній мережі. Для виконання основних операцій, запропонований метод необхідно застосовувати на мережевому шлюзі (маршрутизаторі), котрий контролює всі з'єднання пристроїв Інтернету речей. Основні кроки алгоритму включають наступні особливості:

- Основна вимога до першого кроку – для швидкої класифікації пристроїв Інтернету речей, необхідно отримати якомога менше пакетів.
- Другим кроком є аналіз зібраних даних та побудова поведінкового профілю пристроїв Інтернету речей. Для бінарної класифікації профіль складається з одного або декількох векторів особливостей. Для мультигрупової класифікації особливістю є виконання сегментації трафіку пристроїв, для створення поведінкового профілю пристроїв Інтернету речей, що складається з деякої кількості векторів характеристик. Сегментація трафіку може відбуватися за часовим значення (тобто деяким періодом) або необхідною кількістю пакетів. По кожному отриманому векторові далі проводиться аналіз та обчислюється необхідні показники статистики пакетів пристроїв Інтернету речей. Причому, необхідно враховувати обчислювальні можливості центрального пристрою збору даних. Тому, розроблений алгоритм повинен бути оптимізований до зберігання ключової статистики по кожному пакетові, зменшуючи надлишковість, що своєю чергою може призвести до працездатності центрального пристрою. Дана функціональність реалізовується за допомогою інкрементальної статистики.
- Навчання моделі включає в себе обробку побудованих профілів кожного пристрою. У випадку неконтрольованого машинного навчання, немає необхідності в наперед помічених даних. У випадку

контрольованого машинного навчання пропонується використовувати або вже навчену модель, або виконати необхідний збір даних протягом деякого часу від пристроїв в мережі, що знаходяться під контролем. Після виконання необхідного спостереження, отримані дані помічаються необхідним класом пристроїв адміністратором мережі та отримані помічені вектори використовуються в подальшому для навчання моделі.

- Класифікація нових, раніше не помічених в мережі, пристроїв Інтернету речей застосовується до зібраних поведінкових профілів цих пристроїв. При цьому, класифікатор повинен включати навчену модель, описану в попередньому пункті.

У випадку бінарної класифікації, пропонується використовувати алгоритми машинного навчання, а саме Random Forest. Основна перевага цих алгоритмів – швидкий процес навчання та класифікації. Для даних методів класифікації базовий процес аналізу та побудови поведінкових профілів пристроїв Інтернету речей включає інкрементальний аналіз характеристик пакетів. Під інкрементальним аналізом в даній роботі розуміється постійне оновлення середнього показника характеристики кожної окремої складової пакету.

У випадку мультигрупової класифікації пропонується використовувати підхід нейронної мережі, модель якої будується за допомогою RNN-CNN структури. Основна перевага цих алгоритмів – точність прогнозування необхідного результату. Однак недоліками цих методів – невелика швидкість навчання та ресурсоємність.

3.2 Реалізація алгоритму

Так, як основною складовою алгоритму класифікації є побудова поведінкових профілів пристроїв Інтернету речей, що включають в себе

характеристики потоку пакетів, тому реалізація включає в себе парсинг, тобто аналіз пакетів за допомогою програмної реалізації.

Відповідно до цього, було розроблено програмний застосунок, за допомогою мови програмування Python3 та пакету для роботи з трафіком `dpkt`. `dpkt` - модуль python для швидкого, простого створення / аналізу пакетів з визначенням основних протоколів стеку TCP / IP.

Таким чином, за допомогою даного модулю було встановлено наступні показники:

- типи протоколів, що використовується пристроями;
- доменні іменні серверів;
- TCP сесії;
- поле user-agent;
- IP-адреси зв'язків пристроїв.

Для визначення доменних імен серверів, було проаналізовано протокол DNS, що використовується пристроями для встановлення IP адресів по кожному доменному імені. Щоб відобразити дану інформацію, був використаний власний скрипт, який фільтрував DNS пакети типу A та CNAME. Оскільки CNAME – це псевдонім та не надає необхідної інформації про IP адресу, однак даний псевдонім використовується в подальших запитах, зустрічаючи запит типу A. Тому, інформація по кожному доменному імені включає в себе як IP адреси, так і відповідно його псевдоніми. Для збереження даних було використано такий тип даних, як словник, ключовими словами котрого були доменні іменна, а значенням – список IP адрес. Таким чином, встановлюючи чи було з'єднання з сервером з конкретним доменним ім'ям, достатньо перевіряти IP – адресу кожного пакету зі відомими списком IP-адрес.

Для коректного встановлення TCP сесії кожного пристрою мережі, було використано підхід програмного забезпечення Wireshark, а саме: до однієї і тієї ж сесії належать пакети, що мають наступні однакові параметри, як: IP-адреса відправника, IP-адреса одержувача, порт відправлення та порт отримувача.

Також, не мало важливим є відстеження спеціальних TCP – прапорців. Основними TCP-прапорцями є:

- SYN;
- ACK;
- FIN;
- RST.

Відповідно до алгоритму TCP сесії, виокремують три фази:

- Фаза з'єднання;
- Фаза передачі даних;
- Фаза закінчення з'єднання.

Для встановлення TCP – сесії, цікавить перша та остання фаза, а саме фаза з'єднання та закінчення. Відповідно до цього за ці фази відповідають наступні прапорці SYN та FIN. Тому, було розроблено скрипт, для відновлення TCP-сесії, що відшуковує в TCP пакетах SYN прапорець, фіксує відповідні IP адреси відправника та одержувача, і порти надсилання та призначення. Однак, необхідно зазначити, внаслідок закінчення терміну очікування, TCP пакет з прапорцем SYN може надсилатися знову, тому необхідно брати останній надісланий прапорець SYN перед відповіддю. Так, при закінченні сесії, фіксується прапорець FIN TCP пакету. Однак, важливо також зазначити, що в ході сесії можливе різке її обривання, що супроводжується прапорцем RST TCP-пакету. Цей прапорець також береться до уваги.

Також, було реалізовано алгоритм аналізу поля user-agent. Однак, через складний формат даного поля, що вміщує різні характеристики пристрою, було використано Python3 пакет user-agent. За допомогою даного пакету, є можливість явно отримати атрибути браузера, характеристики операційної системи, характеристики пристрою. Варто зазначити, що також є можливість бінарної відповіді на запитання, що відносяться до перелічених вище характеристик поля user-agent. Реалізація дослідження поля user-agent зображена на рисунку 3.2.

```
def parse_ua(devices):
    for device in devices:
        for ua_item in devices[device]['ua']:
            system = parse(ua_item)
            if system.device.family not in devices[device]['device']:
                devices[device]['device'].append(system.device.family)
            if system.os.family + ' ' + system.os.version_string not in devices[device]['os']:
                devices[device]['os'].append(system.os.family + ' ' + system.os.version_string)
            if system.browser.family + ' ' + system.browser.version_string not in devices[device]['clients']:
                devices[device]['clients'].append(system.browser.family + ' ' + system.browser.version_string)
```

Рисунок 3.2 – Реалізація алгоритму дослідження поля user-agent.

Для реалізації алгоритмів машинного навчання було обрано Python бібліотеку sklearn. Дана бібліотека вміщує реалізацію популярних класичних алгоритмів, включаючи обраний Random Forest. Реалізація алгоритму класифікації за допомогою методу Random Forest відображено на рисунку 3.3.

```
def run_RF(self):
    data_train_x = pd.read_csv(self.train_file, sep=',', header=None, index_col=0)
    data_train_y = []
    y_file = open(self.train_file)
    for line in y_file:
        if line.split(',')[0] in self.classes_new:
            data_train_y.append(self.classes_new[line.split(',')[0]])
        if line.split(',')[0] in self.classes_old:
            data_train_y.append(self.classes_old[line.split(',')[0]])
    data_test_x = pd.read_csv(self.test_file, sep=',', header=None, index_col=0)
    neigh = RandomForestClassifier(n_estimators=1000)
    neigh.fit(data_train_x, data_train_y)
    print(neigh.predict_proba(data_test_x))
    data_test_x['cluster'] = neigh.predict(data_test_x)
    print(data_test_x['cluster'])
    feature_importances = pd.DataFrame(neigh.feature_importances_, index=data_train_x.columns,
                                       columns=['importance']).sort_values('importance', ascending=False)
    print(feature_importances)
```

Рисунок 3.3 – Реалізація бінарної класифікації

На початку алгоритму відбувається зчитування раніше збережених поведінкових профілей пристроїв Інтернету речей та подача отриманих даних на вхід класифікатора. В результаті отримуємо клас відповідного профілю та показники важливості кожної особливості вектору.

