

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

«На правах рукопису»

УДК 004.056.5

«До захисту допущено»

Завідувач кафедри

\_\_\_\_\_ Дмитро ЛАНДЕ

“    ” \_\_\_\_\_ 2023 р.

**Магістерська дисертація**

на здобуття ступеня магістра

за освітньо-науковою програмою «Системи, технології та математичні  
методи кібербезпеки»

зі спеціальності 125 «Кібербезпека»

на тему: **Виявлення ознак атак на критичну інфраструктуру методом  
аналізу мережевого протоколу**

Виконав здобувач ступеня магістра II курсу, групи ФБ-11мн

Таран Вікторія Євгенівна

(прізвище, ім'я, по батькові)

\_\_\_\_\_ (підпис) \_\_\_\_\_

Науковий керівник доцент кафедри інформаційної безпеки, к.т.н.,

Коломицев М. В.

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

\_\_\_\_\_ (підпис) \_\_\_\_\_

Рецензент СЕО “Октал Сек’юріті”, к.т.н., Шантир А. С.

(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.  
Здобувач ступеня магістра \_\_\_\_\_

\_\_\_\_\_ (підпис) \_\_\_\_\_

Київ – 2023 року

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

Рівень вищої освіти – другий (магістерський)  
Спеціальність – 125 «Кібербезпека»  
Освітньо-наукова програма «Системи, технології та математичні методи кібербезпеки»

ЗАТВЕРДЖУЮ  
Завідувач кафедри  
\_\_\_\_\_ Дмитро ЛАНДЕ  
(підпис)  
« \_\_\_\_ » \_\_\_\_\_ 2023 р.

**ЗАВДАННЯ**  
**на магістерську дисертацію здобувачу ступеня магістра**

Таран Вікторії Євгенівні  
(прізвище, ім'я, по батькові)

1. Тема дисертації: Виявлення ознак атак на критичну інфраструктуру методом аналізу мережевого протоколу,  
науковий керівник дисертації Коломицев Михайло Володимирович, к.т.н.,  
доцент кафедри інформаційної безпеки  
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «11» квітня 2023 р. №1482-с

2. Термін подання здобувачем дисертації 12.05.2023 р.

3. Об'єкт дослідження є мережева інфраструктура, зокрема мережі Інтернету та протокол маршрутизації BGP.

4. Предмет дослідження є вплив війни в Україні на мережеву інфраструктуру, зокрема на зміни маршрутизації трафіку та управління мережами в умовах воєнного конфлікту.

5. Перелік завдань, які потрібно розробити 1. Розглянути теоретичні матеріали мережевого протоколу BGP, атак та методів захисту від них. 2. Розглянути методи пошуку аномалій та визначити один з них. 3. Розробити методику

виявлення змін у трафіку BGP. 4. Провести експериментальні дослідження. 5. Проаналізувати результати.

6. Орієнтовний перелік ілюстративного матеріалу – презентація.

7. Орієнтовний перелік публікацій “Виявлення ознак атак на критичну інфраструктуру методом аналізу мережевого протоколу”.

8. Дата видачі завдання 01.01.2023

#### Календарний план

| № з/п | Назва етапів виконання магістерської дисертації                | Термін виконання етапів магістерської дисертації | Примітка |
|-------|----------------------------------------------------------------|--------------------------------------------------|----------|
| 1     | Вибір тематики роботи                                          | 01.01.2023                                       | Виконано |
| 2     | Огляд літератури                                               | 01.01.2023 – 01.02.2023                          | Виконано |
| 3     | Вибір методу пошуку аномалій                                   | 01.02.2023 – 15.02.2023                          | Виконано |
| 4     | Розробка методики виявлення змін у трафіку BGP                 | 15.02.2023 – 28.02.2023                          | Виконано |
| 5     | Збір та обробка даних                                          | 01.03.2023 – 01.04.2023                          | Виконано |
| 6     | Тестування методики                                            | 01.03.2023 – 01.04.2023                          | Виконано |
| 7     | Проведення експериментального дослідження і аналіз результатів | 01.03.2023 – 01.04.2023                          | Виконано |
| 8     | Оформлення дипломної роботи                                    | 01.04.2023 – 30.04.2023                          | Виконано |

Здобувач ступеня магістра

—  —

Вікторія ТАРАН

(власне ім'я, ПРІЗВИЩЕ)

Науковий керівник

—  —

Михайло КОЛОМИЦЕВ

(власне ім'я, ПРІЗВИЩЕ)

## РЕФЕРАТ

Робота обсягом 89 сторінок містить 26 ілюстрацій, 24 таблиць, 25 джерел за переліком посилань та 3 додатка.

У даній роботі було досліджено вплив війни в Україні на стан мережі, зокрема на BGP трафік та наявність збоїв, які пов'язані з військовими діями та кібератаками.

Була розроблена методика для виявлення змін у трафіку, що використовує метод пошуку аномалій відстані Кука. Для цього було зібрано та оброблено дані BGP, з яких було створено датасет. Датасет використовувався для застосування методу пошуку аномалій на мережевих даних. В результаті було знайдено 4 аномальні події за 2 місяці війни.

Отримана методика є корисною для виявлення змін у мережевому трафіку, які можуть бути пов'язані з кібератаками або іншими несправностями в мережі. Це дає можливість оперативно реагувати на такі зміни та запобігати можливим кібератакам. Крім того, методика може бути застосована в межах географічного регіону для виявлення змін у трафіку, які можуть вказувати на подібні проблеми в мережі.

Ключові слова: BGP протокол, BGP дані, аномалії, методи пошуку аномалій, російсько-українська війна, кібератаки, відстань Кука.

## ABSTRACT

This work is comprised of 89 pages, 26 illustrations, 24 tables, 25 references and 3 appendices.

This paper investigated the impact of the war in Ukraine on the state of the network, in particular on BGP traffic and the presence of failures associated with military operations and cyber attacks.

A technique was developed to detect changes in traffic using the Cook's distance anomaly detection method. For this, BGP data was collected and processed, from which a dataset was created. The dataset was used to apply the anomaly detection method to network data. As a result, 4 anomalous events were found in 2 months of the war.

The resulting technique is useful for detecting changes in network traffic that may be associated with cyber attacks or other network malfunctions. This makes it possible to quickly react to such changes and prevent possible cyber attacks. In addition, the technique can be applied within a geographic region to detect changes in traffic that may indicate similar problems in the network.

Keywords: BGP protocol, BGP data, anomalies, anomaly detection methods, Russian-Ukrainian war, cyber attacks, Cook's distance.

## ЗМІСТ

|                                                                             |    |
|-----------------------------------------------------------------------------|----|
| ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,<br>СКОРОЧЕНЬ І ТЕРМІНІВ ..... | 8  |
| ВСТУП .....                                                                 | 9  |
| 1 АНАЛІЗ ОСОБЛИВОСТЕЙ МЕРЕЖЕВИХ ПРОТОКОЛІВ.....                             | 11 |
| 1.1 OSPF протокол .....                                                     | 11 |
| 1.2 EIGRP протокол .....                                                    | 13 |
| 1.3 BGP протокол .....                                                      | 14 |
| 1.3.1 Структура BGP протоколу .....                                         | 17 |
| 1.4 BGP повідомлення .....                                                  | 19 |
| 1.4.1 BGP OPEN.....                                                         | 21 |
| 1.4.2 BGP UPDATE .....                                                      | 22 |
| 1.4.3 BGP NOTIFICATION.....                                                 | 23 |
| 1.4.4 BGP KEEPALIVE .....                                                   | 23 |
| 1.4.5 BGP ROUTE-REFRESH.....                                                | 24 |
| 1.5 Атаки на BGP протокол.....                                              | 24 |
| 1.5.1 Атака викрадення префіксів .....                                      | 25 |
| 1.5.2 Атака витоку маршрутів.....                                           | 26 |
| 1.5.3 Відмова в обслуговуванні .....                                        | 27 |
| 1.5.4 Інші атаки.....                                                       | 28 |
| 1.5.5 Методи захисту від атак .....                                         | 29 |
| 1.5.6 Cloudflare як сервіс від DDoS атак .....                              | 30 |
| Висновки до розділу 1 .....                                                 | 31 |
| 2 АНАЛІЗ МЕТОДІВ ПОШУКУ АНОМАЛІЙ .....                                      | 33 |
| 2.1 Відстань Кука .....                                                     | 33 |

|                                                              |    |
|--------------------------------------------------------------|----|
| 2.2 Інтерквартильний діапазон .....                          | 34 |
| 2.3 Ізольований ліс .....                                    | 37 |
| 2.4 Детермінант мінімальної коваріації .....                 | 39 |
| 2.5 Парето.....                                              | 41 |
| 2.6 Порівняння методів.....                                  | 42 |
| Висновки до розділу 2 .....                                  | 47 |
| 3 РОЗРОБКА МЕТОДИКИ ВИЯВЛЕННЯ ЗМІН У ТРАФІКУ BGP .....       | 48 |
| 3.1 Основні складові методики виявлення змін у трафіку ..... | 48 |
| 3.2 Практична реалізація методики.....                       | 51 |
| 3.3 Аналіз результатів.....                                  | 55 |
| Висновки до розділу 3 .....                                  | 59 |
| 4 РОЗРОБЛЕННЯ СТАРТАП ПРОЕКТУ .....                          | 60 |
| 4.1 Опис ідеї стартап-проєкту.....                           | 60 |
| 4.2 Технологічний аудит ідеї проєкту.....                    | 62 |
| 4.3 Аналіз ринкових можливостей запуску стартап-проєкту..... | 62 |
| 4.4 Розроблення ринкової стратегії проєкту .....             | 68 |
| 4.5 Розроблення маркетингової програми проєкту .....         | 71 |
| Висновки до розділу 4 .....                                  | 73 |
| ВИСНОВКИ.....                                                | 74 |
| ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ.....                                 | 75 |
| Додаток А.....                                               | 78 |
| Додаток Б .....                                              | 82 |
| Додаток В.....                                               | 85 |

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,  
СКОРОЧЕНЬ І ТЕРМІНІВ

BGP – Border Gateway Protocol

OSPF – Open Shortest Path First

EIGRP – Enhanced Interior Gateway Routing Protocol

DDoS – distributed denial-of-service

IGRP – Interior Gateway Routing Protocol

AS – autonomous system

DNS – Domain Name System

DNSSEC – Domain Name System Security Extensions

CDN - content delivery network

## ВСТУП

Актуальність дослідження. Дослідження має на меті вивчення впливу війни в Україні на мережеву інфраструктуру та пов'язати збої в мережі з руйнування в інфраструктурі. Ця війна має значний вплив на регіональну мережеву інфраструктуру, включаючи зміни в маршрутизації трафіку в Інтернеті. BGP (Border Gateway Protocol) є одним з найважливіших протоколів мережевої інфраструктури, і його дослідження може допомогти краще зрозуміти, як війна в Україні впливає на мережеву інфраструктуру. Це дослідження може мати важливе значення для розвитку стратегій управління мережами та резервування маршрутів з метою забезпечення надійності та стабільності мережевої інфраструктури в умовах війни або інших надзвичайних ситуацій.

Крім того, важливо зауважити, що рівень збоїв у мережі залишається значно вищим, ніж зазвичай, навіть під час нормальної роботи мережі, що може бути наслідком руйнування інфраструктури, можливо, спричиненого кібератаками. Додатково варто відзначити, що перед значущими подіями або змінами можна спостерігати збільшення рівня збоїв в BGP трафіку, що також може бути наслідком підготовки до кібератак або інших несправностей в мережі.

Метою дослідження є розробка методики виявлення змін у трафіку BGP. Для досягнення мети необхідно розв'язати задачі дослідження, які полягають в:

- 1) аналізі мережевого протоколу BGP, атак та методів захисту
- 2) аналізі методів пошуку аномалій в BGP даних
- 3) розробці методики виявлення змін у трафіку BGP
- 4) зборі та обробці BGP даних
- 5) аналізі збоїв та пошуку аномалій в наборі даних
- 6) перевірці розробленої методики на наборі даних
- 7) аналізі зв'язку між збоями і зовнішніми подіями

Для розв'язання задачі необхідно вирішити такі завдання:

- 1) Розглянути теоретичні матеріали мережевого протоколу BGP, атак та методів захисту від них
- 2) Розглянути методи пошуку аномалій та визначити один з них
- 3) Провести експериментальні дослідження
- 4) Проаналізувати результати

Об'єктом дослідження є мережева інфраструктура, зокрема мережі Інтернету та протокол маршрутизації BGP.

Предметом дослідження є вплив війни в Україні на мережеву інфраструктуру, зокрема на зміни маршрутизації трафіку та управління мережами в умовах воєнного конфлікту.

При розв'язанні поставлених завдань використовувались такі методи дослідження: аналіз BGP повідомлень, аналіз статистичних даних.

Наукова новизна полягає в розробці методики виявлення змін у трафіку та вивчення впливу воєнного конфлікту на мережеву інфраструктуру.

Практичне значення. Результати даного дослідження пов'язують зміни мережевого трафіку з воєнним конфліктом в Україні основуючись на розробленій в цій роботі методиці.

Апробація результатів та публікації. Результати цієї роботи були представлені в збірнику XXI-ої всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених “Теоретичні і прикладні проблеми фізики, математики і інформатики”.

## 1 АНАЛІЗ ОСОБЛИВОСТЕЙ МЕРЕЖЕВИХ ПРОТОКОЛІВ

В цьому розділі буде представлено аналіз літератури та буде визначено основні терміни та поняття, що належать темі цієї роботи. Спочатку будуть проаналізовані мережеві протоколи динамічної маршрутизації такі, як BGP, OSPF та EIGRP. Далі буде детально розглянуто протокол BGP, його структуру, BGP Update повідомлення. Далі будуть визначені атаки на BGP протокол та методи захисту від них.

Протоколи динамічної маршрутизації в комп'ютерних мережах використовуються для автоматичного визначення оптимальних маршрутів передачі даних між різними вузлами мережі. Основна ідея полягає в тому, що вузли мережі взаємодіють між собою, обмінюючись інформацією про стан мережі та використовуючи цю інформацію для визначення оптимального маршруту.

### 1.1 OSPF протокол

Протокол OSPF (Open Shortest Path First) є одним із сімейства протоколів IP-маршрутизації та є протоколом внутрішнього шлюзу (IGP) для Інтернету, який використовується для розподілу інформації про IP-маршрутизацію в одній автономній системі (AS) у мережі IP.

Протокол OSPF — це протокол маршрутизації за станом зв'язку, що означає, що маршрутизатори обмінюються інформацією про топологію зі своїми найближчими сусідами. Інформація про топологію розповсюджується по всій AS, тому кожен маршрутизатор в AS має повне уявлення про топологію AS. Це зображення потім використовується для обчислення наскрізних шляхів через AS, як правило, з використанням варіанту алгоритму Дейкстри. Таким чином, у протоколі маршрутизації за станом зв'язку адреса наступного переходу, на яку пересилаються дані, визначається шляхом вибору найкращого наскрізного шляху до кінцевого пункту призначення [1].

Основна перевага протоколу маршрутизації стану зв'язку, такого як OSPF, полягає в тому, що повне знання топології дозволяє маршрутизаторам обчислювати маршрути, які задовольняють певним критеріям. Це може бути корисним для цілей організації дорожнього руху, коли маршрути можуть бути обмежені відповідно до певних вимог щодо якості обслуговування. Основним недоліком протоколу маршрутизації стану зв'язку є те, що він погано масштабується, оскільки до домену маршрутизації додається більше маршрутизаторів. Збільшення кількості маршрутизаторів збільшує розмір і частоту оновлення топології, а також тривалість часу, необхідного для розрахунку наскрізних маршрутів. Ця відсутність масштабованості означає, що протокол маршрутизації стану зв'язку непридатний для маршрутизації через Інтернет загалом, що є причиною того, чому IGP направляють трафік лише в межах однієї AS.

Кожен маршрутизатор OSPF розповсюджує інформацію про свій локальний стан (придатні для використання інтерфейси та доступні сусіди, а також вартість використання кожного інтерфейсу) іншим маршрутизаторам за допомогою повідомлення про оголошення про стан зв'язку (LSA). Кожен маршрутизатор використовує отримані повідомлення для створення ідентичної бази даних, яка описує топологію AS.

На основі цієї бази даних кожен маршрутизатор розраховує власну таблицю маршрутизації за допомогою найкоротшого шляху (SPF) або алгоритму Дейкстри. Ця таблиця маршрутизації містить усі пункти призначення, про які знає протокол маршрутизації, пов'язані з IP-адресою наступного переходу та вихідним інтерфейсом.

- Протокол перераховує маршрути при зміні топології мережі за допомогою алгоритму Дейкстри та мінімізує трафік протоколу маршрутизації, який він генерує.
- Він забезпечує підтримку кількох шляхів однакової вартості.

- Він забезпечує багаторівневу ієрархію (дворівневу для OSPF), яка називається «маршрутизація області», так що інформація про топологію в межах певної області AS прихована від маршрутизаторів за межами цієї області. Це забезпечує додатковий рівень захисту маршрутизації та зменшує трафік протоколу маршрутизації.
- Усі обміни протоколами можна автентифікувати, щоб лише довірені маршрутизатори могли приєднатися до обміну маршрутизацією для AS.

## 1.2 EIGRP протокол

Enhanced Interior Gateway Routing Protocol — це розширений протокол дистанційної векторної маршрутизації, заснований на принципах протоколу Interior Gateway Routing Protocol (IGRP). Він має унікальну характеристику, яка покращує робочу здатність і швидко швидкість сходження. Він може визначити вектор відстані найкоротшого шляху та працює за принципом Interior Gateway Routing Protocol, безкласового протоколу маршрутизації. Він використовує такі показники, як пропускна здатність, навантаження та затримки, щоб розрахувати найкоротший оптимальний мережевий маршрут. Це технологічно більш просунутий протокол маршрутизації на основі вектора відстані. Щоб обмінюватися інформацією за допомогою EIGRP, насамперед, маршрутизатори повинні стати сусідами EIGRP, а потім EIGRP використовує багатоадресну адресу для обміну інформацією.

Основна логіка протоколу Enhanced Interior Gateway Routing Protocol базується на концепції автономної системи. У системі, де кожен маршрутизатор має стати сусідами EIGRP і кожна система позначена як сусіди в Enhanced Interior Gateway Routing Protocol матиме той самий налаштований номер системи.

Enhanced Interior Gateway Routing Protocol (EIGRP) — це протокол динамічної маршрутизації, який використовується для прийняття рішень і

налаштування маршрутизаторів. EIGRP надсилає лише поетапні оновлення, що скорочує навантаження на маршрутизатори та зменшує обсяг інформації, що передається. EIGRP — це класичний гібридний протокол, який підтримує безкласову маршрутизацію, він підтримує автоматичне та ручне підсумовування в інтерфейсі з підтримкою EIGRP. Це аутентифікація MD5 на маршрутизаторах, на яких працює EIGRP, а також має максимальну кількість переходів, близьку до 255. EIGRP виконує балансування навантаження на шляхах рівної та нерівної вартості.

Щоб виконувати функції EIGRP, він створює три таблиці, які:

- Стіл сусідів. Містить інформацію про маршрутизатори та відносини сусідства з тими, кого було встановлено. Має такі поля, як Handle, Address, Interface, Hold Time, Uptime, Smooth Round Trip Time, Retransmission Timeout, Queue Count, Sequence Number.
- Таблиця топології. Містить інформацію про всі шляхи до мереж, які розуміють маршрутизатори EIGRP. Таблиця топології містить наступні поля: Passive, Feasible Distance, Advertised distance, Feasible distance.
- Таблиця маршрутизації. Зберігає маршрути, які на даний момент активні для надсилання пакетів до мережі. Він зберігає оптимальний маршрут для місця призначення від відправника. Таблиця маршрутизації містить такі поля D, 90/ 5632. через 11.0.0.2, GigabitEthernet0/1 [2].

### 1.3 BGP протокол

BGP (Border Gateway Protocol) є одним з найважливіших протоколів мережевої інфраструктури Інтернету, який використовується для маршрутизації трафіку між різними автономними системами (AS).

Важливість цього протоколу полягає в тому, що він відповідає за передачу інформації про маршрутизацію між різними провайдерами та мережами, що

дозволяє користувачам Інтернету отримувати доступ до різноманітних ресурсів.

Актуальність BGP протоколу полягає в тому, що залежно від налаштувань та стану мережі, маршрутизація трафіку може бути змінена, що може призвести до відключення деяких регіонів від Інтернету, зниження швидкості передачі даних та інших проблем. Зокрема, у ситуації війни, коли існує ризик знищення мережевої інфраструктури або відключення деяких регіонів від Інтернету, дослідження BGP може допомогти зрозуміти, як можна покращити стійкість мережевої інфраструктури та забезпечити надійну маршрутизацію трафіку в подібних ситуаціях.

Основні характеристики BGP:

- BGP є протоколом маршрутизації між декількома автономними системами (AS). Іншими словами, BGP використовується для обміну маршрутною інформацією між різними провайдерами та організаціями.
- BGP базується на моделі векторної маршрутизації, де кожен маршрутизатор обмінюється інформацією з сусідніми маршрутизаторами про найкоротший шлях до кожної мережі.
- BGP використовується для встановлення зовнішніх маршрутів - маршрутів до мереж, які належать до інших автономних систем.
- BGP забезпечує розподіл маршрутів за критеріями, які встановлюються власниками мереж.
- BGP забезпечує функцію маршрутизації з базою даних, що включає інформацію про мережі та маршрути до цих мереж.
- BGP може використовуватись для встановлення Peer-to-Peer-з'єднань між провайдерами, що дозволяє їм обмінюватись трафіком напряму.
- BGP забезпечує механізми для перевірки коректності маршрутної інформації, що дозволяє запобігти поширенню неправильних маршрутів.
- BGP може використовувати різні протоколи для обміну маршрутною інформацією.

BGP використовується для обміну маршрутною інформацією між автономними системами, які використовують різні маршрутизатори. Кожен маршрутизатор BGP має список всіх мереж, які він може доставити, та список всіх мереж, які він знає від інших маршрутизаторів. Коли мережа стає доступною або відключається, маршрутизатор BGP повідомляє про це інші маршрутизатори [3].

Розглянувши декілька протоколів динамічної маршрутизації, ми склали список, чому BGP вважається важливішим, ніж OSPF і EIGRP:

- Підключення до Інтернету: BGP — це протокол, який забезпечує маршрутизацію між різними автономними системами (AS) у глобальному Інтернеті, забезпечуючи глобальне підключення та зв'язок між мережами, якими керують різні організації, постачальники послуг і країни. OSPF і EIGRP, з іншого боку, використовуються для маршрутизації в межах однієї AS.
- Масштаб і складність: BGP розроблений для обробки великомасштабних і складних вимог до маршрутизації глобального Інтернету, з можливістю обробки величезної кількості маршрутів, різноманітних топологій і кількох провайдерів. OSPF і EIGRP, з іншого боку, зазвичай використовуються в невеликих мережах з меншою кількістю маршрутів і простішою топологією.
- Маршрутизація на основі політики: BGP надає розширені можливості маршрутизації на основі політики, що дозволяє мережевим адміністраторам визначати та впроваджувати складні політики маршрутизації на основі різних атрибутів, таких як шлях AS, атрибути маршруту, рядки спільноти тощо. Це дає змогу детально контролювати рішення щодо маршрутизації та проектування трафіку, які можуть бути недоступні або настільки складні в OSPF та EIGRP.
- Міждоменна маршрутизація: BGP спеціально розроблений для міждоменної маршрутизації, що дозволяє обмінюватися інформацією

про маршрутизацію та узгоджувати маршрути між різними автономними системами. Це важливо для належного функціонування глобальної мережі Інтернет і встановлення однорангових і транзитних відносин між автономними системами. OSPF і EIGRP, з іншого боку, є внутрішніми протоколами маршрутизації, які використовуються в одній AS і не мають можливостей міждоменного маршрутизації.

- Магістральна інфраструктура Інтернету: BGP — це фактично стандартний протокол, який використовується постачальниками послуг Інтернету (ISP) і великими мережами для обміну інформацією про маршрутизацію та встановлення глобального з'єднання. Магістральна інфраструктура Інтернету на основі BGP є важливою частиною глобальної інфраструктури Інтернету, що забезпечує маршрутизацію трафіку між різними регіонами та континентами. OSPF і EIGRP зазвичай використовуються в невеликих мережах і не розроблені для складності та масштабу глобальної мережі Інтернет.

Таким чином, у той час як OSPF і EIGRP є важливими протоколами маршрутизації для внутрішньодоменового маршрутизації в автономних системах, BGP має вирішальне значення для міждоменного маршрутизації в глобальній мережі Інтернет, надаючи розширені можливості маршрутизації, маршрутизацію на основі політики та масштабованість для вирішення складних завдань [4].

### 1.3.1 Структура BGP протоколу

Прикордонний маршрутизатор - це мережевий маршрутизатор, який забезпечує зв'язок між декількома мережами. Він використовується в BGP (Border Gateway Protocol) для обміну маршрутами між різними автономними системами (AS). Кожен AS має свій власний прикордонний маршрутизатор, який відповідає за забезпечення з'єднання з іншими AS та маршрутизацію трафіку між ними.

У BGP кожен прикордонний маршрутизатор використовується для визначення найкращого маршруту до певного призначення. Кожен прикордонний маршрутизатор вирішує, до якого наступного хопу (наступного маршрутизатора в шляху до призначення) він повинен відправити пакет.

Кожен прикордонний маршрутизатор в BGP обмінюється інформацією про доступні маршрути з іншими прикордонний маршрутизатор. Це дозволяє кожному прикордонний маршрутизатор знати про доступні маршрути до призначення та визначати найкращий маршрут.

У BGP кожен прикордонний маршрутизатор також може використовуватися для визначення пріоритету маршрутів. Якщо доступні два маршрути з однаковою метою, прикордонний маршрутизатор може використовувати різні критерії для вибору найкращого маршруту. Наприклад, можуть використовуватися такі критерії, як пропускна здатність, затримка, маршрут з меншою кількістю хопів тощо.

Крім того, прикордонний маршрутизатор в BGP може використовуватися для забезпечення безпеки мережі. Це досягається шляхом фільтрації пакетів, які проходять через маршрутизатор, та контролю доступу користувачів до мережі, щоб запобігти можливим атакам або шкідливому трафіку. Прикордонний маршрутизатор також може застосовувати різні методи шифрування, щоб забезпечити конфіденційність даних, що передаються через мережу.

Отже, прикордонний маршрутизатор є важливим елементом в мережевому інфраструктурі BGP, дозволяючи встановлювати зв'язок між різними AS та забезпечувати ефективну маршрутизацію між ними.

Крім прикордонного маршрутизатору, структура BGP протоколу може бути описана наступними компонентами:

- BGP додаткові поля - це поля, які містять додаткову інформацію про маршрутизацію, наприклад, вартість маршруту, максимальну швидкість передачі даних та інші параметри.

- BGP атрибути - це параметри, які використовуються для визначення маршруту та його вартості. Атрибути можуть бути передані від одного AS до іншого, щоб допомогти встановити оптимальний маршрут.
- BGP сесії - це з'єднання між двома роутерами, які використовують BGP для передачі інформації про маршрутизацію. Сесії можуть бути з'єднаннями між двома роутерами у межах одного AS, або між роутерами з різних AS.
- BGP таблиці маршрутизації - це таблиці, які містять інформацію про маршрутизацію, яку отримали роутери від інших AS через BGP. Таблиці маршрутизації зазвичай зберігаються на кожному роутері, що дозволяє їм використовувати цю інформацію для прийняття рішень щодо маршрутизації трафіку [5].

BGP працює на основі обміну повідомленнями між BGP-маршрутизаторами, що дозволяє кожному маршрутизатору знати про всі маршрути в мережі та вибрати оптимальний маршрут. BGP використовує векторну маршрутизацію, що означає, що кожен маршрутизатор відомий з певним вектором, що включає мережу і наступний маршрутизатор в шляху до цієї мережі.

Під час обміну повідомленнями, маршрутизатори використовують атрибути, що дозволяють встановлювати пріоритети між маршрутами. Наприклад, атрибут "AS path" вказує на список AS, які маршрут пройшов, що дозволяє визначити, який маршрут є більш надійним та має менше перешкод.

BGP може використовуватися для різних цілей, таких як забезпечення надійності мережі, збільшення пропускної здатності, забезпечення балансування навантаження та ін. BGP є важливим протоколом для функціонування Інтернету та забезпечення безпеки та стабільності мережі.

#### 1.4 BGP повідомлення

Існує кілька типів повідомлень BGP:

- OPEN - це перше повідомлення, яке відправляється при з'єднанні між двома BGP-маршрутизаторами. Воно містить параметри, необхідні для налагодження з'єднання між маршрутизаторами, такі як версія BGP та AS номер.
- UPDATE - це повідомлення використовується для передачі інформації про маршрутизацію між маршрутизаторами BGP. Воно містить інформацію про маршрути, доступні через маршрутизатор, та метрики для визначення найкращого маршруту.
- NOTIFICATION - це повідомлення використовується для повідомлення про помилки або проблеми в процесі BGP-маршрутизації. Якщо маршрутизатор отримує повідомлення NOTIFICATION, він закриває з'єднання з іншим маршрутизатором.
- KEEPALIVE - це повідомлення використовується для підтримки активного з'єднання між маршрутизаторами BGP, якщо протягом певного часу не було передано жодного повідомлення.
- ROUTE-REFRESH – це повідомлення використовується для запиту інформації про маршрутизацію зазначеного сімейства адрес.

Ці повідомлення можуть використовуватись разом або окремо для налагодження та підтримки з'єднання між маршрутизаторами BGP.

Всі перелічені повідомлення мають однаковий заголовок [6]:

- Маркер — 16-байтове поле використовується для розмежування повідомлень BGP. Усі маркери мають бути одиницями.
- Довжина — двобайтове ціле число без знаку вказує загальну довжину повідомлення.
- Тип — це однобайтове ціле число без знаку вказує на код типу повідомлення. Визначено такі коди типів: 1 – OPEN, 2 - UPDATE, 3 - NOTIFICATION, 4 - KEEPALIVE і 5 - ROUTE-REFRESH. Перші чотири перелічені коди визначено в RFC 1771, а останній перерахований код визначено в RFC 2918.

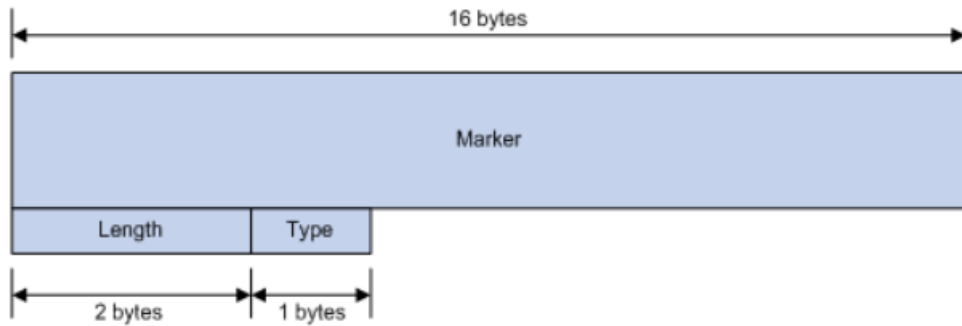


Рисунок 1.1 – Заголовок BGP повідомлень

### 1.4.1 BGP OPEN

Основні поля повідомлення BGP такі:

- Версія — це однобайтове ціле число без знаку вказує номер версії протоколу. Поточна версія BGP – 4.
- Моя автономна система — це двобайтове ціле число без знаку вказує на номер автономної системи відправника.
- Час утримання — під час встановлення однорангових відносин дві сторони домовляються про однаковий час утримання. Якщо протягом часу утримання від однорангової мережі не отримано Keepalive або Update, з'єднання BGP вважається неактивним.
- Ідентифікатор BGP — IP-адреса, яка ідентифікує маршрутизатор BGP.
- Довжина додаткових параметрів встановлюється на 0, якщо жоден необов'язковий параметр недоступний.
- Додаткові параметри — використовуються для багатопрокольних розширень та інших функцій.

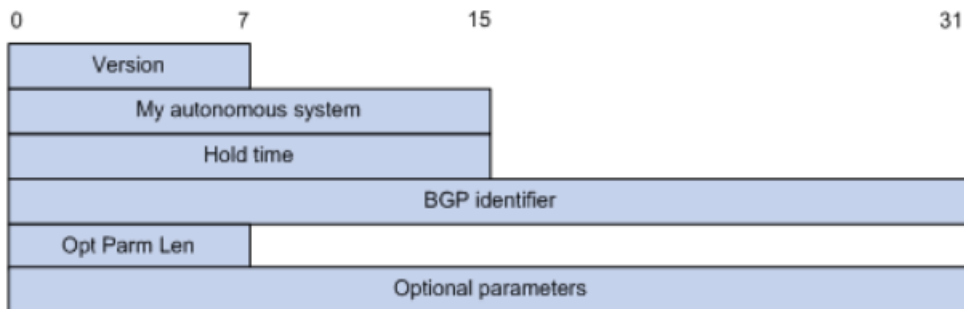


Рисунок 1.2 – Структура BGP OPEN повідомлення

### 1.4.2 BGP UPDATE

Повідомлення UPDATE можна використовувати для оголошення нового маршруту або скасування існуючого. Як вказує назва, оголошення про маршрут використовується для оголошення маршруту до AS, де розміщено мережевий префікс, що розглядається, і дозволяє іншим вузлам знати про альтернативні шляхи зв'язку. Він містить таку інформацію, як:

- Ідентифікатор маршрутизатора моніторингу, тобто номер AS та IP-адреса маршрутизатора BGP, який отримав повідомлення оновлення
- Мітка часу отримання повідомлення
- Префікс мережі
- Шлях (послідовність ASN), який необхідно пройти, щоб досягти вихідної AS префікса.

З іншого боку, відкликання використовується для скасування існуючого шляху маршрутизації. Відкликання містить таку інформацію, як:

- Ідентичність моніторингового маршрутизатора
- Мітка часу отримання повідомлення
- Префікс, який розглядається.

Рисунок 1.3 показує два повідомлення BGP UPDATE; оголошення у верхній частині фігури та відкликання у нижній частині. Поля оголошення: протокол BGP, час Unix у секундах, тип оновлення («A» для оголошення), IP-адреса маршрутизатора, що повідомляє, номер AS AS, що повідомляє, префікс, шлях AS, джерело оновлення. («IGP» для протоколу внутрішнього шлюзу або «EGP» для протоколу зовнішнього шлюзу), наступний стрибок, локальні налаштування, бажаний шлях, згруповані маршрути зі схожими політиками, атомарний агрегатор і агрегатор. Атомарний агрегатор — це позначка, яка показує, що деяку інформацію було втрачено, коли AS агрегував префікси. Поле агрегатора ідентифікує AS і маршрутизатор, які об'єднали префікси, отримані від різних вузлів, у єдиний префікс. Поля відкликання: протокол BGP, час Unix у секундах, тип оновлення («W» для

відкликання), IP-адреса анонсуєного маршрутизатора, номер AS анонсуєного AS, відкликаний префікс.

```
BGP4MP|1064963009|A|129.250.0.11|2914|202.128.64.0/19|
  2914 701 18864 9548|IGP|129.250.0.11|0|58|2914:420
  2914:2000 2914:3000|NAG||
```

```
BGP4MP|1064963016|W|64.200.199.3|7911|200.49.32.0/22
```

Рисунок 1.3 – Приклад BGP UPDATE повідомлення

### 1.4.3 BGP NOTIFICATION

Основні поля BGP NOTIFICATION повідомлення:

- Код помилки — тип сповіщення.
- Підкод помилки — конкретна інформація про характер повідомленої помилки.
- Дані — використовуються для діагностики причини сповіщення. Вміст поля даних залежить від коду та підкоду помилки. У поле даних можуть бути записані помилкові дані. Довжина поля даних є змінною.



Рисунок 1.4 – Структура BGP NOTIFICATION повідомлення

### 1.4.4 BGP KEEPALIVE

BGP не використовує механізм підтримки активності на основі транспортного протоколу, щоб визначити, чи доступні однорангові мережі. Натомість повідомлення KEEPALIVE обмінюються між одноранговими вузлами досить часто, щоб не призвести до закінчення таймера утримання. Розумний максимальний час між повідомленнями KEEPALIVE становитиме одну третину інтервалу часу утримання. Повідомлення KEEPALIVE не повинні надсилатися частіше, ніж одне на секунду. Реалізація може регулювати швидкість, з якою вона надсилає повідомлення KEEPALIVE, як функцію інтервалу часу утримання.

Якщо узгоджений інтервал часу утримання дорівнює нулю, то періодичні повідомлення KEEPALIVE не повинні надсилатися.

Повідомлення KEEPALIVE складається лише з заголовка повідомлення та має довжину 19 октетів [7].

#### 1.4.5 BGP ROUTE-REFRESH

Повідомлення про оновлення маршруту надсилається одноранговому вузлу, щоб запитати інформацію про маршрутизацію вказаного сімейства адрес. Структура:

- AFI — ідентифікатор родини адрес.
- Res — Зарезервовано; встановити 0.
- SAFI — наступний ідентифікатор родини адрес.