Побудова моделі нейронної мережі виконана за допомогою бібліотеки TensorFlow. На рисунку 3.4 відображено реалізацію обраної RNN-CNN моделі.

```

def model_get(self, size, b, c):

    inputs = Input(shape=(b,c))

    layer = LSTM(self.n_features, return_sequences=True)(inputs)
    layer = LSTM(self.n_features, return_sequences=True)(layer)

    layer = Reshape((1,b,c))(layer)

    layer = Conv2D(32, (3,3), data_format="channels_first", padding="same", activation="relu")(layer)
    layer = MaxPooling2D(data_format='channels_first')(layer)

    layer = Flatten(data_format="channels_first")(layer)
    layer = Dropout(0.8)(layer)
    predictions = Dense(size, activation='softmax')(layer)

    network = Model(inputs=inputs, outputs=predictions)
    network.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])

    return network

```

Рисунок 3.4 – Реалізація RNN-CNN моделі

Дана реалізація включає реалізацію двошарової LSTM рівня, Conv2D, Maxpooling2D та Dense (в якості Fully Connection Layer) рівнів, враховуючи вхідні розмірності векторів, необхідні перетворення векторів в процесі проходження по кожному рівні моделі та кількість класів, на які необхідно класифікувати.

4.3 Результати дослідження

Для виконання необхідних обчислень було обрано два наступних набори даних, що були отримані з відкритих джерел. Кожен з отриманих наборів даних вміщують в собі різні типи пристроїв. Структуру кожного набору даних відображено на рисунку 3.5.

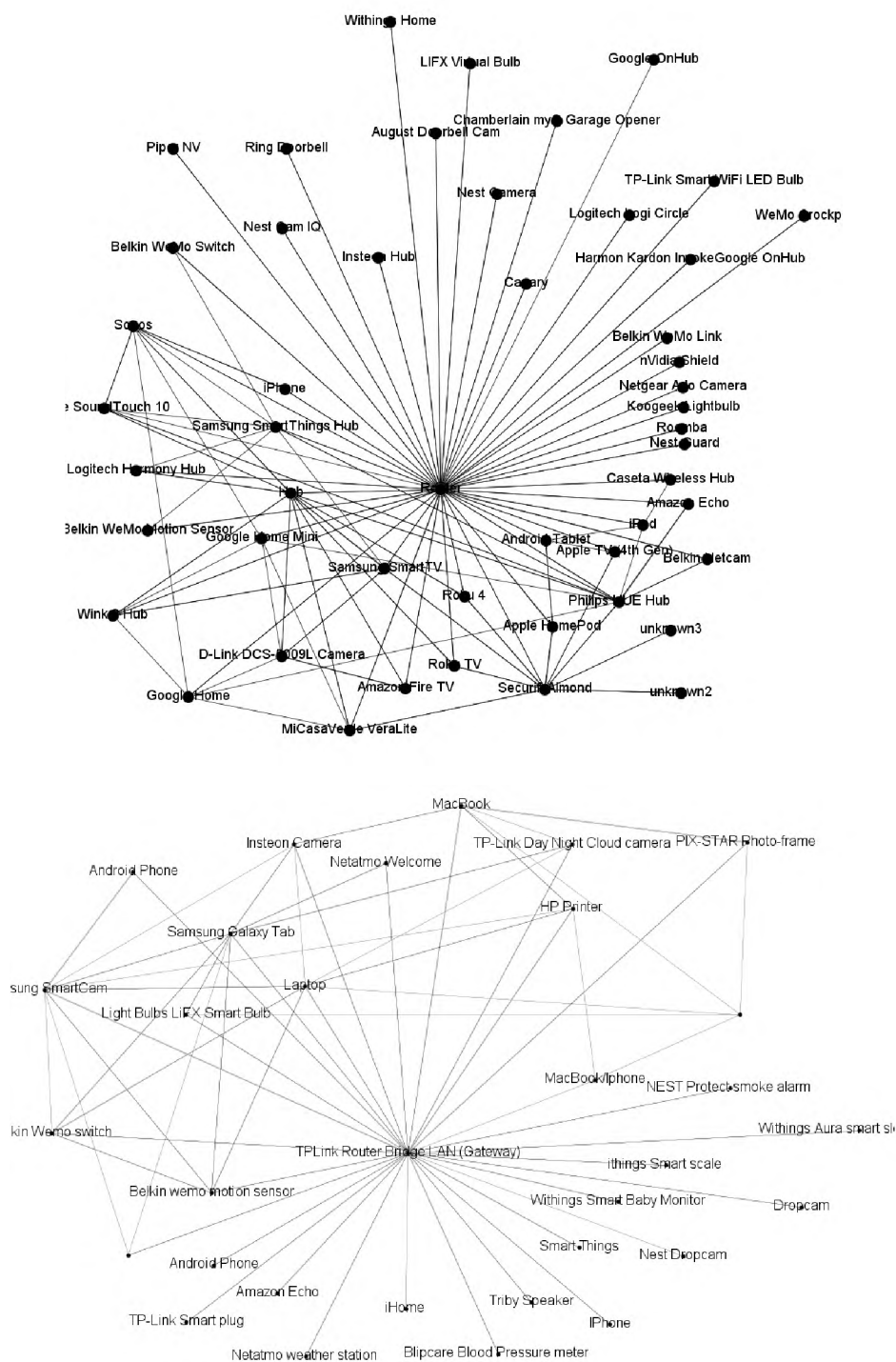


Рисунок 3.5 – Структури двох різних наборів даних.

Так, набір [37] включає в себе 54 пристрої, а набір [38] – 32 пристрої.

Для бінарної класифікації в якості побудови та навчання моделі було обрано набір даних, що вміщував найбільшу кількість пристроїв Інтернету речей [37]. В ході аналізу, було встановлено наступне. Кількість вузькоцільових пристроїв для даного набору даних становила 34, багатоцільових – 20. Трафік

вміщує 4 дні спостереження, загальний розмір складає ~100 ГБ. Для побудови поведінкових профілів кожного пристрою, було обрано період спостереження 4 дні.

Для тестування, було обрано [38] набір даних, в ході аналізу було встановлено, що кількість вузькоцільових пристроїв становить 22 та багатоцільових – 9. Трафік вміщує 20 днів спостереження. Загальний розмір складає 16 ГБ.

Поведінковий профіль в режимі бінарної класифікації включав в себе наступні особливості: кількість унікальних DNS запитів; кількість типів протоколів, що використовуються в процесі комунікації; тип пристрою за допомогою поля user-agent; кількість унікальних з'єднань в локальній мережі; сила зв'язності пристроїв в локальній мережі; кількість доменних імен, з котрими пристрій має TCP з'єднання; кількість невідомих з'єднань; середній час TCP з'єднання; середнє значення кількості трафіку за одну TCP сесію; кількість всіх з'єднань.

Результати бінарної класифікації наведено в таблиці 4.1. Параметр кількості дерев при класифікації встановлений 1000, параметр $m = 4$.

Таблиця 4.1 - Результат бінарної класифікації з межею в 21000 пакетів

Реально \ Прогноз	Запропоноване рішення		Попередні рішення	
	Вузькоцільові (Positive)	Багатоцільові (Negative)	Вузькоцільові (Positive)	Багатоцільові (Negative)
Вузькоцільові (Positive)	TP = 22	FP = 2	TP = 22	FP = 5
Багатоцільові (Negative)	FN = 0	TN = 7	FN = 0	TN = 4

Результат класифікації, демонструє, що алгоритм досягає стабільно високої точності (таблиця 4.2). Варто зазначити, під час спостереження було

виявлено оптимальну межу в 21000 перших зібраних пакетів для кожного пристрою. Хибними виявилися два рішення, що віднесли акустичні системи з підвищеною вбудованою функціональністю до вузькоцільових пристроїв, в силу пасивної (неактивної) поведінки в мережі.

Таблиця 4.2 – Показники точності класифікації

Показник	Запропоноване рішення	Попередні рішення
Recall	100%	100%
Precision	92%	81%
Accuracy	94%	84%
F1-score	96%	89%

Recall - це ймовірність того, що фактичний пристрій буде успішно класифікований як такий; Precision - це ймовірність того, що вузькоцільовий пристрій є справді вузькоцільовим; F1-оцінка - це єдиний індекс продуктивності.

Однак, при підвищенні межі кількості пакетів для класифікації до 560000, отримуємо всього одну помилку (таблиця 4.3), при цьому наступні показники складають: $Recall = 100\%$; $Precision = 96\%$; $Accuracy = 97\%$ $F1 - score = 98\%$.