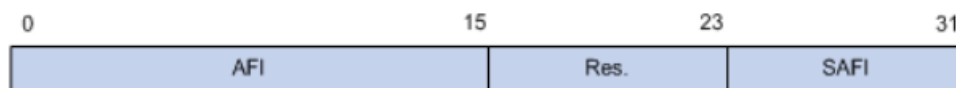


Рисунок 1.5 - Структура BGP ROUTE-REFRESH повідомлення

#### 1.5 Атаки на BGP протокол

Протягом останніх 30 років протокол BGP дозволив мережевим операторам застосовувати найрізноманітніші політики маршрутизації між AS. Дивовижно, як цей протокол ефективно підтримував постійно зростаючу кількість підмереж і автономних систем, а також еволюцію Інтернету від переважно ієрархічної структури, що складається з клієнтів і провайдерів, до структури, де піринг і IXP стають більш важливими з кожним днем.

Незважаючи на всі його хороші якості, BGP демонструє кілька вразливостей, які, якщо їх використати, можуть спричинити хвилі по всьому Інтернету. Корінь проблеми полягає в тому, що BGP було задумано на ранній стадії розвитку Інтернету, коли було лише кілька гравців. Отже, його дизайн не враховував захист від навмисних або випадкових помилок, тому шкідливі або неправильно налаштовані джерела можуть потенційно поширювати

підроблену інформацію про маршрутизацію по всьому Інтернету, використовуючи цей недолік захисту. Ще гірше, джерелом підробленої або зловмисної інформації про маршрутизацію може бути або справжній одноранговий вузол BGP, або фальшивий одноранговий вузол, оскільки BGP працює на TCP/IP і, отже, піддається кожній класичній атаці TCP/IP, такий як IP-спуфінг.

Частково проблему можна вирішити, застосувавши криптографічну автентифікацію на кожному вузлі BGP, але це не допоможе зупинити розповсюдження фальшивої інформації по всьому Інтернету з законних неправильно налаштованих джерел (витік маршрутів), із законних джерел, які або не застосовували криптографічну автентифікацію на усі або з джерел, які навмисно оголосили фальшиву інформацію про маршрутизацію (викрадення префіксів).

#### 1.5.1 Атака викрадення префіксів

Викрадення префіксів – це навмисне створення фальшивої інформації про маршрутизацію; Причин, що стоять за ними, є безліч, які важко зрозуміти:

- Зловмисник може оголосити маршрути, щоб порушити роботу служб, що працюють поверх IP-простору, охопленого маршрутами, або захопити трафік для аналізу конфіденційної інформації, що надходить до цієї служби.
- Зловмисник також може просто оголосити маршрути зі створеним AS-шляхом, щоб показати підроблені сусідні з'єднання на відомих веб-сайтах.
- Або, що ще гірше, зловмисник може захопити трафік, щоб маніпулювати потоковими пакетами за власним бажанням, або просто захоче використати невикористані маршрути для створення спаму.

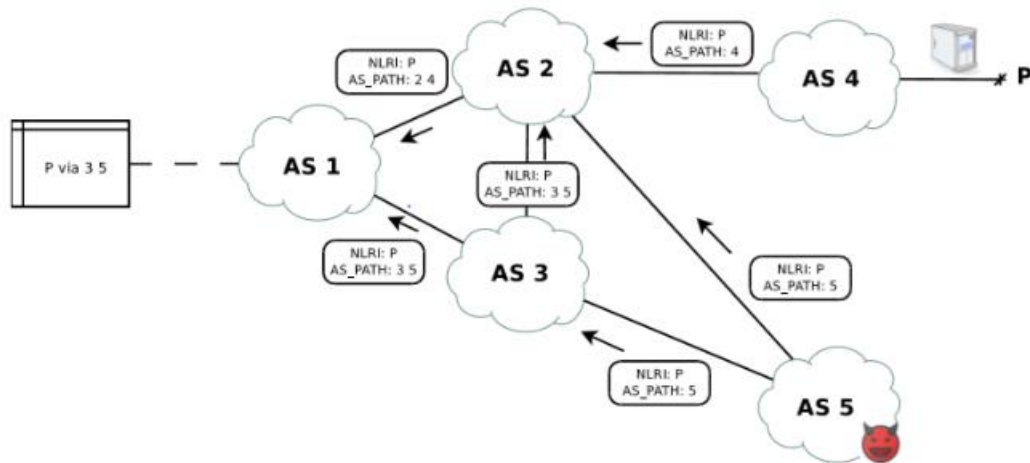


Рисунок 1.6 – Приклад топології мережі для атаки викрадення префіксів

Давайте розглянемо наведений вище сценарій, щоб краще зрозуміти, як можна виконувати викрадення префіксів. AS 5 є зловмисником і підключений до Інтернету через двох постачальників: AS 2 і AS 3. AS 1 є клієнтом AS 2 і постачальником AS 3, тоді як AS 4 є одноранговим AS 2, а AS 2 є постачальником AS 3. Нарешті, ми припускаємо, що AS 2 правильно налаштував свої вхідні фільтри BGP, тоді як AS 1 і AS 3 мають слабку конфігурацію фільтра (якщо така є).

У цьому першому сценарії AS 5 оголосить мережу P, яка належить і вже оголошена AS 4. Через описані вище конфігурації фільтрів повідомлення про оновлення, оголошене AS 5, буде видалено AS 2, тоді як воно буде прийняте AS 2. AS 3. Потім AS 3 повідомить про це своїм постачальникам (AS 1 і AS 2). AS 2 знову відкине пакет через фільтри, тоді як AS 1 прийме його. Якщо процес прийняття рішення про BGP AS 1 вибере як найкращий маршрут шлях від AS 5, тоді трафік від AS 1 до AS 5 буде надіслано до зловмисника, а не до належного власника [8].

### 1.5.2 Атака витоку маршрутів

Витоки маршрутів — це ненавмисне створення фіктивної інформації про маршрутизацію, спричинене неправильними конфігураціями маршрутизатора, такими як помилки в конфігурації фільтра або неправильне

походження чужої мережі. Навіть якщо це ненавмисно, наслідки витоку маршруту можуть бути такими ж, як і викрадення префікса.

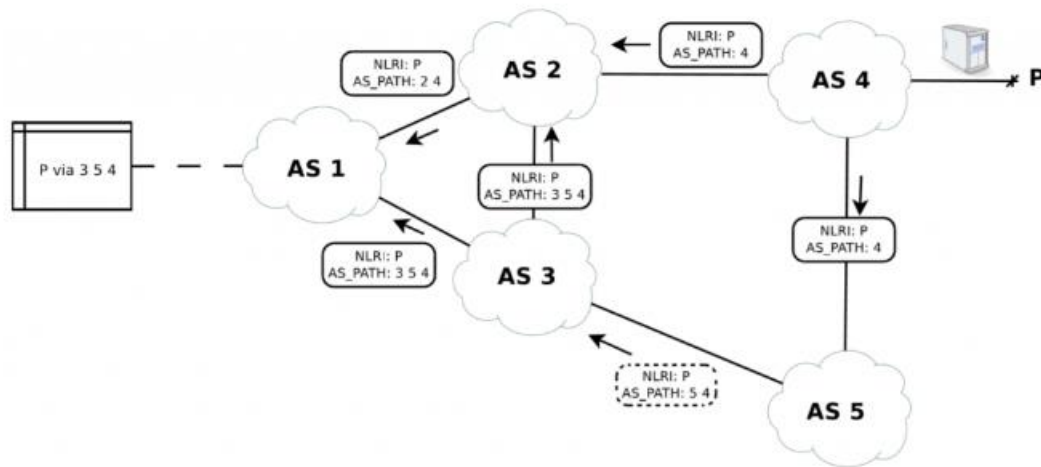


Рисунок 1.7 - Приклад топології мережі для атаки витоку маршрутів

Розглянемо ту саму топологію, яку ми використовували у прикладі викрадення префікса, з тією різницею, що AS 5 тепер є звичайним оператором мережі, який просто застосовував неправильні фільтри BGP, наприклад «прийняти все від мого постачальника, сповістити про все моєму постачальнику». Це, на жаль, нерідкий випадок, і це помилка, яку можуть зробити кілька AS під час переходу від одного постачальника до кількох постачальників.

Через цю помилку тепер AS 5 поширюватиме все, що отримує від свого постачальника, іншому постачальнику. Потім цей фрагмент інформації про маршрутизацію пошириться по всьому Інтернету, і AS почнуть маршрутизувати трафік залежно від результату процесу прийняття рішення щодо BGP кожної AS.

### 1.5.3 Відмова в обслуговуванні

DDoS-атака на атрибути BGP передбачає переповнення мережі трафіком шляхом маніпулювання значеннями атрибутів BGP. Ось деякі загальні метрики атрибутів BGP, на які можна націлити таку атаку:

- AS\_PATH: цей атрибут визначає шлях автономної системи (AS), яким маршрут BGP пройшов до місця призначення. Зловмисник може маніпулювати атрибутом AS\_PATH, щоб ввести неправдиву інформацію в мережу, що призведе до перенаправлення трафіку до шкідливих місць призначення.
- NEXT\_HOP: цей атрибут визначає IP-адресу наступного маршрутизатора для перенаправлення трафіку на шляху до місця призначення. Маніпулюючи атрибутом NEXT\_HOP, зловмисник може спричинити перенаправлення легітимного трафіку до підробленого або зловмисного пункту призначення.
- ORIGIN: цей атрибут визначає, як було створено маршрут BGP, або всередині AS, або ззовні з іншої AS. Зловмисник може маніпулювати атрибутом ORIGIN, щоб ввести неправдиву інформацію в мережу, що призведе до перенаправлення трафіку до шкідливих місць призначення.
- COMMUNITY: Цей атрибут використовується для групування маршрутів BGP у спільноти, які можна використовувати для застосування політик маршрутизації. Зловмисник може маніпулювати атрибутом COMMUNITY, щоб обійти політику безпеки та спрямувати трафік до зловмисних цілей.
- LOCAL\_PREF: цей атрибут використовується для вказівки переваги маршруту BGP в AS. Зловмисник може маніпулювати атрибутом LOCAL\_PREF, щоб вплинути на рішення щодо маршрутизації та спричинити перенаправлення трафіку до шкідливих місць призначення.

#### 1.5.4 Інші атаки

До списку атак на BGP протокол можна віднести [9]:

- Порушення конфіденційності: дані маршрутизації, що передаються в BGP, передаються відкритим текстом, тому підслуховування є можливою атакою проти конфіденційності даних маршрутизації. (Маршрутні дані конфіденційність не є звичайною вимогою.)

- Відтворення: BGP не забезпечує захист від відтворення повідомлення.
- Вставка повідомлення: BGP не забезпечує захист від вставки повідомлень. Однак, оскільки BGP використовує TCP, коли з'єднання повністю встановлено, повідомлення вставлено сторонньою особою вимагатиме точного передбачення порядкового номера або атаки з крадіжкою сесії.
- Видалення повідомлення: BGP не забезпечує захист від видалення повідомлень.
- Модифікація повідомлення: BGP не забезпечує захист від модифікації повідомлень. Синтаксично правильну модифікацію і не змінену довжину корисного навантаження TCP загалом неможливо виявити.
- Людина посередині: BGP не забезпечує захист від атаки посередині.

#### 1.5.5 Методи захисту від атак

Є кілька заходів, які можна вжити для захисту BGP від атак. Ось список деяких загальних засобів захисту та пом'якшення [10]:

- Запровадження найкращих практик безпеки BGP: це включає в себе впровадження контролю доступу, щоб обмежити, які пристрої можуть підключатися до маршрутизаторів BGP, використання безпечних протоколів для сеансів BGP і фільтрацію вхідного та вихідного трафіку BGP для запобігання несанкціонованому доступу.
- Використовуйте перевірку маршруту BGP: перевірку маршруту BGP можна використовувати, щоб переконатися, що маршрути, рекламовані сусідом BGP, є законними та авторизованими. Це можна зробити за допомогою таких інструментів, як Resource Public Key Infrastructure (RPKI).
- Реалізація BGP Flowspec: BGP Flowspec дозволяє мережевим адміністраторам створювати правила фільтрації трафіку на основі певних критеріїв, таких як IP-адреса джерела, IP-адреса призначення,

протокол і номер порту. Це може допомогти запобігти DDoS-атакам, фільтруючи зловмисний трафік на межі мережі.

- Впровадити BGP Damping: BGP Damping — це техніка, яка може бути використана для придушення реклами нестабільних маршрутів. Це може допомогти запобігти розвороту маршруту та зменшити вплив DDoS-атак на маршрутизатори BGP.
- Впровадити BGPsec: BGPsec — це протокол безпеки, який забезпечує криптографічну перевірку повідомлень маршруту BGP. Це може допомогти запобігти захопленню BGP та іншим атакам, які покладаються на маніпуляції з атрибутами BGP.
- Використовуйте інструмент моніторингу та аналізу BGP: інструмент моніторингу та аналізу BGP може допомогти виявити й попередити мережових адміністраторів про потенційні атаки на BGP, дозволяючи швидко пом'якшити будь-які проблеми.
- Підтримуйте актуальність маршрутизаторів і програмного забезпечення BGP: регулярне оновлення маршрутизаторів і програмного забезпечення BGP може допомогти гарантувати, що будь-які відомі вразливості або проблеми безпеки виправляються та усуваються.

Це лише кілька прикладів заходів, які можна вжити для захисту BGP від атак. Конкретні засоби захисту та пом'якшення, що впроваджуються, залежатимуть від потреб і вимог кожної окремої мережі.

#### 1.5.6 Cloudflare як сервіс від DDoS атак

Cloudflare — мережа доставки контенту (CDN) і постачальник послуг безпеки, який пропонує низку засобів захисту від атак BGP. Деякі способи, за допомогою яких Cloudflare може допомогти захистити BGP, включають [11]:

- Захист від DDoS: Cloudflare пропонує низку послуг захисту від DDoS, включаючи захист на мережевому рівні, який може виявляти та пом'якшувати DDoS-атаки на основі BGP.

- Anycast-маршрутизація: Cloudflare використовує anycast-маршрутизацію для розподілу трафіку між декількома центрами обробки даних, знижуючи ризик збою однієї точки та підвищуючи стійкість мережі.
- Оптимізація маршруту BGP: Cloudflare використовує оптимізацію маршруту BGP для оптимізації маршрутів BGP і мінімізації ризику перенаправлення або викрадення трафіку.
- DNSSEC: Cloudflare підтримує DNSSEC, який забезпечує криптографічну перевірку відповідей DNS і допомагає запобігти атакам на основі DNS, які можна використовувати для перенаправлення трафіку або запуску інших атак на маршрутизатори BGP.
- Фільтрація протоколу прикордонного шлюзу: Cloudflare фільтрує трафік BGP, щоб блокувати відомі шкідливі IP-адреси та запобігати несанкціонованому доступу до маршрутизаторів BGP.

Загалом Cloudflare може забезпечити додатковий рівень захисту для маршрутизаторів BGP і допомогти зменшити ризик атак на BGP.

## Висновки до розділу 1

Таким чином, в цьому розділі ми провели аналіз літератури та визначили основні терміни та поняття, пов'язані з темою нашої роботи. Ми розглянули протоколи динамічної маршрутизації, серед яких OSPF, EIGRP і BGP. Також ми ознайомились з структурою BGP протоколу та BGP UPDATE повідомлення, як основні складові цього протоколу. Далі ми визначили різні атаки на BGP протокол та методи захисту від них.

Аналіз літератури дозволив нам отримати глибше розуміння протоколу BGP, його функцій та особливостей. Ми також визначили основні терміни та поняття, які будуть використовуватися в нашій роботі, що допоможе забезпечити чіткість та зрозумілість подальших досліджень.

Детальний огляд структури BGP та BGP UPDATE повідомлення дав нам чітке уявлення про те, як цей протокол працює та взаємодіє з іншими вузлами мережі. Ми також розглянули можливі атаки на BGP протокол та методи захисту, які можуть бути використані для запобігання цим атакам.

Отримані результати аналізу літератури та визначення основних термінів та понять становлять основу для подальших досліджень у нашій роботі з вивчення BGP протоколу та його захисту. Детальний огляд структури та функціонування BGP допоможе нам краще зрозуміти принципи роботи цього протоколу.

Оскільки BGP протокол важливий, стан його функціонування свідчить про стан інфраструктури загалом. Однак через розподілену та динамічну природу BGP у даних BGP можуть виникати аномалії або неправильності, що може вплинути на стабільність, безпеку та продуктивність Інтернету. Таким чином, виявлення аномалій у даних BGP включає різні методи, включаючи статистичний аналіз, машинне навчання, перевірку та фільтрацію маршрутів, аналіз на основі поведінки та спільні підходи. Ці методи можуть допомогти виявити потенційні аномалії в даних BGP і допомогти у виявленні та пом'якшенні проблем, пов'язаних з BGP, забезпечуючи стабільність і безпеку глобальної інфраструктури маршрутизації Інтернету. Для цього важливо розглянути і застосовувати методи пошуку аномалій.

## 2 АНАЛІЗ МЕТОДІВ ПОШУКУ АНОМАЛІЙ

Розглянемо завдання виявлення змін в трафіку як завдання пошуку аномалій у наборі даних. Проаналізуємо такі методи, як відстань Кука, інтерквартильний діапазон (IQR), ізольований ліс, детермінант мінімальної коваріації (MCD) та парето. Далі виділимо набір критеріїв, на основі яких виберемо метод для даного дослідження. Важливо зазначити, що при аналізі BGP даних за допомогою методів пошуку аномалій необхідно шукати аномалії в часовому ряді, тому були використані відповідні методи для аналізу.

### 2.1 Відстань Кука

Відстань Кука — це вимірювання впливу. Він визначає ступінь впливу певного спостереження на моделі. Коли справа доходить до аномалій, дуже впливовим є спостереження, випадок або новизна. Правильне використання відстані Кука допоможе точно побачити викиди, які знаходяться в наборі даних. Видалення цих викидів може допомогти вам створити кращу базову лінію того, що є нормальним для набору даних і полегшує виявлення аномалій. Якщо прогнози однакові з відповідним спостереженням або без нього, то спостереження не впливає на модель регресії. Якщо прогнози сильно відрізняються, коли спостереження не включено в аналіз, то спостереження є впливовим. Відстань Кука вимірює ефект видалення певного спостереження. Точки даних із великими залишками та високим кредитним плечем можуть спотворити результат і точність регресії. Відстань Кука — це масштабована зміна підігнаних значень. Кожен результуючий елемент у діагностичному обчисленні є нормалізованою зміною вектора коефіцієнтів через видалення спостереження. У практичному звичайному аналізі найменших квадратів відстань Кука можна використовувати кількома способами: для вказівки точок даних, які особливо варто перевірити на

достовірність; щоб вказати регіони простору проектування, де було б добре мати можливість отримати більше точок даних.

Точки з великою відстанню Кука вважаються такими, що заслуговують на більш детальне вивчення під час аналізу. Впливові випадки зазвичай не є проблемою, якщо їх видалення з набору даних залишить оцінки параметрів по суті незмінними [12].

$$D_i = \frac{\sum_{j=1}^n (\hat{Y}_j - \hat{Y}_{j(i)})^2}{p \text{ MSE}},$$

Рисунок 2.1 – Математичний вираз для розрахунку відстані Кука

У виразі  $\hat{Y}_j$  є прогнозом повної регресійної моделі для спостереження  $j$ ,  $\hat{Y}_{j(i)}$  це передбачення для спостереження  $j$  від переобладнаної регресійної моделі, в якій спостереження  $i$  було пропущено,  $p$  - кількість підігнаних параметрів у моделі,  $\text{MSE}$  це середня квадратична помилка регресійної моделі.

Для випадку простої лінійної регресії наступні алгебраїчно еквівалентні вирази:

$$D_i = \frac{e_i^2}{p \text{ MSE}} \left[ \frac{h_{ii}}{(1 - h_{ii})^2} \right],$$

$$D_i = \frac{(\hat{\beta} - \hat{\beta}^{(-i)})^T (X^T X) (\hat{\beta} - \hat{\beta}^{(-i)})}{(1 + p)s^2},$$

Рисунок 2.2 – Математичні вирази для розрахунку відстані Кука

У виразах  $h_{ii}$  це важіль, тобто  $i$ -й діагональний елемент матриці  $X(X^T X)^{-1} X^T$ ,  $e_i$  є залишком (тобто різницею між спостережуваним значенням і значенням, яке відповідає запропонованій моделі).

## 2.2 Інтерквартильний діапазон

Це враховує розміщення точок даних у межах першого і третього квартилей нормальними. Усі точки даних 1,5 або більше разів поза цими квартилями враховується викидами. Ця методика зазвичай відображається графічно за допомогою коробчатої ділянки.

Хоча існує лише одна формула, існують різні методи визначення квартилів. Ви отримаєте різне значення інтерквартильного діапазону залежно від використовуваного методу. Тут ми обговоримо два найпоширеніші методи. Ці методи відрізняються залежно від того, як вони використовують медіану. Ексклюзивний метод виключає медіану при визначенні  $Q1$  і  $Q3$ , тоді як інклюзивний метод включає медіану при ідентифікації квартилів.

Процедура визначення медіани відрізняється залежно від того, парний чи непарний ваш набір даних. Якщо у вас непарна кількість точок даних, медіана – це значення в середині вашого набору даних. Ви можете вибрати між інклюзивним і ексклюзивним методом. При парній кількості точок даних посередині є два значення, тому медіана є їхнім середнім. У цьому випадку частіше використовують ексклюзивний метод.

Хоча немає єдиної думки щодо найкращого методу визначення інтерквартильного діапазону, ексклюзивний інтерквартильний діапазон завжди більший, ніж інклюзивний інтерквартильний діапазон.

Ексклюзивний інтерквартильний діапазон може бути більш відповідним для великих вибірок, тоді як для невеликих вибірок інклюзивний інтерквартильний діапазон може бути більш репрезентативним, оскільки це вужчий діапазон.

Формула для визначення інтерквартильного діапазону бере значення третього квартиля та віднімає значення першого квартиля  $IQR = Q3 - Q1$ . Так само інтерквартильний діапазон – це область між 75-м і 25-м перцентилем ( $75 - 25 = 50\%$  даних).

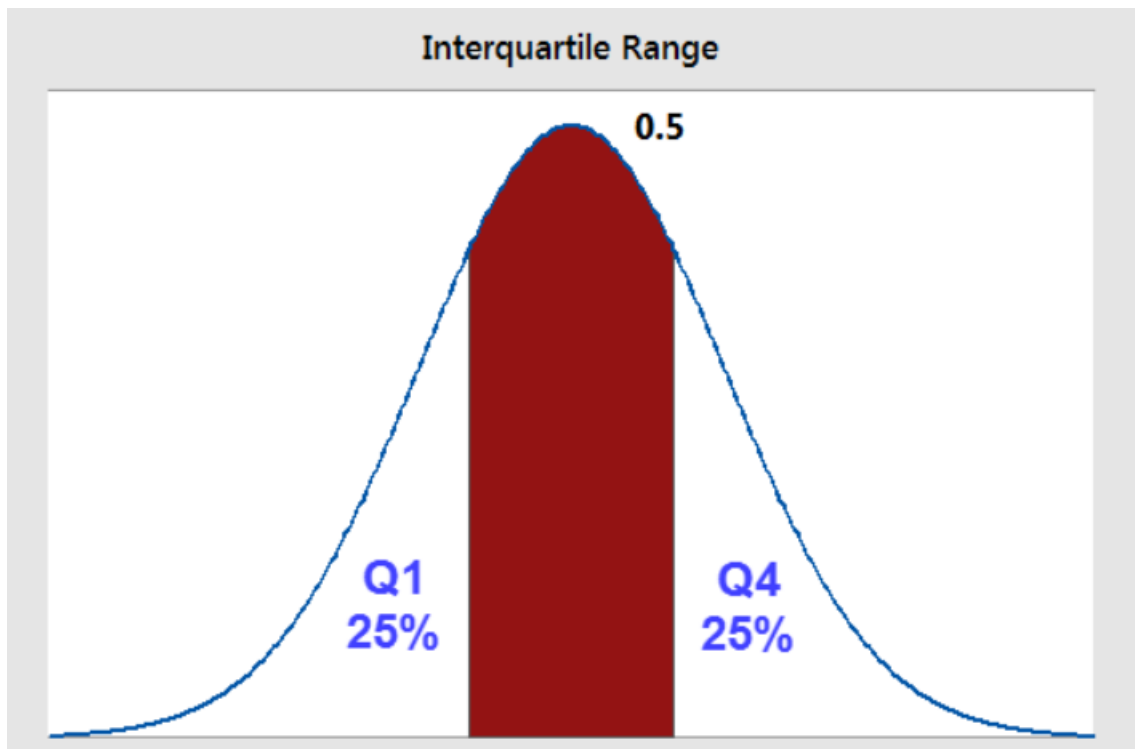


Рисунок 2.3 – Пошук інтерквартильного діапазону

Використовуючи формулу IQR, нам потрібно знайти значення Q3 і Q1. Для цього просто впорядкуйте дані від низького до високого та розділіть значення на чотири рівні частини. Для прикладу ми обрали набір даних нижче і розділили на квартилі. Інтерквартильний діапазон поширюється від значення Q1 до значення Q3. Для цього набору даних інтерквартильний діапазон становить  $39 - 20 = 19$ .

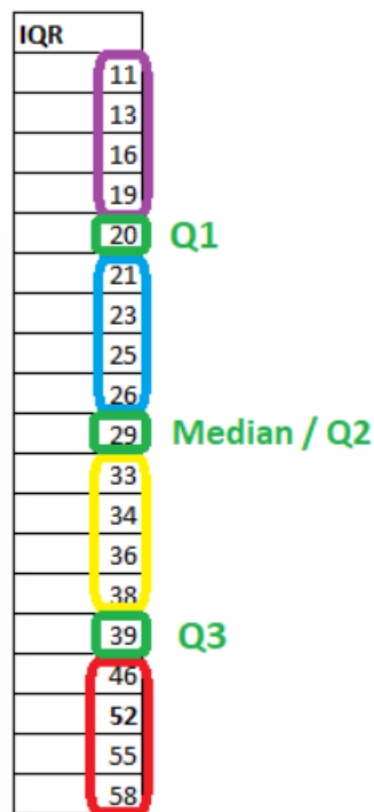


Рисунок 2.4 – Набір даних для пошуку інтерквартильного діапазону

Зауважте, що різні методи та статистичні програми знайдуть дещо різні значення Q1 і Q3, що впливає на інтерквартильний діапазон. Ці варіації походять від альтернативних способів знаходження процентилів [13].

### 2.3 Ізольований ліс

Цей метод явно визначає аномалії замість профілювання нормальних точок даних. Ізольований ліс, як і будь-який метод ансамблю дерев, будується на основі дерев рішень. У цих деревах розділи створюються шляхом спочатку випадкового вибору функції, а потім вибору випадкового значення розділення між мінімальним і максимальним значенням вибраної функції.

В принципі, викиди зустрічаються рідше, ніж регулярні спостереження, і відрізняються від них за значеннями (вони лежать далі від регулярних спостережень у просторі ознак). Ось чому за допомогою такого випадкового поділу їх слід ідентифікувати ближче до кореня дерева (менша середня

довжина шляху, тобто кількість ребер, які має пройти спостереження в дереві, що йде від кореня до кінцевого вузла), з меншою кількістю необхідні розколи.

Ідею ідентифікації нормального та ненормального спостереження можна побачити на рисунку 2.5. Нормальна точка (ліворуч) потребує ідентифікації більше розділів, ніж аномальна точка (праворуч) [14].

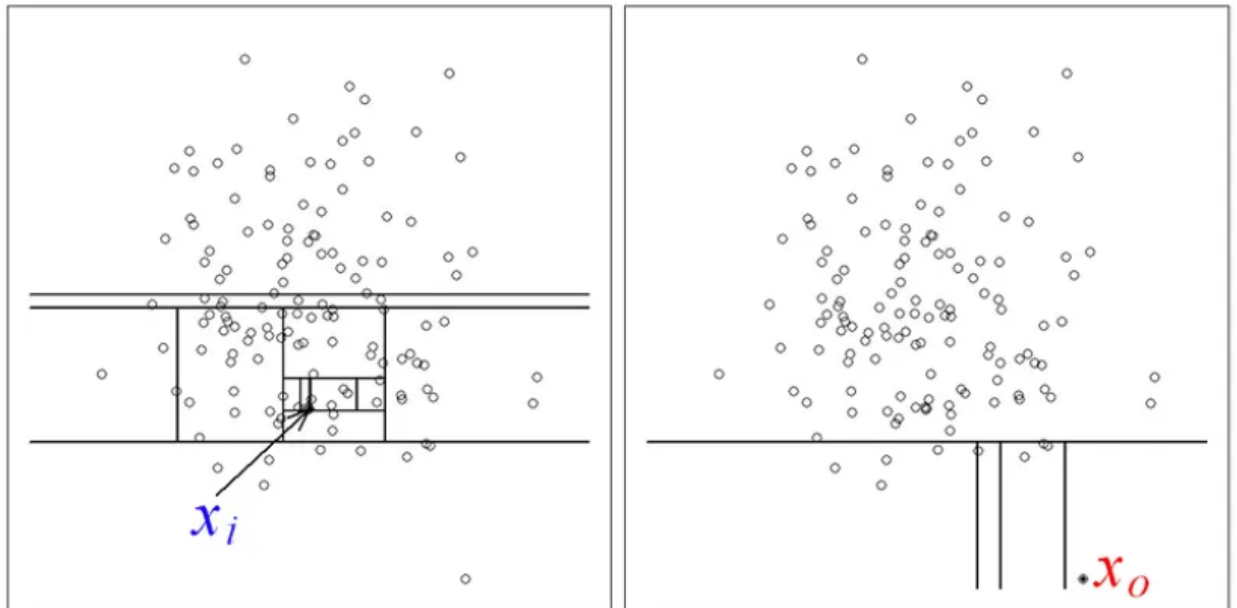


Рисунок 2.5 – Нормальні і аномальні точки в датасеті

Як і для інших методів виявлення викидів, оцінка аномалії потрібна для прийняття рішення. У випадку ізолюваного лісу це визначається за наступною формулою:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}}$$

Рисунок 2.6 – Формула оцінки аномалії в методі ізолюваного лісу

У виразі  $h(x)$  це довжина шляху спостереження  $x$ ,  $c(n)$  — середня довжина шляху невдалого пошуку в бінарному дереві пошуку, а  $n$  — кількість зовнішніх вузлів.

Кожному спостереженню присвоюється бал аномалії, і на його основі можна прийняти таке рішення:

- Оцінка, близька до 1, свідчить про аномалії.
- Оцінка набагато менша за 0,5 свідчить про нормальні спостереження.
- Якщо всі бали близькі до 0,5, то вся вибірка не має чітко виражених аномалій.

## 2.4 Детермінант мінімальної коваріації

Метод мінімального коваріаційного визначника (MCD) — це надійний засіб оцінки багатовимірного розташування та розкиду, для якого доступний швидкий алгоритм. Оскільки оцінка коваріаційної матриці є наріжним каменем багатьох багатовимірних статистичних методів, MCD є важливим будівельним блоком при розробці надійних багатовимірних методів. Він також служить зручним і ефективним інструментом для виявлення викидів. Розглянуто оцінювач MCD, а також його основні властивості, такі як афінна еквіваріантність, значення розбивки та функція впливу. Ми обговорюємо його обчислення та перераховуємо застосування та розширення MCD у прикладній та методологічній багатовимірній статистиці. Описано два останніх розширення MCD. Перший — це швидкий детермінований алгоритм, який успадковує надійність MCD, будучи майже афінним еквіваріантним. Другий адаптований до даних великої розмірності, можливо, з більшою кількістю вимірів, ніж випадків, і включає регуляризацію для запобігання сингулярним матрицям.

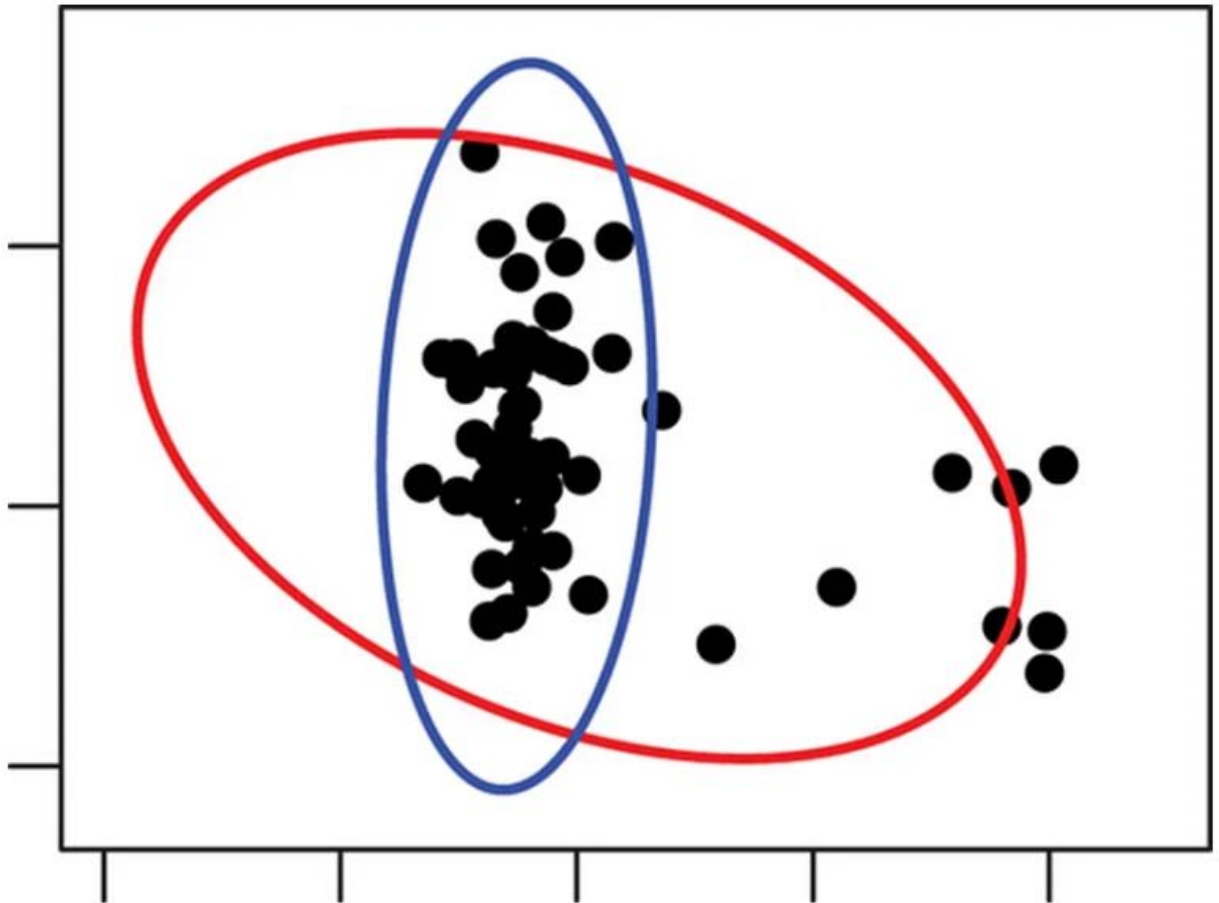


Рисунок 2.7 – Візуальне представлення MCD методу

MCD — це метод оцінки середнього значення та коваріаційної матриці таким чином, щоб мінімізувати вплив аномалій. Ідея полягає в тому, щоб оцінити ці параметри на основі підмножини даних, які були вибрані як вільні від аномалій.

Для цього береться набір підвбірок даних заданого розміру та з них визначається середнє значення та коваріаційна матриця. Потім оцінки зберігаються для підмножини, коваріаційна матриця якої має найменший визначник.

Ідея мінімізації детермінанта полягає в тому, що детермінант коваріаційної матриці вимірює, наскільки широким є розподіл. Тому MCD вибирає найбільш щільно розподілену підмножину даних. Це робиться для того, щоб виключити аномалії, які, ймовірно, лежать далі від решти даних [15].

## 2.5 Парето

PDA використовує концепцію оптимальності за Парето для виявлення аномалій без необхідності вибору вагових коефіцієнтів для різних критеріїв. Оптимальність за Парето є типовим методом визначення оптимальності, коли може існувати кілька суперечливих критеріїв для порівняння елементів. Кажуть, що предмет може бути оптимальним за Парето, якщо не існує іншого елемента, кращого або рівного за всіма критеріями. Елемент, який є оптимальним за Парето, є оптимальним у звичайному розумінні за певної комбінації, необов'язково лінійний, критеріїв.