Таблиця 4.3 - Результат бінарної класифікації з межею в 560 000 пакетів

Прогноз \ Реально	Запропоноване рішення	
	Вузькоцільові (Positive)	Багатоцільові (Negative)
Вузькоцільові (Positive)	TP = 22	FP = 1
Багатоцільові (Negative)	FN = 0	TN = 8

Також важливо помітити, що в обох результатах відсутні помилки, пов'язані з класифікуванням вузькоцільових пристроїв як багатоцільових. Це означає, що даний підхід здатен виявити за допомогою поведінкових профілів

потоків мережевих пакетів кожного пристрою вузькоцільові та багатоцільові пристрої в різних мережах.

При цьому, обчисливши ступінь важливості кожного показника поведінкового профілю, можна побачити наступні результати, що наведені в таблиці 4.4. Варто зазначити, сума всіх значень ступенів важливості рівна 1.

Таблиця 4.4 - Ступінь важливості показників поведінкового профілю

Показник	Ступінь важливості
Кількість унікальних DNS запитів	0.28
Кількість доменних імен, з котрими пристрій має TCP з'єднання	0.18
Кількість невідомих з'єднань	0.14
Кількість всіх з'єднань	0.099
Сила зв'язності пристроїв в локальній мережі	0.070
Кількість з'єднань в локальній мережі	0.065
Кількість типів протоколів, що використовуються в процесі комунікації	0.049
Тип пристрою за допомогою поля user-agent	0.044
Середній час TCP з'єднання	0.04
Середній трафік за одну TCP сесію	0.033

Отримані результати підтверджують важливість запропонованих припущень щодо ефективного розділення пристроїв Інтернету речей різного типу.

Порівнюючи різні алгоритми машинного навчання можна побачити наступне. Технологія випадкового лісу демонструє кращу точність класифікації в порівнянні з іншими. Наведені спостереження наведено в таблиці 4.5.

Таблиця 4.5 - Порівняння ефективності різних методів машинного навчання

Технологія машинного навчання	Точність
Random Forest	94%
Decision Tree	87%
kNN	78%

На рисунку 3.6 відображено робочу характеристику приймача для різних алгоритмів машинного навчання.

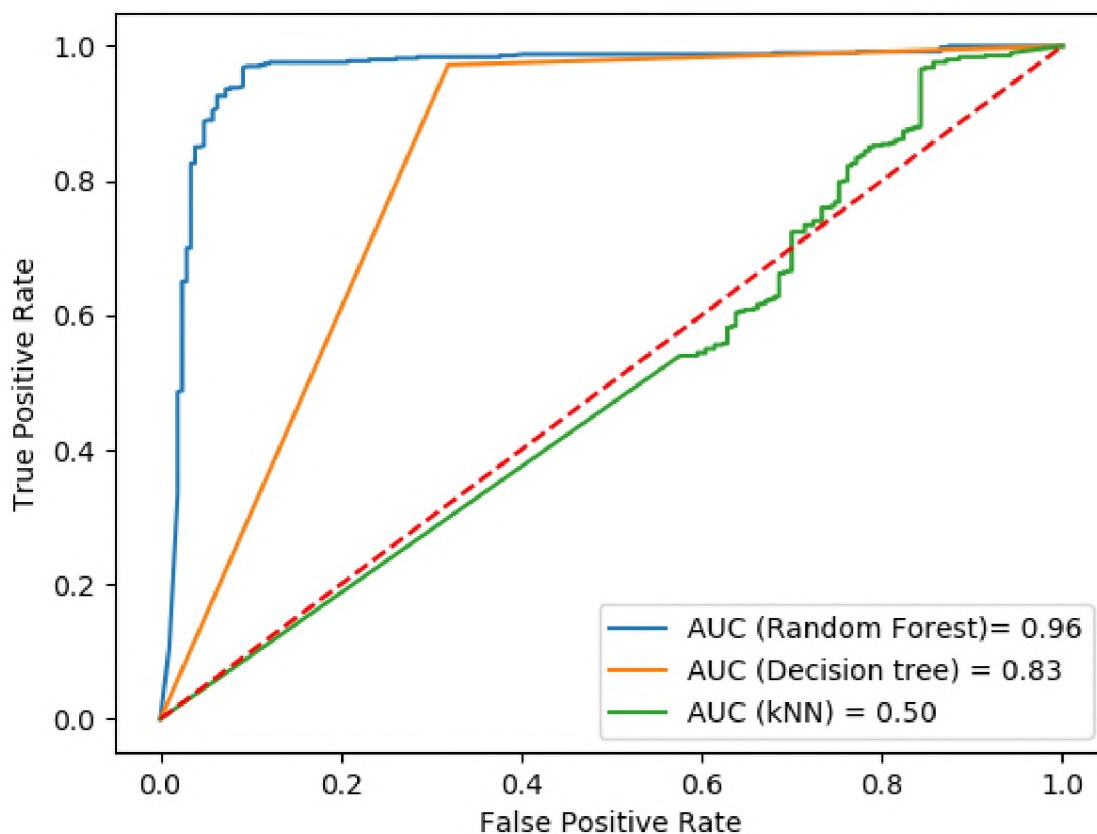


Рисунок 3.6 – Робоча характеристика приймача різних алгоритмів машинного навчання

У випадку мультигрупової класифікації, було розглянуто розподілення пристроїв Інтернету речей на запропоновані універсальні класи.

Так, в таблиці 3.6 наведені результати валідації для класифікації пристроїв Інтернету речей, використовуючи різні комбінації характеристик потоку пакетів по кожному сегментові різного часового діапазону. Експерименти проводилися за трьох різних режимів налаштувань моделі:

- Режим 1: розмір сегменту - 5 хв, кількість сегментів для навчання – 6;
- Режим 2: розмір сегменту - 5 хв, кількість сегментів для навчання – 15-20;
- Режим 3: розмір сегменту - 1 хв, кількість сегментів для навчання – 10.

Для кожного експерименту використовувалися обидва набори даних, кожен для навчання та валідації. Для набору даних [37] для навчання було обрано 3 дні та для валідації – наступні 3-4 години; для набору [38] – 7 днів навчання та 1 день валідації.

Таблиця 4.6 - Результати унікальної класифікації пристроїв Інтернету речей

[illegible]

Продовження таблиці 4.6

1	2	3				4					
Кількість службових пакетів	+	+	+	+	-	-	-	-	-	-	+
Середня довжина службових пакетів	+	+	+	+	+	+	+	+	+	+	+
Максимальна довжина службових пакетів	+	+	+	+	+	+	+	+	+	+	+
Кількість унікальних MAC	-	+	+	+	-	-	-	-	-	-	+
Кількість DNS пакетів	-	+	+	+	+	+	+	+	+	+	+
Кількість NTP пакетів	-	+	+	+	+	+	+	+	+	+	+
Кількість унікальних IP адрес	-	-	+	+	-	-	-	-	-	-	+
Кількість ICMP пакетів	-	-	+	+	+	+	+	+	+	+	+
Кількість ARP пакетів	-	-	+	+	+	+	+	+	+	+	+
Кількість HTTP пакетів	-	-	-	+	-	-	+	+	-	-	+
Кількість HTTPS пакетів	-	-	-	+	-	-	+	+	-	-	+
Кількість отриманих пакетів	-	-	-	-	-	+	+	+	-	+	+
Кількість надісланих пакетів	-	-	-	-	-	+	+	+	-	+	+
Кількість mDNS пакетів	-	-	-	-	-	-	-	+	+	+	+
Результат класифікації	81%	88%	90%	90%	86%	88%	90%	91%	88%	92%	90%

Згідно з отриманими результатами, найбільш високий показник точності валідації класифікації сягнув 92% точності, використовуючи при цьому один набір даних для навчання та валідації, в різних конфігураціях.

Для класифікації пристроїв на універсальні групи, [37] набір даних використовувався для навчання та [38] набір даних для тестування. Варто зазначити, що обидва набори даних включають різні пристрої, таким чином імітуючи класифікацію раніше не підключених до мережі пристроїв. Набір даних для навчання вибирався в силу більшості вміщуваних в собі пристроїв Інтернету речей.

В таблиці 4.7 наведено приклади результатів класифікації пристроїв на групи, що запропоновано в попередніх роботах, в таблиці 4.8 наведено результат класифікації пристроїв на універсальні групи (жирним кольором виділено пристрої, що не відповідають результуючому класу).