Підхід PDA передбачає створення діад, що відповідають відмінностям між парами даних зразків за всіма показниками відмінності. Набори парето-оптимальних діад, які називаються фронтами Парето, потім обчислюються. Перший фронт Парето (глибина один) - це набір недомінованих діад. Другий фронт Парето (друга глибина) виходить шляхом видалення цих недомінованих діад, тобто відшаровування першого фронту і переобчислення першого фронту Парето з тих, що залишилися. Цей процес триває до тих пір, поки не залишиться діад. Таким чином, кожна діада призначається фронту Парето на деякій глибині. Номінальні та аномальні проби розташовані поблизу різних глибин фронту Парето; таким чином обчислення фронтальних глибин діад, що відповідають тестовому зразку, може розрізняти між ними номінальні та аномальні проби.

Розглянемо таку задачу: задано  $n$  елементів, позначених множиною  $S$ , і  $K$  критеріїв оцінки кожного елемента, позначених функціями  $f_1, \dots, f_K$ , виберіть  $x \in S$ , що мінімізує  $[f_1(x), \dots, f_K(x)]$ . У більшості налаштувань неможливо визначити один елемент  $x$ , який одночасно мінімізує  $f_i(x)$  для всіх  $i \in \{1, \dots, K\}$ . Мінімізатор можна знайти, об'єднавши критерії  $K$  за допомогою лінійної комбінації  $f_i$  і знайшовши мінімум комбінації. Різні вибори (невід'ємних) ваг у лінійній комбінації можуть призвести до різних мінімізаторів; набір елементів, які є мінімізаторами під деякою лінійною

комбінацією, можна створити, наприклад, за допомогою пошуку в сітці за вагами.

Більш потужний підхід передбачає знаходження набору оптимальних за Парето елементів. Кажуть, що елемент  $x$  суворо домінує над іншим елементом  $x^*$ , якщо  $x$  не перевищує  $x^*$  за кожним критерієм і  $x$  є меншим за  $x^*$  принаймні за одним критерієм. Це співвідношення можна записати як  $x \succ x^*$  якщо  $f_i(x) \leq f_i(x^*)$  для кожного  $i$  та  $f_i(x) < f_i(x^*)$  для деяких  $i$ . Набір оптимальних за Парето елементів, званий фронтом Парето, — це набір елементів у  $S$ , над якими не домінує строго інший елемент у  $S$ . Він містить усі мінімізатори, знайдені за допомогою лінійних комбінацій, але також включає інші елементи, які не можна знайти шляхом лінійних комбінацій. Позначимо  $F1$  фронт Парето, який ми називаємо першим фронтом Парето. Другий фронт Парето можна побудувати шляхом знаходження елементів, які не є строго домінованими жодним із елементів, що залишилися, які є членами множини  $S \setminus F1$ . Більш загально визначення  $i$ -й фронт Парето виглядає  $s \setminus \left( \bigcup_{j=1}^{i-1} F_j \right)$  [16].

## 2.6 Порівняння методів

Для вибору потрібного методу пошуку аномалій необхідно порівняти методи між собою, спочатку розглянувши методології пошуку, плюси та мінуси таких підходів.

Таблиця 2.1 – Порівняння методів пошуку аномалій за методом пошуку

| Метод         | Методологія пошуку                                    | Плюси                                                                         | Мінуси                                    |
|---------------|-------------------------------------------------------|-------------------------------------------------------------------------------|-------------------------------------------|
| Відстань Кука | Виявляє впливові спостереження на основі їх впливу на | - Спеціально розроблено для визначення впливових точок у регресійному аналізі | - Припускає лінійний зв'язок між змінними |

Продовження таблиці 2.1

|                 |                                                                        |                                                                                                                                                                                   |                                                                                                                                                                                                                                                           |
|-----------------|------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                 | коефіцієнти регресії                                                   | <ul style="list-style-type: none"> <li>- Враховує як величину, так і напрямок ефекту</li> <li>- Добре працює з моделями лінійної регресії</li> </ul>                              | <ul style="list-style-type: none"> <li>- Чутливий до припущень моделі та викидів</li> <li>- Може не працювати добре для нелінійних зв'язків</li> </ul>                                                                                                    |
| IQR             | Використовує різницю між 75-м і 25-м процентилям для виявлення викидів | <ul style="list-style-type: none"> <li>- Простий і зрозумілий</li> <li>- Надійність до екстремальних значень у наборі даних</li> <li>- Добре працює з нерівними даними</li> </ul> | <ul style="list-style-type: none"> <li>- Враховує лише поширення даних у межах інтерквартильного діапазону</li> <li>- Ігнорує інші статистичні показники</li> <li>- Може не працювати належним чином для даних із кількома піками або режимами</li> </ul> |
| Ізольований ліс | Будує ансамбль дерев рішень для ізоляції аномалій                      | <ul style="list-style-type: none"> <li>- Здатність обробляти дані великого розміру</li> <li>- Швидкий і масштабований для великих наборів даних</li> </ul>                        | <ul style="list-style-type: none"> <li>- Може погано працювати з наборами даних із даними низької розмірності або даними зі складними</li> </ul>                                                                                                          |

Кінець таблиці 2.1

|        |                                                                                           |                                                                                                                                                                         |                                                                                                                                                                   |
|--------|-------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|        |                                                                                           | - Не вимагає припущень щодо розподілу даних                                                                                                                             | зв'язками<br>- Може бути чутливим до налаштування гіперпараметрів                                                                                                 |
| MCD    | Використовує надійну оцінку середнього та коваріації для виявлення багатовимірних викидів | - Надійний до викидів і стійкий до високих точок кредитного плеча<br>- Добре працює з даними великого розміру<br>- Може обробляти набори даних із нелінійними зв'язками | - Припускає, що більшість даних генерується з багатовимірного нормального розподілу<br>- Може не працювати належним чином для даних із важкими хвостами розподілу |
| Парето | Ранжує точки даних на основі їх внеску в кумулятивний відсоток від загальної кількості    | - Простий і зрозумілий<br>- Надає візуальне представлення точок даних, які найбільше вносять вклад у загальну суму                                                      | - Визначає лише найбільших учасників<br>- Ігнорує інші статистичні показники<br>- Може не працювати добре для наборів даних із великою кількістю учасників        |

Також ми зібрали список критеріїв для порівняння перелічених методів пошуку аномалій. Критерії використовуються для оцінки продуктивності та ефективності різних методів виявлення аномалій у наборах даних. Ці критерії служать вказівками або еталонними показниками для оцінки придатності та надійності методів виявлення аномалій для конкретних програм або наборів даних.

Критерії зазвичай використовуються для порівняння та порівняння різних методів виявлення аномалій, а також для прийняття обґрунтованих рішень про те, який метод може найкраще підходити для конкретного випадку використання. Вони допомагають об'єктивно оцінити сильні та слабкі сторони різних методів і вибрати найбільш відповідний метод на основі конкретних вимог або обмежень.

Таблиця 2.2– Порівняння методів за набором критеріїв

| Критерій                            | Відстань Кука | IQ R | Ізольовані й ліс | MCD                         | Парето |
|-------------------------------------|---------------|------|------------------|-----------------------------|--------|
| Виявляє викиди                      | Так           | Так  | Так              | Так                         | Так    |
| Багатоваріантний                    | Ні            | Ні   | Так              | Так                         | Ні     |
| Однофакторний                       | Так           | Так  | Ні               | Ні                          | Так    |
| Стійкий до шумних даних або викидів | Ні            | Так  | Так              | Так                         | Ні     |
| Обробляє багатовимірні дані         | Ні            | Ні   | Так              | Так                         | Ні     |
| Припускає розподіл даних            | Лінійний      | N/A  | N/A              | Багатоваріантний нормальний | N/A    |

Кінець таблиці 2.2

|                                      |     |    |     |     |     |
|--------------------------------------|-----|----|-----|-----|-----|
| Чутливість до припущень моделі       | Так | Ні | Ні  | Ні  | Ні  |
| Вимагає налаштування гіперпараметрів | Ні  | Ні | Так | Так | Ні  |
| Забезпечує візуальне представлення   | Ні  | Ні | Ні  | Ні  | Так |

Підсумовуючи, відстань Кука є переконливим вибором для виявлення аномалії на основі критеріїв, представлених у таблиці. Його чутливість до впливових моментів, стійкість до викидів, можливість інтерпретації, знайомство з регресійним аналізом, встановлені властивості та доступність реалізації в популярних пакетах статистичного програмного забезпечення роблять його сприятливим варіантом для виявлення аномалій у наборах даних.

Відстань Кука особливо корисна при роботі з наборами даних, які можуть містити впливові точки даних або дані з шумом, оскільки вона забезпечує кількісну міру впливу кожної точки даних на оцінки параметрів моделі. Його стійкість до викидів гарантує, що він може точно ідентифікувати аномалії навіть за наявності незвичайних або схильних до викидів точок даних.

Крім того, відстань Кука є добре відомим методом із довгою історією використання в регресійному аналізі, що робить його знайомим для практиків і легкодоступним у багатьох статистичних програмних пакетах. Його можливість інтерпретації дозволяє легко зрозуміти та повідомити результати зацікавленим сторонам або членам команди.

Загалом відстань Кука є надійним і ефективним методом виявлення аномалій у наборах даних і може бути цінним інструментом у різних програмах, де виявлення впливових точок або аномалій представляє інтерес.

## Висновки до розділу 2

Проаналізовано ряд методів, таких як відстань Кука, інтерквартильний діапазон (IQR), ізольований ліс, детермінант мінімальної коваріації (MCD) та парето.

Метод відстані Кука використовується для виявлення великих аномалій, які можуть впливати на регресійну модель. Інтерквартильний діапазон (IQR) дозволяє виявляти аномалії на основі міжквартильного розмаху. Ізольований ліс - це метод, який базується на використанні випадкових лісів для виділення аномалій. Детермінант мінімальної коваріації (MCD) використовується для виявлення аномалій в багатовимірних даних, враховуючи коваріаційну структуру даних. Парето - це метод, який дозволяє визначити найважливіші аномалії на основі принципу Парето.

Кожен з цих методів має свої переваги та обмеження і може бути застосований в різних випадках виявлення аномалій в трафіку. Вибір оптимального методу залежить від специфіки даних та поставленої задачі. Дослідження різних методів дозволило нам зробити аналіз їх ефективності та можливостей застосування в контексті виявлення змін в трафіку. Спираючись на задачі дослідження та датасет, в даній роботі буде використано метод відстані Кука.

### 3 РОЗРОБКА МЕТОДИКИ ВИЯВЛЕННЯ ЗМІН У ТРАФІКУ BGP

В цьому розділі представлено методику виявлення змін у трафіку BGP. Ми перевіримо гіпотезу, що зміни мережевого трафіку пов'язані з руйнуванням інфраструктури внаслідок війни. Ми розглянемо та реалізуємо методику виявлення змін у трафіку. В ході роботи ми опрацюємо BGP повідомлення та встановимо критерії для підрахунку кількості збоїв мережі. Крім цього, буде використано метод пошуку аномалій відстань Кука на даних мережевого трафіку за час війни та проаналізовано результати.

#### 3.1 Основні складові методики виявлення змін у трафіку

У роботі пропонується власна методика виявлення змін у трафіку BGP з використанням методу пошуку аномалій відстань Кука, яка реалізована у наборі скриптів на мові програмування Python.

Опис кроків методики виявлення змін у трафіку BGP:

1. Збір та обробка даних BGP: В першому етапі здійснюється збір даних BGP з джерела [17], які стосуються мережевого трафіку. Ці дані підлягають обробці, а саме розпаковці, щоб перетворити в читабельний формат.
2. Вибір критерію для збою мережі: У другому етапі вибирається критерій, який відповідає збою мережі, який планується виявити.
3. Відбір BGP повідомлень згідно критерію з використанням алгоритму: Далі, на основі вибраного критерію, відбираються BGP повідомлення, які відповідають збою мережі. Це етап включає фільтрацію повідомлень за певними атрибутами та порівняння поточних значень зі значеннями в минулому.
4. Перетворення проміжних результатів в датасет та використання методу пошуку аномалій відстань Кука: Після відбору BGP повідомлень, проміжні результати перетворюються в датасет, який може використовуватися в методі пошуку аномалій. Метод відстані Кука

використовується для виділення аномальних подій в цьому датасеті.

Цей метод дозволяє виявити відхилення від типових значень і визначити події, які були аномальними в наборі даних.

Перед тим, як перейдемо до алгоритму відбору BGP повідомлень, визначимо критерій для даного дослідження.

Критерій. В даній роботі збоєм вважається тривалий період, протягом якого відмова від трансляції BGP не супроводжується оголошенням для даного префікса IP; на практиці це означає, що префікс недоступний протягом цього періоду. В даному аналізі я беру до уваги лише відключення, які тривають більше 5 хвилин. Збої, які тривають менше цього порогового значення, вважаються тимчасовими, випадковими збоями та ігноруються.

Алгоритм для відбору BGP повідомлень порівнює BGP UPDATE повідомлення між собою, щоб знайти збої в мережі, які задовольняють критерій. Алгоритм складається з наступних етапів:

1. Зчитування BGP повідомлення.
2. Перевірка приналежності мережі до українського сегменту.
3. Базуючись на типі повідомлення, існує 2 сценарії:
  - a. Якщо повідомлення типу A і префікс мережі не знаходиться в базі даних, то даний префікс мережі зберігається з часом створення повідомлення.
  - b. Якщо повідомлення типу W і префікс мережі знаходиться в базі даних, то ми витягуємо з бази даних час створення попереднього повідомлення і перевіряємо, чи час збою перевищує вказаний критерій. Якщо так, то об'єкт додається в окремо базу зі збоями.

Наприкінці, префікс мережі видаляється з бази даних.

Візуально алгоритм виглядає наступним чином:

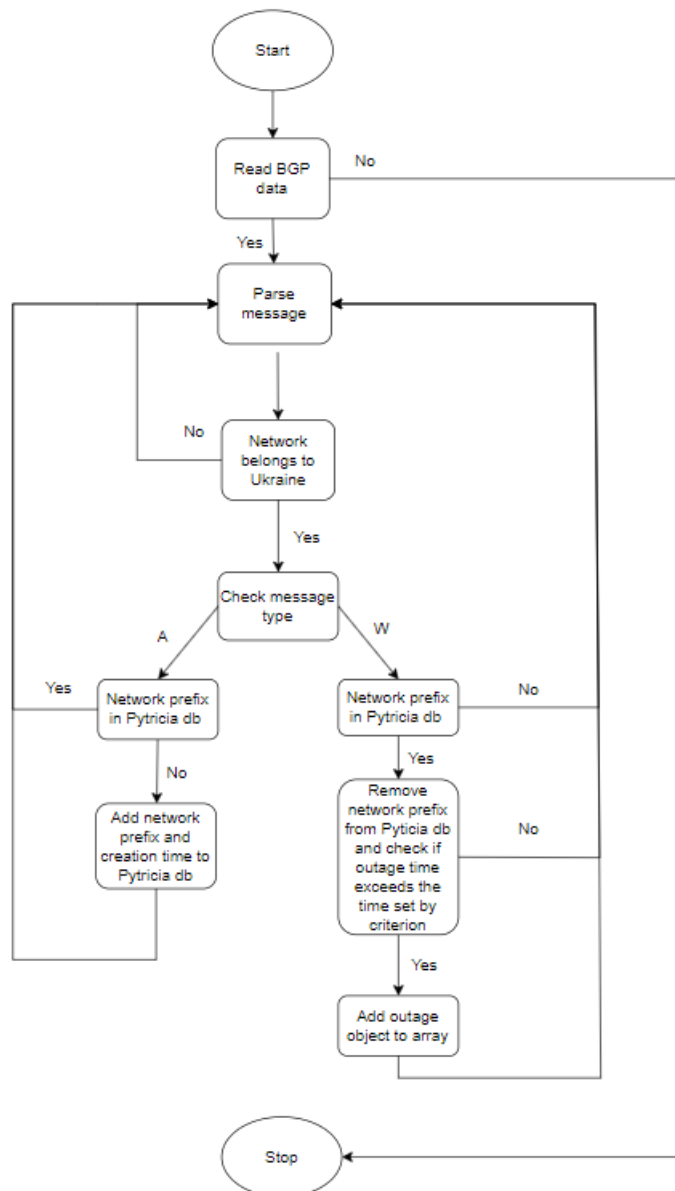


Рисунок 3.1 – Блок-схема

Останнім елементом методики є використання відстані Кука. Для проведення перевірки гіпотези та виявлення аномалії був застосований метод відстані Кука на попередньо отриманих даних. Задача його використання це виявити спостереження, які мають надмірний вплив на результуючу модель, а також визначити потенційні аномалії, які можуть впливати на вірогідність висновків, зроблених на основі досліджуваних даних.

Перед використанням метода, треба сформувати датасет з BGP даних у форматі дата та кількість збоїв. Для цього ми беремо попередньо отриманий

масив даних та підраховуємо кількість збоїв за кожен день. Обробивши весь масив, ми отримаємо датасет, на якому можемо використати відстань Кука.

### 3.2 Практична реалізація методики

Першим кроком розробленої методики є збір та обробка BGP даних. Повідомлення BGP були взяті з 24 лютого 2022 року по 30 квітня 2022 року та були оброблені з використанням `bgpdump` утиліти [18].

```
url = 'http://archive.routeviews.org/bgpdata/%s/UPDATES/updates.%s.%s.bz2' % (ymdot, ymd, hm)
print(url)
subprocess.run(["wget", "-r", "-nc", "-P", dir, url])
```

Рисунок 3.2 – Завантаження даних BGP

```
file = '/Users/viktoriia/Downloads/Diploma/Practice/archive.routeviews.org/bgpdata/%s/UPDATES/updates.%s.%s.bz2' % (ymdot, ymd, hm)
file2 = '/Users/viktoriia/Downloads/Diploma/Practice/files/%s.%s' % (ymd, hm)
print(file)
if os.path.exists(file) and os.path.isfile(file):
    print(file)
    print('\nFile: ', file)
    convert = 'bgpdump -m %s > %s' % (file, file2)
    os.system(convert)
```

Рисунок 3.3 – Обробка даних BGP з використанням `bgpdump`

Далі було використано python скрипт, який зчитував кожне повідомлення, відбирав українські мережі і фіксував збій.

```
def extract_bgp_message(message):
    # Extract the fields from the message using the regular expression pattern
    fields = message.strip().split("|")
    type = fields[2]
    prefix = ipaddress.ip_network(fields[5]).exploded
    timestamp = int(fields[1].split('.')[0])
    return type, prefix, timestamp
```

Рисунок 3.4 – Обробка повідомлення BGP

```
def is_ukrainian_prefix(prefix):
    try:
        city_rec = geo_reader.city(prefix.split('/')[0])
        country_name = city_rec.country.name
    except:
        return ''
    return country_name
```

Рисунок 3.5 – Визначення приналежності мережі до України

```

if type == "A" and not pt.has_key(prefix):
    pt.insert(prefix,timestamp)
elif type == "W" and pt.has_key(prefix):
    duration = timestamp - pt[prefix]
    if duration > outage_time:
        # Create a dictionary object to store the outage information
        outage_object = {
            'prefix': prefix,
            'start_time': pt[prefix],
            'end_time': timestamp
        }
        db.append(outage_object)
    #Remove the prefix from the PyTricia object
    pt.delete(prefix)

```

Рисунок 3.6 – Фіксування збою

Для наступного етапу методики було отримано масив об'єктів з префіксом мережі, початком і кінцем збою. Проміжні дані розробленої методики представлені далі у вигляді графіків.