Таблиця 4.7 - Результати класифікації пристроїв Інтернету речей на класи, запропоновані в попередніх роботах

Клас пристроїв	Пристрої
Хаби	Dropcam , Smart Things, Netatmo Welcome , Withings Smart Baby Monitor , Netatmo weather station
Камери	PIX-STAR Photo-frame
Тригери та перемикачі	TP-Link Day Night Cloud Camera , Belkin Wemo Motion Sensor, TP-LINK Smart plug
Світлові лампи	
Електроніка	Tribby Speaker, Samsung Galaxy Tab , Amazon Echo, Samsung SmartCam , HP Printer, Laptop , Android Phone , Belkin wemo switch, Light Bulbs LiFX Smart Bulb ,
Гаджети	MacBook/Iphone

Таблиця 4.8 - Результати класифікації пристроїв Інтернету речей на запропоновані універсальні класи

Клас пристроїв	Пристрої
1	2
Контроллери	Smart Things, Amazon Echo

Продовження таблиці 4.8

1	2
Камери	Dropcam, Netatmo Welcome, Belkin Wemo Motion Sensor , TP-Link Day Night Cloud Camera, Insteon Camera, Nest Dropcam
Електронні командні пристрої	Belkin wemo switch, PIX-STAR Photo-frame, HP Printer, Withings Smart Baby Monitor , Tribby Speaker, iHome, Withings Smart scale, Light Bulbs LiFX Smart Bulb, Blipcare Blood Pressure meter
Сенсори	Netatmo weather station, TP-LINK Smart plug , Withings Aura smart sleep sensor, NEST Protect smoke alarm
Високотехнологічні пристрої	Samsung Galaxy Tab, Laptop, Android Phone, Samsung SmartCam , MacBook/Iphone,

Відповідно до отриманих результатів, точність класифікації пристроїв на універсальні класи сягає близько 84%, що в порівнянні з класифікацією пристроїв Інтернету речей на запропоновані класи в попередніх роботах на 10% вище. Однак, однією з причин не досить високої класифікації пристроїв Інтернету речей може бути часова різниця наборів даних, що складає 2 роки. Для більшості пристроїв Інтернету речей може бути значно оновлена працездатність.

Підсумовуючи, необхідно зазначити наступне. У випадку бінарної класифікації, отримуємо швидку систему. Однак, дана система є не працездатною у випадку мультигрупової класифікації. Для вирішення цієї проблеми доцільно використовувати технології нейронних мереж.

Однак, необхідно враховувати обчислювальну складність запропонованих алгоритмів. Так, в ході експериментів, аналіз потоку пакетів та навчання моделі нейронних мереж найбільшим набором даних склало 6 годин (24 години набору даних), та використання RAM при цьому склало 1500-2000 МБ. Виходом з цієї ситуації є використання наперед навченої моделі. В порівнянні з алгоритмами

машинного навчання, аналіз потоку пакетів пристроїв Інтернету речей тривав близько 1 години (7 днів спостереження) та використання RAM склало 200-300 МБ.

Висновки до розділу 3

Отже, з результатів дослідження, можна зробити висновок, що запропонований алгоритм є досить продуктивним. Як наслідок, різниця в точності бінарної класифікації пристроїв на вузькоцільові та багатоцільові між запропонованою моделлю та моделлю, запропонованій в попередніх дослідженнях склала 10%. Тестування на різних наборах даних показує, що модель більш стійка до різних мереж (точність сягнула 94%). У випадку мультигрупової класифікації точність валідації розділення пристроїв в межах одного набору даних зросла на 15%, у випадку класифікації пристроїв на запропоновані універсальні групи використовуючи різні набори даних для тренування та тестування, точність зросла (в порівнянні з існуючими рішеннями) на 10% і склала 84%. При цьому, необхідно брати до уваги факт часової відмінності обох наборів даних, що складає 2 роки. Для більшості пристроїв Інтернету речей може бути значно оновлена працездатність.

4 РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ

4.1 Опис ідеї проекту

В даній роботі описано основні складові стартап-проекту. Так, в таблиці 4.1 представлено загальний опис стартап-проекту та основні напрямки впровадження; в таблиці 4.2 було проаналізовано основні техніко-економічні переваги запропонованої ідеї в порівнянні з існуючими пропозиціями.

Таблиця 4.1 – Опис ідеї стартап-проекту

<i>Зміст ідеї</i>	<i>Напрямки застосування</i>	<i>Вигоди для користувача</i>
Класифікація пристроїв Інтернету речей використовуючи особливості потоку мережевих пакетів кожного з них	Житлові будинки, що включають пристрої Інтернету речей (в рамках Smart Home).	Полегшене адміністрування, шляхом виокремлення груп пристроїв та застосування необхідних дій по кожній групі, без необхідності контролювати кожен пристрій.
		Побудова мережевої карти підключень пристроїв Інтернету речей
	Офісні приміщення та державні установи України, що включають пристрої Інтернету речей (в рамках Smart City)	Застосування базової безпеки пристроїв шляхом ізоляції, наприклад, за допомогою створення окремих VLAN.
		Мінімізація фінансових втрат.

Таблиця 4.2 - Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п / п	Техніко- економічні характери- стики ідеї	(потенційні) товари/концепції конкурентів		W (слабка сторона)	N (нейт- ральна сторона)	S (сильна сторона)
		Мій проект	IoT Sentiel			
1	2	3	4	5	6	7
1.	Еконо- мічні	Витрати на розробку програмного забезпечення (парсерів)	Витрати на розробку програмного забезпечення з використан- ням систем віртуалізації	Слабкий рівень марке- тинго- вих рішень	Алгорит- мічна схожість запропо- нованих рішень	Просто- та та дешеви- зна запро- понова- ного підходу
2.	Технічні	Використан- ня різних алгоритмів машинного навчання	Використання класичного алгоритму машинного навчання	Немає	Класич- ний алгоритм машин- ного навчання Random Forest	Можли- вість застосу- вання штучної нейрон- ної мережі
3.	Техноло- гічні	Використан- ня потужностей маршрутиза- торів	Використання потужностей вискотехноло- гічних маршрутиза- торів	Зміни до проши- вки маршру- тизатора	Базовий функціо- нал маршру- тизаторів	Дешеви- зна

Продовження таблиці 4.2

1	2	3	4	5	6	7
4.	Надійності	Висока точність класифікації пристроїв (~96%)	Досить висока точність класифікації (~81%)	Можливість пониження пропускну здантності маршрутизатора	Метод забезпечення ізоляції	Використання заводських властивостей маршрутизаторів без додаткових інструментів
5.	Ергономічні	Консольний вивід інформації та заводський інтерфейс маршрутизаторів	Консольний вивід інформації та інтерфейс, що надає розробник SDN контролера	Часто недружній заводський інтерфейс маршрутизаторів	Немає	Універсальність

Продовження таблиці 4.2

1	2	3	4	5	6	7
6.	Транспор- табель- ності	Швидкість встановлення та налаштуван- ня системи	Швидкість та налаштування системи може зайняти тривалий проміжок часу	Немає	Налашту- вання моделі класифі- кації	Швид- кість надання послуг при почат- ковому налаш- туванні
7.	Екологі- чності	Відсутність побічного впливу	Відсутність побічного впливу	Немає	Відсут- ність витрат на додаткові екологіч- ні заходи	Немає

4.2 Технологічний аудит ідеї проекту

В таблиці 4.3 було проведено аналіз технологічної здійсненності проекту, що включає опис всіх необхідних технологій для створення та реалізації проекту, наявність всіх згаданих технологій у відкритому доступі для розробників проекту.

З аналізу всіх необхідних складових, було зроблено висновок щодо можливості реалізації з технічної точки зору запропонованої ідеї

Таблиця 4.3 - Технологічна здійсненність ідеї проекту

<i>№ n/n</i>	<i>Ідея проекту</i>	<i>Технології її реалізації</i>	<i>Наявність технологій</i>	<i>Доступність технологій</i>
1	Збір та обробка мережевих пакетів пристроїв Інтернету речей	Використання програмно реалізованих бібліотек (dpkt)	Потрібно встановити та використати необхідні інструменти розробника	Так, ця технологія є доступною
2	Класифікація пристроїв Інтернету речей алгоритмами машинного навчання	Побудова моделі та реалізація алгоритмів машинного навчання	Потрібно встановити необхідні інструменти розробника для цілей машинного навчання (sklearn, TensorFlow)	Так, ця технологія є доступною
3	Забезпечення базової безпеки пристроїв	Використання заводських можливостей маршрутизаторів для налаштування VLAN	Потрібно придбати необхідне обладнання	Так, ця технологія є доступною
4	Внесення змін до заводської прошивки маршрутизатора	Відкрите програмне забезпечення для мережевого обладнання (OpenWRT)	Потрібно встановити необхідні інструменти розробника (OpenWRT)	Так, ця технологія є доступною

Обрана технологія реалізації ідеї проекту: так як для реалізації ідеї проекту всі технології є наявними та доступними, тому обираються всі вище описані технології.

4.3. Аналіз ринкових можливостей запуску стартап-проекту

В таблиці 4.4 представлено аналіз ринку, що включає теперішній стан та темпи розвитку.

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап-проекту

<i>№ n/n</i>	<i>Показники стану ринку (інструментів класифікації)</i>	<i>Характеристика</i>
1	Кількість головних гравців, од	5
2	Загальний обсяг продаж, грн/ум.од	1.5 млн ум.од
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу	Потреба у кадрах із високим рівнем знань у сфері програмування мережевих пристроїв
5	Специфічні вимоги до стандартизації та сертифікації	Потребує сертифікації зміненого коду мережевого обладнання
6	Середня норма рентабельності в галузі, %	Не менше 150

За попередніми аналізом, ринок для ефективних інструментів класифікації пристроїв Інтернету речей є дуже привабливим, однак існують чіткі вимоги до швидкодії. У таблиці 4.5 представлено характеристику основних споживачів запропонованого продукту.

Таблиця 4.5 – Характеристика потенційних клієнтів стартап-проекту.

<i>№ n/n</i>	<i>Потреба, що формує ринок</i>	<i>Цільова аудиторія (цільові сегменти ринку)</i>	<i>Відмінності у поведінці різних потенційних цільових груп клієнтів</i>	<i>Вимоги споживачів до товару</i>
1	Виявлення та класифікація пристроїв Інтернету речей, що підключені до мережі та застосування до них базової безпеки шляхом ізоляції для запобігання їх компрометації та створення ботнетів.	<ul style="list-style-type: none"> - Власники домівок, що вміщують пристрої Інтернету речей (в рамках створення Smart Home) - Компанії та органи державної влади, що займаються створенням розумних середовищ (в рамках Smart city) - Розробники програмного забезпечення 	<p>Для кожної з описаних категорій існують конкретні варіанти поведінки:</p> <ul style="list-style-type: none"> • Клієнти надають перевагу простим та точним інструментам. • Невелика кількість коштів у клієнтів для розгортання системи захисту. • Вимоги регулятора (державні установи) до сертифікації. 	<ul style="list-style-type: none"> - точність класифікації - швидкодія мережевого обладнання - простота налаштування

Таблиці 4.6 та 4.7 відображають потенційні фактори загроз, що можуть виникнути в процесі реалізації ідеї та фактори можливостей, що матимуть прямий вплив на підвищення якості продукту.

Таблиця 4.6 – Фактори загроз

<i>№ n/n</i>	<i>Фактор</i>	<i>Зміст загрози</i>	<i>Можлива реакція компанії</i>
1	Ціноутворення конкурентної продукції.	Політика цін конкурентів на аналогічні рішення, відкритість досліджень ефективності	Пошук ефективніших особливостей у зв'язку з технічним прогресом пристроїв.
2	Зниження точності класифікації	Технічний розвиток пристроїв Інтернету речей, що змінює поведінку (зміна поведінки ключових особливостей)	Пошук ефективних підходів до розробки інструмента для підвищення якості продукту.
3	Зниження заводських характеристик пристроїв	Зменшення пропускної спроможності мережевих пристроїв	Пошук всіх можливих підходів до ефективного використання наявних ресурсів (технічного обладнання).
4	Недостатня кількість досліджених пристроїв Інтернету речей	Зниження якості класифікації через недостатність знань про нові пристрої	Поповнення знань, шляхом знаходження нових наборів даних

Таблиця 4.7 - Фактори можливостей

<i>№ n/n</i>	<i>Фактор</i>	<i>Зміст можливості</i>	<i>Можлива реакція компанії</i>
1	Поява високотехнологічного мережевого обладнання	Збільшення точності класифікації за рахунок використання більш ресурсоємних та продуктивніших алгоритмів машинного навчання	Швидке впровадження нових можливостей алгоритмів машинного навчання для класифікації на цих пристроях
2	Поява нових оптимізованих алгоритмів машинного навчання	Оптимізація існуючих та поява нових алгоритмів машинного навчання	Застосування нових алгоритмів на практиці та покращення якості продукції
3	Поява нових відкритих наборів даних.	Збільшення знань про поведінку пристроїв Інтернету речей.	Побудова удосконалених моделей алгоритмів машинного навчання.
4	Залучення відомих компаній мережевого обладнання до співпраці.	Об'єднання спільних зусиль на створення продукту.	Поява високотехнологічних мережевих пристроїв з вбудованою функціональністю класифікації. Створення унікальних продуктів.

В таблиці 4.8 описано результати аналізу особливостей запропонованої політики конкуренції на ринку.

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку.

<i>Особливості конкурентного середовища</i>	<i>В чому проявляється дана характеристика</i>	<i>Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)</i>
1	2	3
Чиста конкуренція	Динаміка цін, залежить від ефективності та швидкодії побудованого інструменту; конкуренція зростає через відкритість дослідження даного напрямку.	Вдосконалення підходу до класифікації, виокремлюючи додаткові характерні особливості пристроїв в силу постійного оновлення технологічності.
За рівнем конкурентної боротьби - міжнародна	Надання послуг в межах України та закордоном.	Проведення активної реклами для привертання уваги провідних виробників мережових обладнань та пристроїв Інтернету речей.
За галузевою ознакою - міжгалузева	Рішення може бути однаково застосовано в різних сферах діяльності.	Розширення сфери застосування інтелектуального продукту, шляхом вивчення потреб кожної сфери діяльності в пристроях Інтернету речей.
Конкуренція за видами товарів: - товарно-видова	Продукт відрізняється від інших рішень точністю та якістю прогнозування та надійністю забезпечення безпеки.	Надання послуг додаткової точності, шляхом удосконалення моделі класифікації пристроями Інтернету речей конкретної сфери діяльності.

Продовження таблиці 4.8

1	2	3
За характером конкурентних переваг - цінова	Ціна продукту корелюється за рахунок відкритості наукових досліджень	Надання інструменту класифікації орієнтовну на досягнення безпеки пристроїв в мережі за рахунок доступних технологій та меншої вартості.
За інтенсивністю - марочна	Використання унікального ідентифікатору продукту.	Наголос на постійному вдосконаленні точності класифікації певного продукту Інтернету речей.

В таблиці 4.9 наведено основні пункти, що характеризують умови конкуренції продукту на ринку.

Таблиця 4.9 - Аналіз конкуренції в галузі за М. Портером

1	2	3	4	5	6
Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	IoT Sentiel	Багато дослідницьких продуктів: IoT Keeper, DIoT, IoT or NoT	Постійна підтримка продукту та його удосконалення.	Власники розумних середовищ (в рамках Smart Home / Smart City тощо)	Товари-замінники не присутні на ринку

Продовження таблиці 4.9

1	2	3	4	5	6
Висновки:	Продукт є досить малокоштовним через використання конкурентом технологій віртуалізації.	Існує великий потенціал виходу на ринок, так як існує багато дослідницьких проектів, що здатні дуже швидко комерціалізуватися	Постачальник має привілеї на ринку через кількість проданих товарів та позитивним зворотнім зв'язком.	Клієнти диктують умови на ринку; при закупівлі приділяється увага на цінову пропозицію та швидкодію.	Обмеження на ринку відсутні.

Відповідно до аналізу конкурентної ситуації на ринку запропонованих інструментів класифікації, даний проект має досить потужні можливості. Сильними сторонами даного інструменту є висока точність класифікації пристроїв, навіть незважаючи на рік їх виготовлення, використання точної класифікації за допомогою технологій нейронних мереж та відносно невелике споживання обчислювальних ресурсів при класифікації, що позначається на швидкодії; широкий попит до пристроїв Інтернету речей сприяє швидкому розповсюдженню даного продукту через швидкий темп росту атак на розумні середовища.

У таблиці 4.10 відображено основні опорні точки конкурентоспроможності запропонованого продукту на ринку.