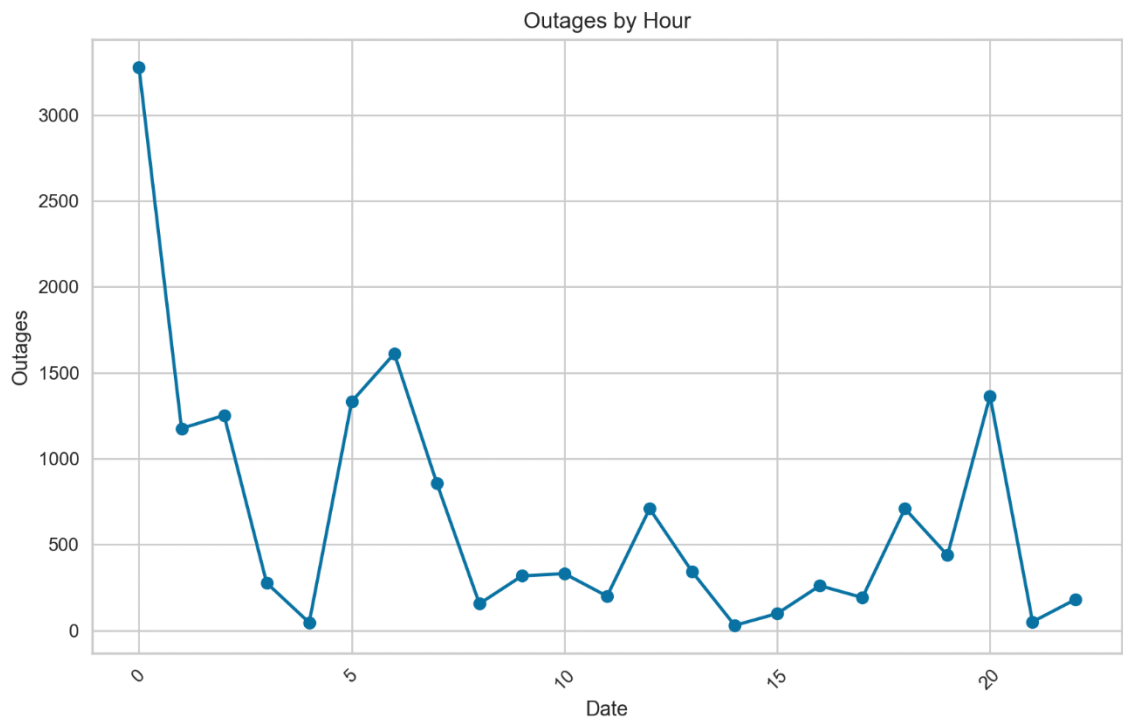


Рисунок 3.7 – Графік збоїв мережі за 24 лютого 2022 року погодинно

За першу годину доби виявлено було понад 3000 збоїв в протоколі маршрутизації BGP, який використовується для обміну мережевою інформацією між автономними системами в Інтернеті. Ці збої можуть бути пов'язані з різними аспектами функціонування BGP, такими як неправильні налаштування, неправильна передача мережевих маршрутів, атаки з метою перекривлення мережевого трафіку тощо.

Згідно з цим графіком, ми можемо спостерігати пікову кількість збоїв на початку дня перед повномасштабним вторгненням, що може вказувати на підготовчі дії до кібератак.

Середня кількість збоїв BGP за годину становить приблизно 1000. Це може вказувати на високий рівень нестабільності в роботі BGP в мережі, що може мати наслідки для передачі даних в Інтернеті та може вимагати швидкої реакції операторів мережі для виправлення проблем. Збої BGP можуть викликати відмови у мережевому з'єднанні, неправильна маршрутизація мережевого трафіку, зниження продуктивності мережі та впливати на доступність різних ресурсів в Інтернеті.

У ситуаціях з воєнним конфліктом можуть відбуватися зміни в мережевій інфраструктурі, зокрема в областях, які стали зонами бойових дій.

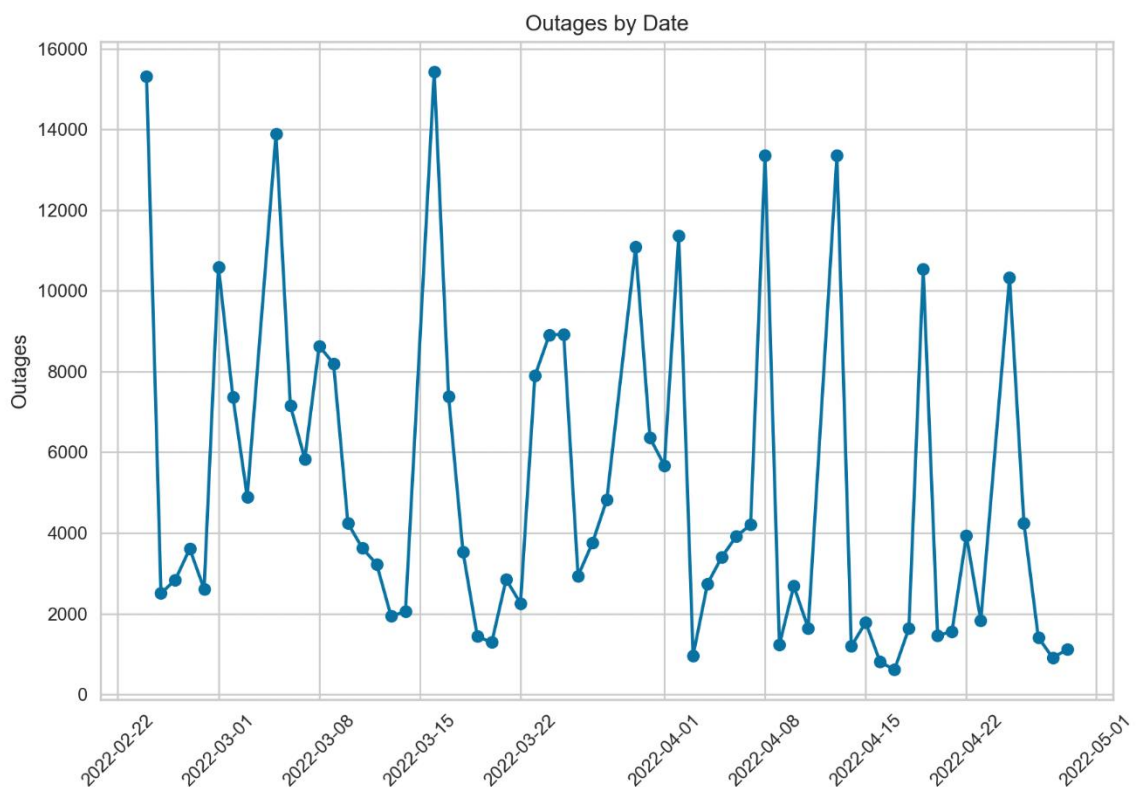


Рисунок 3.8 – Графік збоїв мережі з 24 лютого 2022 року по 30 квітня 2022 року

Протягом визначеного періоду, середня кількість збоїв BGP досягає більше 8000 одиниць за день. Це вказує на високий рівень нестабільності мережевої інфраструктури, пов'язаної з воєнним конфліктом.

Зокрема, виділяється п'ять днів, коли кількість збоїв перевищує 14000 одиниць і більше. Це ще більш серйозний рівень збоїв, який може вказувати на значні труднощі в мережевій інфраструктурі, включаючи можливі відключення мережевих вузлів, блокування передачі мережевих маршрутів, кібератаки та інші проблеми.

Цей рівень збоїв має серйозний вплив на доступність та функціонування мереж та інтернет-послуг, зокрема на роботу підприємств, установ, організацій та громадян, які використовують мережевий зв'язок для своєї діяльності. Враховуючи велику кількість збоїв, особливо на підвищеному рівні в певні дні, відновлення мережевого зв'язку може бути складним і вимагати відповідних заходів безпеки та ретельного моніторингу для вирішення проблем та відновлення нормальної роботи мережі [25].

Наступним етапом розробленої методики є формування датасету для методу пошуку аномалій з попереднього масиву даних:

```
def parse_db():
    # Specify the start and end times of the time range
    data_list = []
    files = sorted([os.path.join(files_dir, f) for f in os.listdir(files_dir) if os.path.isfile(os.path.join(files_dir, f))])
    for file in files:
        print(file)
        db = json.load(open(file))
        db = sorted(db, key=lambda d: d["start_time"])
        #
        ## Count the number of outages that occurred during the time range
        temp_date = datetime.fromtimestamp(db[0]["start_time"], pytz.timezone("UTC")).day
        num_outages = 0
        day_month_year_format = '%Y-%m-%d'
        for outage in db:
            start = datetime.fromtimestamp(outage["start_time"], pytz.timezone("UTC")).day
            if start != temp_date:
                date_insert = datetime.fromtimestamp(outage["start_time"], pytz.timezone("UTC")) - timedelta(days=1)
                data_dict = {
                    "date": date_insert.strftime(day_month_year_format),
                    "outages": num_outages
                }
                data_list.append(data_dict)
                temp_date = start
                num_outages = 0
                num_outages += 1
        with open("data1.json", "w") as f:
            json.dump(data_list, f)
```

Рисунок 3.9 – Формування датасету

Для того щоб використати метод відстані Кука, ми маємо встановити пакет Yellowbrick, який є одним з популярних пакетів візуалізації даних в

мові програмування Python [19]. Для цього можна використовувати менеджер пакетів, такий як `pip`, або інший зручний спосіб встановлення пакетів.

Після встановлення `Yellowbrick` пакету, ми можемо імпортувати його в нашу програму або ноутбук. При цьому ми також маємо імпортувати наш датасет, який ми попередньо сформували. Ми можемо використовувати зручні функції та класи з `Yellowbrick` для візуалізації та аналізу наших даних.

Далі, ми можемо перетворити наш датасет в формат, придатний для використання методу відстані Кука. Це може включати масштабування даних, заміну відсутніх значень, кодування категоріальних ознак, або інші попередні операції, які можуть бути необхідні для підготовки даних перед застосуванням методу.

Після попередньої підготовки даних, ми можемо використовувати метод відстані Кука з `Yellowbrick` для аналізу наших даних в останньому етапі розробленої методики. Це може включати створення графіків, діаграм, діагностичних діаграм, або інших візуалізаційних засобів, які допоможуть нам проаналізувати взаємозв'язки та виявити аномалії в наших даних, використовуючи метод відстані Кука як інструмент аналізу [24].

```
def cook_distance():
    df = pd.read_json('data1.json')

    # Fit a linear regression model
    X = df['outages'].values.reshape(-1, 1)
    y = range(len(X))
    model = LinearRegression().fit(X, y)

    # Visualize Cook's distance for each observation
    visualizer = CooksDistance()
    visualizer.fit(X, y)
    visualizer.show()
```

Рисунок 3.10 – Використання методу відстань Кука

### 3.3 Аналіз результатів

В ході експериментального дослідження було розроблено і перевірено методику виявлення змін у трафіку BGP. Спочатку було зібрано дані з протоколу міждоменного маршрутизатора BGP. Ці дані були піддані обробці для підготовки до аналізу.

Далі, на основі зібраних даних було сформовано датасет, що складався з рядків, які містили інформацію про дати та кількість збоїв в мережі на кожен день або певний проміжок часу.

Далі, був використаний метод пошуку аномалій, зокрема метод відстані Кука, для перевірки гіпотези або виявлення потенційних аномалій в наших даних.

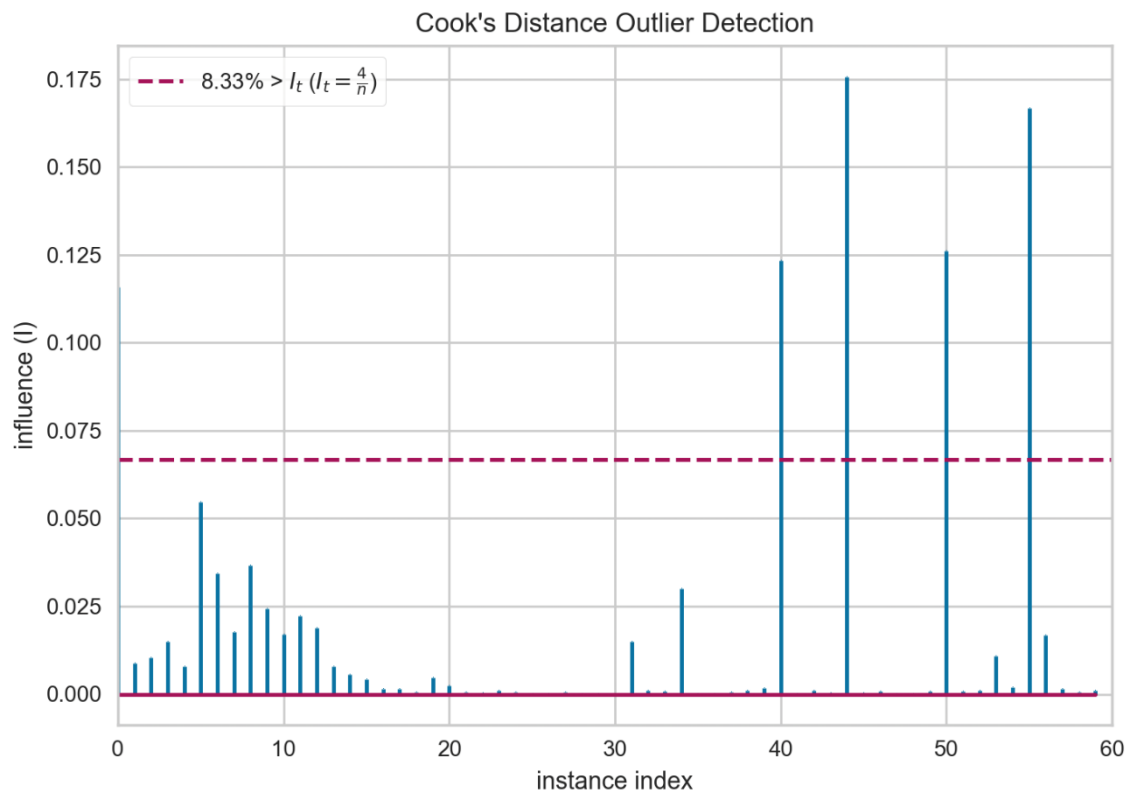


Рисунок 3.11 – Графік аномалій

Згідно графіку було визначено лінію впливу, яка дорівнює 7% та 4 аномалії, які перевищують цю лінію. Визначені аномалії відповідають наступним датам:

- 8 квітня 2022 року
- 13 квітня 2022 року
- 19 квітня 2022 року
- 25 квітня 2022 року

Ми маємо перевірити події, які відбувались в інформаційному просторі або на полі битви в ці дні, щоб підтвердити гіпотезу.

8 квітня 2022 року, внаслідок російської військової агресії проти України, російські військові нанесли ракетний удар по залізничному вокзалу у місті Краматорськ, розташованому на сході Донецької області, що знаходиться на сході України. В результаті цього обстрілу якого загинули 58 цивільних жителів і 100 зазнали поранень.

Зазначений залізничний вокзал у Краматорську був використаний для евакуації цивільного населення з районів, що перебували під вогнем конфлікту. Наслідком ракетного удару було зруйновано частину вокзалу та прилеглої території. Близько чотирьох тисяч цивільних осіб перебували на вокзалі в момент обстрілу [20].

13 квітня 2022 року в оточеному місті Маріуполь на сході України відбулась важлива подія, коли один з підрозділів 36-ї окремої бригади морської піхоти Військово-морських сил ЗСУ імені контрадмірала Михайла Білінського зміг успішно з'єднатися з силами полку "Азов", що було важливим кроком в організації спротиву військовим діям російської агресії. Однак, деяка частина морської піхоти стала жертвами полонення внаслідок жорстоких бойових дій. Доля решти морпіхів залишається нез'ясованою, оскільки частина з них намагалася вибратися з оточеного міста, щоб продовжити опір на фронті [21].

Ця подія свідчить про складність та напруженість воєнного становища в Маріуполі, де збройні сили України вели відчайдушну боротьбу з російською агресією, намагаючись протистояти оточенню та забезпечити виживання своїх військових підрозділів. Це також вказує на тяжкі наслідки війни, включаючи втрати в ув'язнених та загиблих серед військовослужбовців, а також на гуманітарну ситуацію в місті, де цивільне населення сталкивалося з високим ризиком травмування та втратами.

Також цього дня російський ракетний крейсер «Москва», що на початку вторгнення пропонував здатися захисникам острова Зміїний, зазнав обстрілу українськими військовими, які випустили на нього 2 ракети Нептун. Корабель отримав пошкодження та пізніше затонув під час шторму. Згодом

це призвело до зростання частоти ракетних обстрілів української території з боку російських військ, розглядається як помста за затонення крейсера.

19 квітня в місті Маріуполь на сході України сталася жахлива подія, коли російські окупанти вчинили бомбардування лікарні, розташованої в районі заводу "Азовсталь" [22]. Внаслідок цього нападу була скинута потужна бомба, яка призвела до руйнування будівлі, і залишила під завалами приблизно 300 людей, серед яких були і діти.

Ця трагедія свідчить про жорстокість та безжалісність російської агресії, яка спричинила велику кількість цивільних жертв, включаючи невинних дітей, які потрапили під бомбардування. Атаки на медичні заклади, що мають захищений статус відповідно до міжнародного гуманітарного права, є грубим порушенням прав людини та міжнародних норм.

25 квітня сталися жахливі атаки на вокзали та залізничну інфраструктуру в п'яти містах України, зокрема Жмеринці, Козятині та Львові [23]. Ці напади призвели до серйозного пошкодження транспортної системи країни та мали значний вплив на мобільність населення.

Російські окупанти вчинили ці атаки на цивільні об'єкти, такі як вокзали, які мають важливе значення для забезпечення зв'язку та пересування громадян. Ці напади є ще одним жорстоким прикладом безжалісної агресії з боку Росії, спрямованої проти мирних міст та населення України.

Ці атаки на цивільну інфраструктуру, такі як вокзали та залізничні станції, які мають захищений статус згідно міжнародних норм та конвенцій, є серйозним порушенням міжнародного гуманітарного права. Вони призводять до глобального враження на мобільність, економіку та повсякденне життя населення, а також загрожують безпеці та добробуту місцевого населення.

Загалом можна стверджувати, що гіпотеза вірна, так як в ці дні відбувались жахливі події пов'язані з російською агресією. Одним з факторів, що саме ці події є аномальними, свідчить розвиненість українського інформаційного простору. Кожного дня медійні люди

поширюють правду щодо дій окупантів в соціальних мережах, таких як Твіттер та Фейсбук. Це є однією з причин завантаженості BGP маршрутів.

### Висновки до розділу 3

На підставі аналізу BGP повідомлень, підрахунку кількості збоїв мережі за встановленими критеріями та використання методу пошуку аномалій відстань Кука на даних мережевого трафіку, можна зробити висновок, що гіпотеза про зв'язок між змінами мережевого трафіку та руйнуванням інфраструктури внаслідок війни підтверджена. Аналіз показав, що в період війни відбувалися значні зміни в мережевому трафіку, пов'язані зі збоями мережі, що можуть бути безпосередньо пов'язані з руйнуванням інфраструктури внаслідок військових дій. Результати дослідження дозволяють зробити припущення про взаємозв'язок між військовими діями та впливом на мережевий трафік в досліджуваному регіоні. Однак, додаткові дослідження та аналіз можуть бути проведені для більш деталізованого розуміння цього зв'язку та підтвердження цих результатів. Ми успішно перевірили розроблену методику та пов'язали аномалій з важливими воєнними подіями.

Узагальнюючи, збільшення рівня збоїв в BGP трафіку може бути попередником значних подій або змін у мережі, що може бути наслідком підготовки до кібератак або інших несправностей в мережі, як це сталося на початку 24 лютого 2022 року.

## 4 РОЗРОБЛЕННЯ СТАРТАП ПРОЕКТУ

### 4.1 Опис ідеї стартап-проєкту

На початку розробки проєкту проводиться аналіз та формулювання змісту ідеї стартапу, можливих напрямків застосування та переваг, якими може зацікавитися потенційний користувач. Головна ідея, напрямки застосування та вигоди для користувача зібрані у Таблиці 4.1. Основні відмінності від існуючих аналогів та можливих замінників також представлені у Таблиці 4.2.

Таблиця 4.1 – Опис ідеї, напряму та вигоди стартап-проєкту

| Зміст ідеї                                                            | Напрямок застосування                     | Вигоди для користувачів                                                                          |
|-----------------------------------------------------------------------|-------------------------------------------|--------------------------------------------------------------------------------------------------|
| Розробка та впровадження методики виявлення змін в мережевому трафіку | Компанії та підприємства різних масштабів | Пошук змін в мережевому трафіку                                                                  |
|                                                                       |                                           | Реалізація проєкту з використанням Python забезпечує швидкий час відгуку та оптимальну швидкодію |
|                                                                       |                                           | Працює з великим набором даних                                                                   |

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

| № п/п | Техніко-економічні | Потенційні товари/концепції конкурентів | W (слабка сторона) | N (нейтраль) | S (сильна сторона) |
|-------|--------------------|-----------------------------------------|--------------------|--------------|--------------------|
|       |                    |                                         |                    |              |                    |

Продовження таблиці 4.2

|   | характерис<br>тики ідеї                                                  | Мій<br>проект                                                                                                     | Конкурент                                                                                                    |                                                                                                           | на<br>сторона)                                                                                                                        |                                                                                |
|---|--------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| 1 | Швидка<br>обробка<br>великого<br>обсягу<br>BGP update<br>повідомле<br>нь | Використ<br>ання<br>аналітики<br>маршруті<br>в BGP на<br>основі<br>журналів<br>мережеви<br>х<br>пристроїв         | Аналіз<br>змін в<br>маршрутах<br>BGP на<br>основі<br>різних<br>алгоритмів<br>або<br>статистичн<br>их методів | Можливі<br>ложнопозитивні або<br>ложнодійс<br>ні<br>результати                                            | Потенція<br>л для<br>інтеграції<br>з різними<br>типами<br>мережови<br>х<br>пристроїв                                                  | Висока<br>ефективн<br>ість<br>виявленн<br>я<br>аномалій<br>в<br>реальному часі |
| 2 | Автоматиз<br>ація<br>процесу<br>виявлення<br>змін в<br>маршрутах<br>BGP  | Використ<br>ання<br>аналітики<br>мережево<br>го<br>трафіку<br>для<br>виявленн<br>я змін в<br>BGP<br>маршрута<br>х | Не<br>спеціалізу<br>ється                                                                                    | Вимоги до<br>високої<br>обчислюва<br>льної<br>потужності<br>для<br>обробки<br>великого<br>обсягу<br>даних | Можливіс<br>ть<br>налаштув<br>ання<br>правил<br>виявленн<br>я<br>аномалій<br>для<br>різних<br>сценаріїв<br>мережево<br>го<br>трафіку. | Ефективн<br>ість<br>обробки<br>даних                                           |

## Кінець таблиці 4.2

|   |                                        |                            |                    |                                                   |                                           |                            |
|---|----------------------------------------|----------------------------|--------------------|---------------------------------------------------|-------------------------------------------|----------------------------|
| 3 | Використання алгоритму пошуку аномалій | Використання відстані Кука | Не спеціалізується | Можливі ложнопозитивні або ложнодійсні результати | Можливість обрати поріг для відбору даних | Ефективність обробки даних |
|---|----------------------------------------|----------------------------|--------------------|---------------------------------------------------|-------------------------------------------|----------------------------|

## 4.2 Технологічний аудит ідеї проєкту

У цьому підрозділі проведено аудит технологій для реалізації ідеї проєкту. Результати аналізу можливості здійснення ідеї проєкту зібрані в Таблиці 4.3.

Таблиця 4.3 - Технологічні можливості здійснення ідеї

| № п/п | Ідея стартап-проєкту | Технології реалізації | Наявність технології | Доступність технології |
|-------|----------------------|-----------------------|----------------------|------------------------|
| 1     | Збір BGP даних       | Wireshark             | Доробити             | Доступні               |
| 2     | Обробка повідомлень  | Bgpdump, Python       | Наявні               | Доступні               |
| 3     | Виявлення аномалій   | Відстань Кука, Python | Наявні               | Доступні               |

## 4.3 Аналіз ринкових можливостей запуску стартап-проєкту

В даному підрозділі проведено аналіз попиту, включаючи наявність, динаміку розвитку ринку та обсяг попиту. Також були визначені ринкові можливості, які можна використати під час впровадження проєкту, а також

ринкові загрози, які можуть перешкодити реалізації проекту. Результати цього аналізу зібрані в Таблиці 4.4.

Таблиця 4.4 - Характеристика потенційного ринку для проекту

| №<br>п/п | Показники стану ринку                                      | Характеристика |
|----------|------------------------------------------------------------|----------------|
| 1        | Кількість основних гравців                                 | Не відомо      |
| 2        | Загальний обсяг продажів грн/ум.од.                        |                |
| 3        | Динаміка ринку                                             | Зростає        |
| 4        | Наявність обмежень для виходу                              | Немає          |
| 5        | Специфічні умови для сертифікації чи дотримання стандартів | Немає          |
| 6        | Середня норма рентабельності на ринку, %                   | >120%          |

З цього аналізу можна зробити висновок про відкритість ринку, який зараз зростає. Додатково, в даній сфері не було знайдено конкурентів, що робить ринок інвестиційно привабливим.

Таблиця 4.5 – Потенційні клієнти продукту

| №<br>п/п | Потреба, що формує ринок                     | Цільова аудиторія                                      | Відмінності у поведінці різних потенційних цільових груп клієнтів  | Вимоги споживачів до товару         |
|----------|----------------------------------------------|--------------------------------------------------------|--------------------------------------------------------------------|-------------------------------------|
| 1        | Збільшення кількості атак на мережі компанії | 1.Підприємства, які займаються захистом від кібератак. | Перша група зацікавлення в розширенні власних послуг відповідно до | Якісне виявлення аномалій в трафіку |

Кінець таблиці 4.5

|  |  |                                         |                                                                                                                                 |  |
|--|--|-----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|--|
|  |  | 2.Підприємства,<br>що були<br>атаковані | потреб<br>замовників.<br>Друга, власне, і<br>є замовниками,<br>що бажають<br>уникнути<br>репутаційних<br>та фінансових<br>втрат |  |
|--|--|-----------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|--|

Таблиця 4.6 - Фактори загроз

| №<br>п/п | Фактор                               | Зміст загрози                                                                                      | Можлива реакція<br>компанії                                                                          |
|----------|--------------------------------------|----------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| 1        | Можливі випадки хибного<br>виявлення | Набір даних<br>може бути<br>недостатнім або<br>методика відбору<br>даних може<br>давати хибні дані | Постійне<br>покращення<br>методики<br>відбору даних.<br>Впровадження<br>мінімуму для<br>набору даних |
| 2        | Відсутність попиту                   | Недооцінення<br>аналізу BGP<br>трафіку                                                             | Участь в<br>спеціалізованих<br>конференціях,<br>запуск рекламної<br>компанії                         |

Таблиця 4.7 – Фактори можливостей

| №<br>п/п | Фактор                    | Зміст можливості                            | Можлива реакція<br>компанії        |
|----------|---------------------------|---------------------------------------------|------------------------------------|
| 1        | Удосконалення<br>методики | Тестування інших методів<br>пошуку аномалій | Перехід на нову<br>версію методики |

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

|                                                   |                                                                                             |                                                                                              |
|---------------------------------------------------|---------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| Особливості<br>конкурентного<br>середовища        | В чому проявляється<br>дана характеристика                                                  | Вплив діяльності<br>підприємства (можливі<br>дії компанії, щоб бути<br>конкурентоспроможною) |
| Вказати тип<br>конкуренції -<br>олігополія        | Відсутня конкуренція                                                                        | Вузька спеціалізація<br>забезпечує якість і<br>актуальністю                                  |
| За рівнем конкурентної<br>боротьби -національна   | Послуги<br>надаватимуться також<br>за межами території<br>України                           | Вдосконалення<br>методики                                                                    |
| За галуззю -<br>міжгалузева                       | В послугах зацікавлені<br>компанії із різних<br>галузей                                     | Пошук нових клієнтів                                                                         |
| За видами товарів -<br>товарно-видова             | Надана послуга є дуже<br>специфічною, що<br>робить її унікальною                            | Надання послуг у сфері<br>кібербезпеки                                                       |
| За характером<br>конкурентних переваг -<br>цінова | Основна мотивація для<br>замовників –<br>забезпечення безпеки<br>власної<br>інфраструктури, | Надання послуг                                                                               |

Кінець таблиці 4.8

|                               |                                       |                           |
|-------------------------------|---------------------------------------|---------------------------|
|                               | швидке реагування на<br>можливі атаки |                           |
| За інтенсивністю -<br>марочна | Привабливість послуги                 | Вдосконалення<br>методики |

Таблиця 4.9 - Аналіз галузевої конкуренції за М. Портером

| Складові<br>і аналізу | Прямі<br>конкуренти<br>в галузі                 | Потенційні<br>конкуренти<br>в галузі                    | Постачальники                                   | Клієнти                                                          | Товари<br>замінники              |
|-----------------------|-------------------------------------------------|---------------------------------------------------------|-------------------------------------------------|------------------------------------------------------------------|----------------------------------|
|                       | Немає                                           | Компанії з<br>надання<br>послуг<br>кібербезпеки         | Компанії з<br>надання<br>послуг<br>кібербезпеки | Підприємства,<br>стурбовані<br>станом<br>власної<br>кібербезпеки | Немає                            |
| Висновки              | Низька<br>інтенсивність<br>боротьби<br>за ринок | Висока<br>можливість<br>успішного<br>виходу на<br>ринок | Не диктують<br>умови<br>роботи<br>ринку         | Стурбовані<br>станом<br>власної<br>кібербезпеки                  | Немає<br>обмежень<br>на<br>ринку |

Таблиця 4.10 - Обґрунтування факторів конкурентоспроможності

| № п/п | Фактор<br>конкурентоспроможності | Обґрунтування                                                  |
|-------|----------------------------------|----------------------------------------------------------------|
| 1     | Наявність прямих конкурентів     | Аналогічних спеціалізованих<br>послуг немає в вільному продажі |

Кінець таблиці 4.10

|   |                       |                                                          |
|---|-----------------------|----------------------------------------------------------|
| 2 | Якість виявлення змін | Застосовуються алгоритми, які підходять до цієї методики |
|---|-----------------------|----------------------------------------------------------|

Таблиця 4.11 - Порівняльний аналіз сторін проекту

| № п/п | Фактор конкурентоспроможності                     | Бали 1-20 | Рейтинг товарів-конкурентів у порівнянні з ... (назва підприємства) |   |   |   |   |   |   |
|-------|---------------------------------------------------|-----------|---------------------------------------------------------------------|---|---|---|---|---|---|
|       |                                                   |           | 3                                                                   | 2 | 1 | 0 | 1 | 2 | 3 |
| 1     | Наявність прямих конкурентів – унікальний продукт | 15        |                                                                     |   |   | + |   |   |   |
| 2     | Якість виявлення змін                             | 15        |                                                                     |   |   | + |   |   |   |

Таблиця 4.12 — SWOT- аналіз стартап-проекту

|                                                                                                                                                                    |                                                                                                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| Сильні сторони: завдяки вузькій спеціалізації є проект є унікальним. Також перевагами є ціна , відсутність конкуренції та можливість застосування у різних галузях | Слабкі сторони: треба проводити багато експериментальних досліджень, занадто вузька спеціалізація |
| Можливості: вдосконалення методики, підвищення ціни за продукт                                                                                                     | Загрози: поява конкурентів                                                                        |

Таблиця 4.13 — Альтернативи ринкового впровадження стартап-проекту

| №<br>п/п | Альтернатива (орієнтовний комплекс заходів) ринкової поведінки | Ймовірність отримання ресурсів | Строки реалізації |
|----------|----------------------------------------------------------------|--------------------------------|-------------------|
| 1        | Оформлення патенту                                             | Висока                         | До 8 місяців      |
| 2        | Реклама                                                        | Середня                        | До 3 місяців      |
| 3        | Підписання договору з клієнтами                                | Висока                         | До 1 року         |
| 4        | Вдосконалення методики і реалізації                            | Висока                         | До 1 року         |

#### 4.4 Розроблення ринкової стратегії проекту

Таблиця 4.14 — Вибір цільових груп потенційних споживачів

| №<br>п/п | Опис профілю цільової групи потенційних клієнтів      | Готовність споживачів сприйняти продукт | Орієнтовний попит в межах груп | Інтенсивність конкуренції в секторі | Простота входу |
|----------|-------------------------------------------------------|-----------------------------------------|--------------------------------|-------------------------------------|----------------|
| 1        | Підприємства, що надають послуги у сфері кібербезпеки | Є попит                                 | Середній                       | Низька                              | Низька         |
| 2        | Фізичні особи та компанії, стурбовані                 | Є попит                                 | Високий                        | Низька                              | Низька         |

Кінець таблиці 4.14

|  |                                |  |  |  |  |
|--|--------------------------------|--|--|--|--|
|  | станом власної<br>кібербезпеки |  |  |  |  |
|--|--------------------------------|--|--|--|--|

Таблиця 4.15 - Визначення базової стратегії розвитку

| Обрана<br>альтернатива<br>розвитку<br>проекту | Стратегія<br>охоплення<br>ринку | Ключові<br>конкурентоспроможні<br>позиції відповідно до<br>обраної альтернативи | Базова<br>стратегія<br>розвитку                                                  |
|-----------------------------------------------|---------------------------------|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Реклама                                       | Конференції,<br>блог            | Великі перспективи<br>та визначні якості<br>наявного рішення                    | Підвищувати<br>якість<br>надаваних<br>послуг,<br>активна<br>рекламна<br>кампанії |

Таблиця 4.16 — Визначення базової стратегії конкурентної поведінки

| Чи є проект<br>«першопрохідцем»<br>на ринку? | Чи буде<br>компанія<br>шукати нових<br>споживачів, | Чи буде<br>компанія<br>копіювати<br>основні | Стратегія<br>конкурентної<br>поведінки |
|----------------------------------------------|----------------------------------------------------|---------------------------------------------|----------------------------------------|
|----------------------------------------------|----------------------------------------------------|---------------------------------------------|----------------------------------------|

Кінець таблиці 4.16

|    |                                      |                                          |                  |
|----|--------------------------------------|------------------------------------------|------------------|
|    | або забирати існуючих у конкурентів? | характеристики товару конкурента, і які? |                  |
| Ні | Знайти нових і старих                | Частково                                 | Стратегія лідера |

Таблиця 4.17 — Визначення стратегії позиціонування

|                                     |                           |                                                                              |                                                                                          |
|-------------------------------------|---------------------------|------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| Вимоги до товару цільової аудиторії | Базова стратегія розвитку | Ключові конкурентоспроможні позиції власного стартап-проекту                 | Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових) |
| Швидкість, якість                   | Стратегія диференціації   | Консультація з методів виявлення аномалій. Виявлення змін з високою точністю | Швидкість і якість                                                                       |

#### 4.5 Розроблення маркетингової програми проекту

Таблиця 4.18 — Визначення ключових переваг концепції потенційного товару

| Потреби                                        | Вигоди, що пропонує товар | Ключові переваги над конкурентами |
|------------------------------------------------|---------------------------|-----------------------------------|
| Виявлення атак і швидке реагування під час них | Точність і швидкість      | Вузька спеціалізація              |

Таблиця 4.19 - Опис 3 рівнів моделі товару

| Рівні                       | Сутність та складові                            |  |  |
|-----------------------------|-------------------------------------------------|--|--|
| Товар за задумом            | Система для виявлення змін у трафіку            |  |  |
| Товар у реальному виконанні | Властивості                                     |  |  |
|                             | Збір та обробка BGP даних                       |  |  |
|                             | Виявлення змін                                  |  |  |
|                             | До продажу: активне рекламування власних послуг |  |  |

Кінець таблиці 4.19

|                                                                        |                                                                                 |
|------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Товар із підкріпленням                                                 | Після продажу: участь у спеціалізованих конференціях, реклама власного продукту |
| За рахунок чого потенційний товар буде захищено від копіювання: патент |                                                                                 |

Таблиця 4.20 — Визначення меж встановлення ціни

|                                  |                                |                                          |                      |
|----------------------------------|--------------------------------|------------------------------------------|----------------------|
| Рівень цін на товари - замітники | Рівень цін на товари - аналоги | Рівень доходів цільової групи споживачів | Верхня та нижня межі |
| -                                | -                              | Високий                                  | 5000-15000 у.о.      |

Таблиця 4.21 - Формування системи збуту

|                                                 |                                                |                       |                          |
|-------------------------------------------------|------------------------------------------------|-----------------------|--------------------------|
| Специфіка закупівельної поведінки цільових груп | Функції збуту, які має виконувати постачальник | Глибина каналів збуту | Оптимальна система збуту |
| Власна система збуту                            | Консультації                                   | Продаж користувачам   | Продаж без посередників  |

Таблиця 4.22 — Концепція маркетингових комунікацій

| Специфіка поведінки цільових клієнтів | Канали комінукацій якими користуються клієнти | Ключові пропозиції для позиціонування               | Завдання рекламного повідомлення |
|---------------------------------------|-----------------------------------------------|-----------------------------------------------------|----------------------------------|
| Консервативна поведінка               | Пошта                                         | Виявлення змін у трафіку BGP, швидке виявлення атак | Масова розсилка                  |

#### Висновки до розділу 4

У четвертому розділі дипломної роботи було проведено аналіз перспектив запуску стартап-проекту, спрямованого на надання послуг в галузі виявлення атак на відмову в обслуговуванні в мережах Інтернету речей. В результаті дослідження було виявлено, що запуск проекту можливий, але може зустріти певні труднощі.