Таблиця 4.10 - Обґрунтування факторів конкурентоспроможності

<i>№ п/п</i>	<i>Фактор конкурентоспроможності</i>	<i>Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)</i>
1	Ціна	Ціна продукту залежить лише від якості програмного коду, тому вона не є великою. Також, код буде застосовуватися на доступному на сьогоднішній день мережевому обладнанні.
2	Якість	Продукт демонструє досить точні результати, використовуючи при цьому не значну кількість обчислювальних ресурсів, що також демонструє швидкодію.
3	Гнучкість рішення	Кожен споживач має змогу підлаштувати модель для своєї сфери діяльності, при цьому він матиме можливість донавчати модель.
4	Універсальність рішення	На відміну від конкурентів, є можливість застосування продукту безпосередньо на існуючий у покупця мережевий пристрій, зменшуючи затрати на нове обладнання.

Таким чином, виходячи з результатів аналізу, можна побачити запропоновані вище особливості продукту будуть дуже близькими до потенційних споживачів, що робить даний проект конкурентоспроможним серед усіх наявних на сьогоднішній день рішень.

У таблиці 4.11 наведено також порівняльний аналіз сильних сторін продукту з рішеннями конкурентів. У таблиці 4.12 описано запропоноване рішення з точки зору SWOT.

Таблиця 4.11 - Порівняльний аналіз сильних та слабких сторін запропонованого рішення.

№ n/ n	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з запропонованим рішенням						
			-3	-2	-1	0	+1	+2	+3
1	Ціна	15		+					
2	Якість	17			+				
3	Гнучкість рішення	7					+		
4	Універсальність рішення	11		+					

Таблиця 4.12 - SWOT-аналіз стартап-проекту

<p><i>Сильні сторони:</i> висока точність класифікації пристроїв навіть незважаючи на рік їх виготовлення, використання точної класифікації за допомогою технологій нейронних мереж та відносно невелике споживання обчислювальних ресурсів при класифікації, що позначається на швидкодії</p>	<p><i>Слабкі сторони:</i> слабкі маркетингові кроки; можливі зміни в роботі мережевих пристроїв; формування вручну політик безпеки.</p>
<p><i>Можливості:</i> використання впровадження нових технологій машинного навчання; поповнення бази знань новими пристроями Інтернету речей; залучення провідних виробників мережевого обладнання та ІТ компаній в якості співвиробників.</p>	<p><i>Загрози:</i> цінова конкуренція; наростання відкритих дослідницьких робіт.</p>

При складанні стартап-проекту також було проаналізовано основні шляхи відходу у випадку виникнення кризових ситуацій. Кризовими ситуаціями можуть бути надто велика конкуренція або наднизький попит у продукті.

Таблиця 4.13 - Альтернативи ринкового впровадження стартап-проекту

<i>№ п/п</i>	<i>Альтернатива (орієнтовний комплекс заходів) ринкової поведінки</i>	<i>Ймовірність отримання ресурсів</i>	<i>Строки реалізації</i>
1	Концентрація на певній сфері діяльності покупця	Висока ймовірність отримання ресурсів	3 місяці
2	Концентрація виробництва на конкретній моделі мережевого пристрою	Середня ймовірність отримання ресурсів	1 рік
3	Заклучення контракту з провідним виробником мережевого обладнання	Мала ймовірність отримання ресурсів	3 роки
4	Участь в провідних конференціях та семінарах	Середня ймовірність отримання ресурсів	1 рік

З визначених вище альтернатив ринкового впровадження даного стартап-проекту найбільш доцільним є рішення бути учасником провідних конференцій та семінарів з питань захисту пристроїв Інтернету речей. Період реалізації становитиме приблизно 1 рік.

4.4. Розроблення ринкової стратегії проекту

У таблиці 4.14 відображено аналіз потенційних груп споживачів запропонованого програмного рішення.

Таблиця 4.14 - Вибір цільових груп потенційних споживачів.

<i>№ п / п</i>	<i>Опис профілю цільової групи потенційних клієнтів</i>	<i>Готовність споживачів сприйняти продукт</i>	<i>Орієнтовний попит в межах цільової групи (сегменту)</i>	<i>Інтенсивність конкуренції в сегменті</i>	<i>Простота входу у сегмент</i>
1	Житлові приміщення (в рамках Smart Home)	Клієнти зацікавлені в продукті, однак не готові витратити великі кошти	Середній рівень попиту, пов'язаний з коштовністю.	Високий рівень конкуренції, через недовільно великий рівень знань про пристрої	Є складність входу, яка пов'язана із великим відкритим дослідницьким напрацюванням
2	Офісні приміщення та державні установи (в рамках Smart City)	Клієнти готові придбати рішення, оскільки керують цінним активом.	Високий попит, що пов'язаний з кібербезпекою	Низький рівень конкуренції	Є складність, яка пов'язана із складністю поставленої задачі споживача.
Які цільові групи обрано: обидві цільові групи, а саме: житлові приміщення (в рамках Smart Home), офісні приміщення та державні установи (в рамках Smart City).					

Відповідно до аналізу та вибору основних цільових груп було обрано стратегію масового маркетингу. В таблицях 4.15, 4.16 та 4.17 визначено стратегії

розвитку, стратегії позиціонування та стратегії конкурентної поведінки запропонованого продукту відповідно.

Таблиця 4.15 - Визначення базової стратегії розвитку

<i>№ п/ п</i>	<i>Обрана альтернатива розвитку проекту</i>	<i>Стратегія охоплення ринку</i>	<i>Ключові конкурентоспроможні позиції відповідно до обраної альтернативи</i>	<i>Базова стратегія розвитку</i>
1	Фронтальний наступ	Стратегія масового маркетингу	Сильні сторони продукту	Стратегія диференціації

Таблиця 4.16 - Визначення стратегії позиціонування

<i>№ п/ п</i>	<i>Вимоги до товару цільової аудиторії</i>	<i>Базова стратегія розвитку</i>	<i>Ключові конкурентоспро- можні позиції власного стартап- проекту</i>	<i>Вибір асоціацій, які мають сформувані комплексну позицію власного проекту</i>
1	2	3	4	5
1	Купівля готового продукту для застосування, що здатне виконувати надійно і точно свої функціональні задачі, а саме: класифікацію пристроїв Інтернету	Стратегія диферен- ціації	Ключові (сильні) характеристики продукту: вартість, точність, універсальність.	Зручний інструмент керування мережевими потокми. Точність класифікації пристроїв;

Продовження таблиці 4.16

1	2	3	4	5
	речей. Отримані результати дуже важливі для забезпечування базової безпеки шляхом створення на маршрутизаторах окремих VLAN. Для державних установ необхідне підтвердження надійності наданого функціоналу шляхом сертифікації.			швидкість виконання операцій; універсальність продукту; низька кошовність.

Таблиця 4.17 - Визначення базової стратегії конкурентної поведінки

<i>№ п/п</i>	<i>Чи є проект «першопрохідцем» на ринку?</i>	<i>Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?</i>	<i>Чи буде компанія копіювати основні характеристики товару конкурента, і які?</i>	<i>Стратегія конкурентної поведінки</i>
1	Ні	Так	Так, напрацьовані особливості пристроїв.	Стратегія виклику лідера

4.5. Розроблення маркетингової програми стартап-проекту

Для представлення маркетингової програми було проаналізовано та виокремлено основні переваги запропонованого рішення (таблиця 4.18). Також у таблиці 4.19 представлено детальний тривірневий опис моделі товару.

Таблиця 4.18 - Визначення ключових переваг концепції потенційного товару

<i>№ п/ п</i>	<i>Потреба</i>	<i>Вигода, яку пропонує товар</i>	<i>Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити</i>
1	Визначення пристроїв в мережі	Побудова мережевої карти пристроїв.	Визначення взаємозалежності пристроїв та надмірний вплив одне на одного.
2	Розподілення пристроїв на окремі класи	Комплексне застосування заходів безпеки до кожної відокремленої групи.	Виокремлення універсальних класів пристроїв не за сферою застосування, а за базовою функціональністю.
3	Точність класифікації	Високий показник точності результату	Використання алгоритмів машинного навчання в поєднанні з нейронною мережею.
4	Низька вартість та універсальність	Легкість керування, використання наявного обладнання	Використання необхідної прошивки до відповідного обладнання.