Однак, враховуючи активний розвиток Інтернету речей, спостерігається зростання попиту на такі послуги, що забезпечує певну стабільність та перспективність для даного проекту. Крім того, конкуренція в даній галузі на даний момент не є високою, що дозволяє стартапу мати шанси на успіх при виході на ринок.

Таким чином, аналіз перспектив для запуску стартап-проекту показав, що він може бути успішним, хоча необхідно врахувати певні труднощі, які можуть виникнути під час його реалізації.

## ВИСНОВКИ

В результаті проведеної роботи було детально проаналізовано вплив війни в Україні на стан мережі, зокрема розроблено і перевірено методику виявлення змін у трафіку BGP. Перед початком експериментального дослідження, виконано огляд літератури, зокрема протоколу BGP, його структури, BGP повідомлень, атак на протокол та методів захисту від них.

Під час огляду літератури було також ретельно проаналізовано різні методи пошуку аномалій, із яких було вибрано відстань Кука як основний метод. Далі, в ході експериментального дослідження, було розроблено методику виявлення змін у трафіку, під час якої було зібрано та оброблено дані BGP, з них було створено датасет, що використовувався для застосування методу пошуку аномалій на мережевих даних.

Результати дослідження показали виявлення 4 аномальних подій, які мають тісний зв'язок з подіями в російсько-українській війні. Це підтвердило гіпотезу, поставлену на початку роботи, і підтвердило успішну розробку і використання методики для досягнення мети та виконання завдань дослідження. Додатково, ми прийшли до висновку, що висока кількість збоїв в BGP трафіку може передувати важливим подіям або змінам в мережі через підготовку до кібератак.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. What is Open Shortest Path First (OSPF)? [Електронний ресурс] – Режим доступу до ресурсу: <https://www.metaswitch.com/knowledge-center/reference/what-is-open-shortest-path-first-ospf>.
2. Pedamkar. What is EIGRP? [Електронний ресурс] / Pedamkar – Режим доступу до ресурсу: <https://www.educba.com/what-is-eigrp/>.
3. What Is Border Gateway Protocol (BGP)? [Електронний ресурс] – Режим доступу до ресурсу:  
<https://www.fortinet.com/resources/cyberglossary/bgp-border-gateway-protocol>.
4. Difference between EIGRP, OSPF vs BGP Routing Protocols? [Електронний ресурс]– Режим доступу до ресурсу:  
<https://networkjourney.com/difference-between-eigrp-ospf-vs-bgp-routing-protocols/>.
5. Border Gateway Protocol [Електронний ресурс] – Режим доступу до ресурсу: [https://en.wikipedia.org/wiki/Border\\_Gateway\\_Protocol](https://en.wikipedia.org/wiki/Border_Gateway_Protocol).
6. BGP messages formats [Електронний ресурс] – Режим доступу до ресурсу:  
[https://techhub.hpe.com/eginfolib/networking/docs/switches/3600v2/5998-7619r\\_l3-ip-rtnng\\_cg/content/442284290.htm](https://techhub.hpe.com/eginfolib/networking/docs/switches/3600v2/5998-7619r_l3-ip-rtnng_cg/content/442284290.htm).
7. 4.4 KEEPALIVE Message Format [Електронний ресурс] – Режим доступу до ресурсу: <https://www.freesoft.org/CIE/RFC/1771/10.htm>.
8. Vulnerabilities of BGP [Електронний ресурс] – Режим доступу до ресурсу: <https://www.catchpoint.com/blog/bgp-vulnerabilities>.
9. Murphy. BGP Security Vulnerabilities Analysis [Електронний ресурс] / Murphy – Режим доступу до ресурсу: <https://www.rfc-editor.org/rfc/rfc4272>.
10. Sklar. How does BGP routing help to mitigate DDoS attacks? [Електронний ресурс] / Sklar – Режим доступу до ресурсу:

- <https://www.peerspot.com/questions/how-does-bgp-routing-help-to-mitigate-ddos-attacks>.
11. Understanding Cloudflare DDoS protection [Електронний ресурс] – Режим доступу до ресурсу:  
<https://developers.cloudflare.com/support/about-cloudflare/attack-preparation-and-response/understanding-cloudflare-ddos-protection/>.
  12. Cook's Distance [Електронний ресурс] – Режим доступу до ресурсу:  
<https://rpubs.com/DragonflyStats/Cooks-Distance>.
  13. Frost. Interquartile Range (IQR): How to Find and Use It [Електронний ресурс] / Frost – Режим доступу до ресурсу:  
<https://statisticsbyjim.com/basics/interquartile-range/>.
  14. Lewinson. Outlier Detection with Isolation Forest [Електронний ресурс] / Lewinson – Режим доступу до ресурсу:  
<https://towardsdatascience.com/outlier-detection-with-isolation-forest-3d190448d45e>.
  15. Stepanov. Anomaly detection using Minimum Covariance Determinant (MCD) [Електронний ресурс] / Stepanov – Режим доступу до ресурсу:  
<https://medium.com/geekculture/anomaly-detection-using-minimum-covariance-determinant-mcd-41dc28ccc8a>.
  16. Multi-criteria Anomaly Detection using Pareto Depth Analysis [Електронний ресурс] / Hsiao, Xu, Calder, Hero III – Режим доступу до ресурсу: <https://arxiv.org/pdf/1110.3741.pdf>.
  17. Bgpdata [Електронний ресурс] – Режим доступу до ресурсу:  
<http://archive.routeviews.org/bgpdata/>.
  18. bgpdump [Електронний ресурс] – Режим доступу до ресурсу:  
<https://manpages.ubuntu.com/manpages/bionic/man1/bgpdump.1.html>.
  19. Yellowbrick: Machine Learning Visualization [Електронний ресурс] – Режим доступу до ресурсу: <https://www.scikit-yb.org/en/latest/>.
  20. Суспільне. Залізничний вокзал у Краматорську 8 квітня обстріляли з окупованої Донеччини — СБУ [Електронний ресурс] / Суспільне –

- Режим доступу до ресурсу: <https://suspilne.media/241830-zaliznicnij-vokzal-u-kramatorsku-8-kvitna-obstrilali-z-okupovanoi-doneccini-sbu/>.
21. Морпіхи 36-ї бригади в Маріуполі: частина з'єдналася з “Азовом”, частина – ішла на прорив, багато потрапили в полон [Електронний ресурс] – Режим доступу до ресурсу: <https://web.archive.org/web/20220421202842/https://novynarnia.com/2022/04/13/morpihy-36-yi-brygady-v-mariupoli-chastyna-zyednalasya-z-azovom-chastyna-ishla-na-proryv-bagato-potrapyly-v-polon/>.
22. Тарута заявив, що в Маріуполі окупанти скинули бомбу на лікарню в районі Азовсталі. У міськраді спростували, що там були люди [Електронний ресурс] – Режим доступу до ресурсу: <https://web.archive.org/web/20220420051718/https://nv.ua/ukr/ukraine/events/mariupol-azovstal-rosiya-zavdala-udaru-po-likarni-v-rayoni-zavodu-zayaviv-taruta-50235176.html>.
23. Корогодський. 25 квітня окупанти стріляли ракетами по залізниці у чотирьох областях України [Електронний ресурс] / Корогодський – Режим доступу до ресурсу: [https://lb.ua/society/2022/04/25/514665\\_25\\_kvitnya\\_okupanti\\_strilyali\\_raketami.html](https://lb.ua/society/2022/04/25/514665_25_kvitnya_okupanti_strilyali_raketami.html).
24. Mueller. Machine Learning Security Principles: Keep data, networks, users, and applications safe from prying eyes / Mueller, Stephens., 2022. – 450 с.
25. Siddiqui. Did Ukraine suffer a BGP hijack and how can networks protect themselves? [Електронний ресурс] / Siddiqui – Режим доступу до ресурсу: <https://www.manrs.org/2022/03/did-ukraine-suffer-a-bgp-hijack-and-how-can-networks-protect-themselves/>.

## Додаток А

```

import subprocess,os,shutil

ym=""
ymdot=""
ymd=""
hm = "
#CHANGE
dir = '/Users/viktoriia/Downloads/Diploma/Practice/'
year_start = 2022
year_end = 2023
month_start = 2
month_end = 12
day_start = 1
day_end = 31
hour_start = 0
hour_end = 23
min_start = 0
min_end = 45

#download BGP data
def download():
    for year in range(year_start,year_end+1):
        for month in range(month_start,month_end+1):
            for day in range(day_start,day_end+1):
                if (year == 2022 and month == 1) or (year == 2023 and month > 2) or
(year == 2022 and month == 2 and day < 24):
                    continue
                if month < 10:

```

```

        ymdot = str(year)+'.'+str(month)
        ym = str(year)+'0'+str(month)
    else:
        ymdot = str(year)+'.'+str(month)
        ym = str(year)+str(month)
    if day < 10:
        ymd = ym+'0'+str(day)
    else:
        ymd = ym+str(day)
    for hour in range(hour_start, hour_end+1):
        for min in range(min_start, min_end+1, 15):
            if hour < 10:
                hm = '0'+str(hour)
            else:
                hm = str(hour)
            if min < 10:
                hm = hm+'0'+str(min)
            else:
                hm = hm + str(min)
            try:
                url =
'http://archive.routeviews.org/bgpdata/%s/UPDATES/updates.%s.%s.bz2' %
(ymdot, ymd, hm)
                print(url)
                subprocess.run(["wget", "-r", "-nc", "-P", dir, url])
            except:
                print("continue..")

#decompress and parse with bgpdump
def parse():

```

```

for year in [2022,2023]:
    for month in range(1,13):
        for day in range(1,32):
            if (year == 2022 and month == 1) or (year == 2023 and month > 2) or
(year == 2022 and month == 2 and day < 24):
                continue
            if month < 10:
                ymdot = str(year)+'0'+str(month)
                ym = str(year)+'0'+str(month)
            else:
                ymdot = str(year)+''+str(month)
                ym = str(year)+str(month)
            if day < 10:
                ymd = ym+'0'+str(day)
            else:
                ymd = ym+str(day)
            for hour in range(hour_start,hour_end+1):
                for min in range(min_start,min_end+1,15):
                    if hour < 10:
                        hm = '0'+str(hour)
                    else:
                        hm = str(hour)
                    if min < 10:
                        hm = hm+'0'+str(min)
                    else:
                        hm = hm + str(min)
                    #CHANGE
                    file =
'/Users/viktoriiia/Downloads/Diploma/Practice/archive.routeviews.org/bgpdata/%s/
UPDATES/updates.%s.%s.bz2' % (ymdot,ymd,hm)

```

```
file2 =  
'/Users/viktoriiia/Downloads/Diploma/Practice/files/%s.%s' % (ymd,hm)  
print(file)  
if os.path.exists(file) and os.path.isfile(file):  
    print(file)  
    print("\nFile: ', file)  
    convert = 'bgpdump -m %s > %s' %(file,file2)  
    os.system(convert)  
  
download()  
parse()
```

## Додаток Б

```
# Import the required libraries

#!/usr/bin/env python3

from datetime import datetime, timedelta
import pytricia
import os
import ipaddress
from geoip2 import database
pt = pytricia.PyTricia()
import json
import pandas as pd

# files
#CHANGE
geo_reader =
database.Reader('/Users/viktoriia/Downloads/Diploma/Practice/GeoLite2-
City.mmdb', locales=['en'])
files_dir = "/Users/viktoriia/Downloads/Diploma/Practice/files/"

# Initialize the data structures
db = []
outage_time = 300
country = 'Ukraine'
outage_by_Kyiv = 0
outage_by_Kharkiv = 0
```

# Helper function to extract the message type, prefix, and timestamp from a BGP message

```
def extract_bgp_message(message):
    # Extract the fields from the message using the regular expression pattern
    fields = message.strip().split("|")
    type = fields[2]
    prefix = ipaddress.ip_network(fields[5]).exploded
    timestamp = int(fields[1].split('.')[0])
    return type, prefix, timestamp
```

# Helper function to check if a prefix is a Ukrainian network

```
def is_ukrainian_prefix(prefix):
    try:
        city_rec = geo_reader.city(prefix.split('/')[0])
        country_name = city_rec.country.name
    except:
        return " "
    return country_name
```

# Loop over the BGP messages

```
def parse_files():
    files = sorted([os.path.join(files_dir, f) for f in os.listdir(files_dir) if
os.path.isfile(os.path.join(files_dir, f))])
    for file in files:
        print(file)
        with open(file, 'r') as bgp_messages:
            for message in bgp_messages:
                # Extract the message type, prefix and timestamp from the message
                bgp_message = extract_bgp_message(message)
                if bgp_message is None:
```

```

        continue

    type, prefix, timestamp = bgp_message

    # Check if the prefix is a Ukrainian network
    if is_ukrainian_prefix(prefix) != country:
        continue

    if type == "A" and not pt.has_key(prefix):
        pt.insert(prefix, timestamp)
    elif type == "W" and pt.has_key(prefix):
        duration = timestamp - pt[prefix]
        if duration > outage_time:
            # Create a dictionary object to store the outage information
            outage_object = {
                'prefix': prefix,
                'start_time': pt[prefix],
                'end_time': timestamp
            }
            db.append(outage_object)

            # Remove the prefix from the PyTricia object
            pt.delete(prefix)

    print(f'Ukraine: {len(db)}, time: {datetime.now()}')

# save them to a file
with open("dataset1.json", "w") as f:
    json.dump(db, f)

start_time = datetime.now()
print(f'Start time: {start_time}')
parse_files()
end_time = datetime.now()
print(f'End time: {end_time}')

```

## Додаток В

```

import pandas as pd
from yellowbrick.regressor import CooksDistance
from sklearn.linear_model import LinearRegression
import json
from datetime import datetime, timedelta
import os
import pytz
import statsmodels.api as sm
import matplotlib.pyplot as plt

# Load data from json file
files_dir = "/Users/viktoriiia/Downloads/Diploma/Practice/db/"

def parse_db():
    # Specify the start and end times of the time range
    data_list = []
    files = sorted([os.path.join(files_dir, f) for f in os.listdir(files_dir) if
os.path.isfile(os.path.join(files_dir, f))])
    for file in files:
        print(file)
        db = json.load(open(file))
        db = sorted(db, key=lambda d: d["start_time"])
        #
        ## Count the number of outages that occurred during the time range
        temp_date =
datetime.fromtimestamp(db[0]['start_time'],pytz.timezone("UTC")).day
        num_outages = 0
        day_month_year_format = '%Y-%m-%d'

```

```

    for outage in db:
        start =
datetime.fromtimestamp(outage['start_time'],pytz.timezone("UTC")).day
        if start != temp_date:
            date_insert =
datetime.fromtimestamp(outage['start_time'],pytz.timezone("UTC")) -
timedelta(days=1)
            data_dict = {
                "date": date_insert.strftime(day_month_year_format),
                "outages": num_outages
            }
            data_list.append(data_dict)
            temp_date=start
            num_outages = 0
            num_outages += 1
    with open("data1.json", "w") as f:
        json.dump(data_list, f)

```

```

def create_plot():
    data_dict = pd.read_json('data1.json')
    # Extract data from data_dict
    dates = data_dict["date"]
    outages = data_dict["outages"]

    # Create a plot
    plt.figure(figsize=(10, 6))
    plt.plot(dates, outages, marker='o')
    plt.title("Outages by Date")
    plt.xlabel("Date")
    plt.ylabel("Outages")

```

```
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```

```
def cook_distance():
```

```
    df = pd.read_json('data1.json')
    print(df['date'][40])
    print(df['date'][44])
    print(df['date'][50])
    print(df['date'][55])
```

```
    # Fit a linear regression model
```

```
    X = df['outages'].values.reshape(-1, 1)
    y = range(len(X))
    model = LinearRegression().fit(X, y)
```

```
    # Visualize Cook's distance for each observation
```

```
    visualizer = CooksDistance()
    visualizer.fit(X, y)
    visualizer.show()
```

```
def parse_db_by_hour():
```

```
    # Specify the start and end times of the time range
```

```
    data_list = []
```

```
    files = sorted([os.path.join(files_dir, f) for f in os.listdir(files_dir) if
os.path.isfile(os.path.join(files_dir, f))])
```

```
    for file in files:
```

```
        print(file)
```

```
        db = json.load(open(file))
```

```
        db = sorted(db, key=lambda d: d["start_time"])
```

```

#
## Count the number of outages that occurred during the time range
temp_date =
datetime.fromtimestamp(db[0]['start_time'],pytz.timezone("UTC"))
print(temp_date)
print(datetime(2022, 2, 25,tzinfo=pytz.UTC))
if temp_date > datetime(2022, 2, 25,tzinfo=pytz.UTC):
    break
num_outages = 0
temp_date=temp_date.hour
for outage in db:
    start =
datetime.fromtimestamp(outage['start_time'],pytz.timezone("UTC"))
    if start > datetime(2022, 2, 25,tzinfo=pytz.UTC):
        break
    start = start.hour
    if start != temp_date:
        #date_insert =
datetime.fromtimestamp(outage['start_time'],pytz.timezone("UTC")) -
timedelta(days=1)
        data_dict = {
            "date": temp_date,
            "outages": num_outages
        }
        data_list.append(data_dict)
        temp_date=start
        num_outages = 0
        num_outages += 1
with open("data_by_hour.json", "w") as f:
    json.dump(data_list, f)

```

```
data_dict = pd.read_json('data_by_hour.json')  
# Extract data from data_dict  
dates = data_dict["date"]  
outages = data_dict["outages"]
```