Таблиця 4.19 - Опис трьох рівнів моделі товару

<i>Рівні товару</i>	<i>Сутність та складові</i>		
I. Товар за задумом	Класифікація пристроїв Інтернету речей за допомогою алгоритмів машинного навчання для комплексного управління активами, а саме: забезпечення базової безпеки пристроїв шляхом ізоляції за допомогою створення окремих VLAN.		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. Машинне навчання	М	Тх/Тл
	2. Виокремлення універсальних груп пристроїв	М	Тх
	3. Універсальність застосування	М	Вр/Тл/Е
	4. Швидкість роботи	М	Тх
	5. Можливість отримання мережевої карти.	Нм	Тл/Е/Тх
	6. Можливість в деяких випадках встановити назву та ПЗ пристрою	Нм	Тх/Е
	Якість: Стабільність точності класифікації; сертифікація програмного продукту.		
	Електронний ключ доступу до продукту.		
	Марка: IoT Finder		
III. Товар із підкріпленням	До продажу: надання демо-версії програмного забезпечення для початкового ознайомлення.		
	Після продажу: проведення необхідних оновлень.		
За рахунок чого потенційний товар буде захищено від копіювання: використання міжнародного знаку авторського права			

У таблиці 4.20 відображено визначено цінову політику запропонованого рішення, у таблиці 4.21 визначено основні форми збуту товару та у таблиці 4.22 описано основні кроки до залучення майбутніх споживачів.

Таблиця 4.20 – Визначення меж встановлення ціни

<i>№ п/п</i>	<i>Рівень цін на товари- замінники</i>	<i>Рівень цін на товари- аналоги</i>	<i>Рівень доходів цільової групи споживачів</i>	<i>Верхня та нижня межі встановлення ціни на товар/послугу</i>
	12000 ум.од.	15000 ум.од.	10000-12000 ум.од.	8000-8500 ум.од.

Таблиця 4.21 - Формування системи збуту

<i>№ п/п</i>	<i>Специфіка закупівельної поведінки цільових клієнтів</i>	<i>Функції збуту, які має виконувати постачальник товару</i>	<i>Глибина каналу збуту</i>	<i>Оптимальна система збуту</i>
1	Прийняття рішення про необхідність в системі класифікації пристроїв Інтернету речей та пошук пропозицій з мінімальними затратами, що включає мінімізацію кількості посередників.	Виконання запитів потенційних користувачів; пошук популярних каналів збуту з мінімальними вкладеннями; чіткі маркетингові кроки; мінімізація посередництва.	Однорівневий канал збуту (із залученням посередника)	Горизонтальна система збуту - пошук компаній для формування системи збуту

Таблиця 4.22 - Концепція маркетингових комунікацій

<i>№ п/п</i>	<i>Специфіка поведінки цільових клієнтів</i>	<i>Канали комунікації, якими користую ться цільові клієнти</i>	<i>Ключові позиції, обрані для позиціонування</i>	<i>Завдання рекламного повідомлення</i>	<i>Концепція рекламного звернення</i>
1	Дослідження структури продукту, ключових компонентів; пошук можливостей тестування, що в подальшому вплине на рішення про необхідність придбання	Інтернет реклама / реклама у соціальних мережах	Використання ключових характеристик: використання машинного навчання з поєднанням нейронних мереж; швидкість роботи; універсальність використання; немає необхідності у використанні	Показати переваги даного продукту для забезпечення потреб споживача в захисті активів.	Реклама, що націлена на функціональні можливості програмного забезпечення та важливість результату на забезпечення потреб в кібербезпеці

Висновки до розділу 4

Отже, в результаті аналізу потенційної економічної цінності представленого інструменту класифікації пристроїв Інтернету речей, було встановлено, що дане рішення має потенційну можливість комерціалізації. У зв'язку з ростом зацікавленості у подібних системах через велику ймовірність виникнення кіберзагроз, на сьогоднішній день можна спостерігати за високим рівнем попиту на подібні продукти. Таким чином, рентабельність продукту є виправданим на ринку кіберрішень, що в свою чергу створює всі необхідні умови для входу. Однак, існують бар'єри, що пов'язані з вузькими рамками обчислювальних можливостей сучасних мережевих активів, ефективно долаючи які значно підвищується конкурентноспроможність запропонованого рішення.

Альтернативним рішенням запровадження стартап-проекту є пошук компаній-виробників мережевого обладнання, для об'єднання зусиль на створення унікального рішення «з коробки». Я вважаю в доцільності продовження роботи над даним проектом.

ВИСНОВКИ

В результаті роботи було зроблено огляд існуючих підходів до класифікації пристроїв Інтернету речей в мережі. Було встановлено, що не всі запропоновані алгоритми здатні ефективно класифікувати пристрої на різні групи та бути застосованими в реальних умовах.

Алгоритми на основі аналізу особливостей спектру, що отримується шляхом перетворення Фур'є дискретного сигналу передачі службових протоколів виявився непрацездатним в силу різних можливостей налаштувань пристроїв Інтернету речей. Таким чином, часові особливості для класифікації пристроїв є надто складними для аналізу, споживаючи при цьому чималі обчислювальні ресурси. Іншим підходом є аналіз статистичних характеристик потоку пакетів, будуючи відповідно поведінкові профілі кожного пристрою. Тому, саме цей підхід був обраний як основний в даній роботі.

Відповідно до цього, було проаналізовано основні характеристики потоку пакетів пристроїв Інтернету речей та було висунуто основні припущення про поведінкові особливості, які здатні з великою точністю виокремити пристрої того чи іншого класу, включаючи силу зв'язності пристроїв, характеристики пов'язані з DNS пакетами, поле user-agent тощо.

В якості алгоритмів машинного навчання були проаналізовані класичні методи та підходи до побудови ефективних штучних нейронних мереж. Таким чином, в роботі було запропоновано використання керованого алгоритму машинного навчання – Random Forest. Також, було представлено результати порівняння ефективності інших алгоритмів керованого навчання. В якості мультигрупової класифікації, було представлено багат шарову модель нейронної мережі типу RNN-CNN.

Для виконання експериментів, було обрано два відкриті набори даних, кожен з яких складеться з різних пристроїв Інтернету речей. Результати класифікації є наступними. Точність бінарної класифікації пристроїв

збільшилася на 10% в порівнянні з запропонованими раніше алгоритмами та сягнула 94%. При цьому, різні набори даних використовувалися в якості навчальної та тестової вибірки, імітуючи при цьому класифікацію пристроїв, що раніше не спостерігалися в мережі. Таким чином, тестування на різних наборах даних показує, що побудована модель є більш стійкою до різних мереж. На етапі валідації моделі для класифікації пристроїв на універсальні групи в межах одного набору даних, маємо приріст в показнику точності, а саме: точність класифікації зросла на 11%, а на етапі тестування моделі для класифікації на запропоновані універсальні групи пристроїв, точність сягнула 84%, що на 10% вище раніше запропонованих підходів, використовуючи при цьому різні набори даних одночасно. Можливою причиною не досить високої точності є часова різниця наборів даних, що сягає 2 роки. За отриманий проміжок часу могла значно бути оновлена функціональність пристроїв.

Отже, підсумовуючи, необхідно відзначити, що даний алгоритм може бути застосований в реальних умовах для ефективної класифікації пристроїв Інтернету речей. Отриманий результат може бути використаний в подальшому розмежуванні пристроїв для більш гнучкого формування політик безпеки, створення віртуальних мереж (VLAN) та досягнення вищої якості обслуговування і унеможливленні взаємних перешкоджень в роботі.

Ця робота надає поштовх майбутнім дослідженням у сфері безпеки локальних мереж від несанкціонованого впливу на пристрої, а також в сфері забезпечення необхідної продуктивності та якості обслуговування в середовищі Інтернету речей.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

- 1 Luigi Atzori. The Internet of Things: A survey [Текст] / Luigi Atzori, Antonio Iera, Giacomo Morabito // Computer Networks. -2010. -Volume 54. -С. 2787-2805.
- 2 ITU-T. Overview of the Internet of things [Текст] / ITU-T // SERIES Y: Global Information Infrastructure, Internet Protocol Aspects and Next-Generation Networks. -2012. -С. 1-22.
- 3 European Commission, Information Society and Media. Internet of Things in 2020 Roadmap for the Future [Текст] : монгр. / European Commission, Information Society and Media. -2008. -С. 1-27.
- 4 Інтернет речі [Електронний ресурс] . - Режим доступу: https://uk.wikipedia.org/wiki/Інтернет_речей
- 5 Pedro Ruben. Automatic Integration of IoT Devices [Текст] : моногр. / Pedro Ruben. - Lisboa: ISCTE-IUL. -2016. -С. 1-104.
- 6 Kevin Ashton. That “Internet of Things” Thing [Текст] / Kevin Ashton // RFID Journal. -1999. -С. 1
- 7 Dennis Knake. IoT numbers vary drastically: devices and spending in 2020 [Електронний ресурс]. — 2016. — Режим доступу: <https://www.wespeakiot.com/iot-numbers-devices-spending-2020>.
- 8 Dohr A. The Internet of Things for ambient assisted living [Текст] / Dohr A., Modre-Opsrian R., Drobics M., Hayn D., Schreier G. // Proceedings of the Seventh International Conference on Information Technology: New Generations (ITNG). - 2010. -С. 1-7.
- 9 Jara A. J. A Pharmaceutical Intelligent Information System to Detect Allergies and Adverse Drugs Reactions based on Internet of Things [Текст] / Jara A. J., Belchi F. J., Alcolea A. F., Santa J., Zamora-Izquierdo M. A., Gomez-Skarmeta A. F. // Proceedings of the 8th IEEE International Conference on

- Pervasive Computing and Communications Workshops (PERCOM Workshops). – 2010. -C. 1-5.
- 10 AKM Jahangir A. Majumder. A wireless IoT system towards gait detection in stroke patients [Текст] / AKM Jahangir A. Majumder, Yosuf ElSaadany, Mohammed ElSaadany, Donald R. Ucci, Farzana Rahman // 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). – 2017. -C. 1-6.
 - 11 Liu J. Applications of Internet of Things on smart grid in China [Текст] / Liu J., Li X., Chen X., Zhen Y., Zeng L. // Proceedings of the 13th International Conference on Advanced Communication Technology (ICACT). – 2011. -C. 1-5.
 - 12 Darianian M. Smart Home Mobile RFID based internet-of-things systems and services [Текст] / Darianian, M., Michael, M. P. // Proceedings of the Inf Syst Front International Conference on Advanced Computer Theory and Engineering (ICACTE'08). – 2008. -C. 116-120
 - 13 Schaffers H. Smart cities and the future internet: Towards cooperation frameworks for open innovation [Текст] / Schaffers H., Komninos N., Pallot M., Trousse B., Nilsson M., Oliveira A. // The future internet, lecture notes in computer science. – 2011. - Vol. 6656. Berlin: Springer. -C. 431–446.
 - 14 Vicini S. A living lab for Internet of Things vending machines [Текст] / Vicini S., Sanna A., Bellini S. // The impact of virtual, remote, and real logistics labs, communications in computer and information science. – 2012. - Vol. 282. Berlin: Springer. -C. 35–43.
 - 15 Vazquez J. I. Social devices: Autonomous artifacts that communicate on the internet [Текст] / Vazquez J. I., Lopez-de-Ipina D. // The Internet of Things, lecture notes in computer science. – 2008. - Vol. 4952. Berlin: Springer -C. 308–324.
 - 16 Guo B. Living with Internet of Things: The Emergence of Embedded Intelligence [Текст] / Guo B., Zhang D., Wang Z. // Proceedings of the IEEE

- International Conferences on Internet of Things, and Cyber, Physical and Social Computing (iThings/CPSCoM). -2011 -C. 1-9.
- 17 Guo B. Opportunistic IoT: Exploring the social side of the Internet of Things [Текст] / Guo B., Yu Z., Zhou X., Zhang D. // Proceedings of the IEEE 16th International Conference on Computer Supported Cooperative Work in Design (CSCWD). – 2012. -C. 1-6
 - 18 Weisong Shi. Edge Computing: Vision and Challenges [Текст] / Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, Lanyu Xu // IEEE Internet of Things Journal. – 2016. -C. 637 – 646.
 - 19 Andrew Whitmore. The Internet of Things—A survey of topics and trends [Текст] / Andrew Whitmore, Anurag Agarwal, Li Da Xu // Information Systems Frontiers. – 2014. -C. 1-14.
 - 20 T. Yu. Handling a trillion (unfixable) flaws on a billion devices: Rethinking network security for the internet-of-things [Текст] / T. Yu, V. Sekar, S. Seshan, Y. Agarwal, C. Xu // Proceedings of the 14th ACM Workshop on Hot Topics in Networks. - 2015. -C. 1–7.
 - 21 Hafeez. IoT-KEEPER: Securing IoT Communications in Edge Networks [Текст] / Hafeez, Antikainen, Yi Ding, Tarkoma. // Computer Science. – 2018. - C.1-20
 - 22 Noshina Tariq. The Security of Big Data in Fog-Enabled IoT Applications Including Blockchain: A Survey [Текст] / Noshina Tariq, Muhammad Asim, Feras Al-Obeidat, Muhammad Zubair Farooqi, Thar Baker, Mohammad Hammoudeh, Ibrahim Ghafir. // Sensors (Basel). - 2019. -C. 1-34.
 - 23 G. Mezzofiore. A university was attacked by its lightbulbs, vending machines and lamp posts [Электронный ресурс]. — 2017. — Режим доступа: <https://mashable.com/2017/02/13/internet-of-things-university-network-/#9RqajU3A5Oqu>.
 - 24 Mirai [Электронный ресурс]. - Режим доступа: [https://en.wikipedia.org/wiki/Mirai_\(malware\)](https://en.wikipedia.org/wiki/Mirai_(malware))

- 25 H. Guo. IP-Based IoT Device Detection [Текст] / H. Guo, J. Heidemann.// IoT S&P@SIGCOMM 2018 — 2018. -С. 1-7.
- 26 Martin J. Decomposition of MAC address structure for granular device inference [Текст] / Martin J., Ry, E., Beverly R. // Proceedings of the 32nd Annual Conference on Computer Security Applications. - ACSAC '16. – 2016. -С. 78–88.
- 27 Meidan Y. ProfillIoT: A Machine Learning Approach for IoT Device Identification Based on Network Traffic Analysis [Текст] / Meidan Y., Bohadana M., Shabtai A., Guarnizo J. D., Ochoa M., Ole Tippenhauer N., Elovici Y. // SAC 2017: The 32nd ACM Symposium On Applied Computing. – 2017. -С. 1-4.
- 28 Lastovicka M. Passive os fingerprinting methods in the jungle of wireless networks [Текст] / Lastovicka M., Jirsik T., Celeda P., Spacek S., Filakovsky D. // In IEEE/IFIP Network Operations and Management Symposium: Cognitive Management in a Cyber World, NOMS 2018. – 2018. -С. 1-9.
- 29 Thien Duc Nguyen. DIOT: A Federated Self-learning Anomaly Detection System for IoT [Текст] / Thien Duc Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, Ahmad-Reza Sadeghi // Proceedings of the 39th IEEE International Conference on Distributed Computing Systems (ICDCS) — 2019. — С. 1-12.
- 30 A. Sivanathan. Characterizing and Classifying IoT Traffic in Smart Cities and Campuses [Текст] / A. Sivanathan, D. Sherratt, H. Gharakheili, A. Radford, C. W. Vishwanath, V. Sivaraman. // 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). — 2018. -С. 1-6.
- 31 Lei Bai. Automatic Device Classification from Network Traffic Streams of Internet of Things [Текст] : моногр. / Lei Bai, Lina Yao, Salil S. Kanhere, Xianzhi Wang, Zheng Yang. – 2018. -С. 1-9.
- 32 What is MUD? [Электронный ресурс]. - Режим доступа: <https://developer.cisco.com/docs/mud/#!what-is-mud>

- 33 Ayyoob Hamza. Clear as MUD: Generating, Validating and Applying IoT Behavioral Profiles (Technical Report) [Текст] / Ayyoob Hamza, Dinesha Ranathunga, H. Habibi Gharakheili, Matthew Roughan, Vijay Sivaraman // IoT S&P@SIGCOMM 2018 - 2018. -С. 1-10.
- 34 S. Brin. The anatomy of a large-scale hypertextual web search engine. [Текст] / S. Brin, L. Page. // Computer Networks and ISDN Systems. - 1998. -С. 107-117.
- 35 Syed Junaid Nawaz. Quantum Machine Learning for 6G Communication Networks: State-of-the-Art and Vision for the Future [Текст] / Syed Junaid Nawaz, Shree K. Sharma, Shurjeel Wyne, Mohammad N. Patwary, Md Asaduzzaman // IEEE Access. – 2019. -С. 46317 – 46350.
- 36 T. K. Ho. Random decision forests [Текст] / T. K. Ho // ICDAR '95 Proceedings of the Third International Conference on Document Analysis and Recognition. - 1995. -С. 278.
- 37 Data [Электронный ресурс]. - 2018. - Режим доступа: <https://yourthings.info/data>.
- 38 Data collected for IEEE TMC 2018 [Электронный ресурс]. – 2016. -Режим доступа: <https://iotanalytics.unsw.edu.au/iottraces>