ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ
КАФЕДРА ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМ ТА МЕРЕЖ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «АНАЛІЗ МЕРЕЖЕВОГО ТРАФІКУ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-РОЗКЛАДАННЯ»

на здобуття освітнього ступеня магістра
зі спеціальності 172 Електронні комунікації та радіотехніка
(код, найменування спеціальності)
освітньо-професійної програми Телекомунікаційні системи та мережі
(назва)

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

_____ Дмитро ГОЛОВНЯ
(підпис) *(Ім'я ПРІЗВИЩЕ)*

Виконав: здобувач вищої освіти гр. ТСДМ-61
Дмитро ГОЛОВНЯ
Ім'я, ПРІЗВИЩЕ

Керівник: Дмитро МИРОНОВ
к.т.н. *Ім'я, ПРІЗВИЩЕ*

Рецензент: _____
Ім'я, ПРІЗВИЩЕ

Київ 2025

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
Навчально-науковий інститут Телекомунікацій

Кафедра ___Телекомунікаційних систем та мереж_____
Ступінь вищої освіти __магістр_____
Спеціальність ___172 Електронні комунікації та радіотехніка _____
Освітньо-професійна програма _____Телекомунікаційні системи та мережі_____

ЗАТВЕРДЖУЮ
Завідувач кафедри __ТСМ_____
_____ Віктор ЗАЇКА
«____» _____ 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Головні Дмитру Олексійовичу
(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи: Аналіз мережевого трафіку з використанням вейвлет-розкладання
керівник кваліфікаційної роботи Дмитро МИРОНОВ, к.т.н.
(Ім'я, ПРІЗВИЩЕ науковий ступінь, вчене звання)
затверджені наказом Державного університету інформаційно-комунікаційних технологій
від «30» 10. 2025 р. № 467

2. Строк подання кваліфікаційної роботи «___» _____ 2025 р.

3. Вихідні дані до кваліфікаційної роботи:
1. Вейвлет-розкладання.
2. Дзеркалення трафіка.
3. Аналіз мережевого трафіку.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
1. Поняття вейвлет-аналізу в телекомунікаціях.
2. Поняття дзеркалення трафіка.
3. Методи оптимізації аналізу мережевого трафіку з використанням вейвлет-перетворення.
4. Реалізація методики та моделювання результатів аналізу мережевого трафіку.

5. Перелік ілюстративного матеріалу: ___презентація на____слайдах_____

6. Дата видачі завдання «___» _____ 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Підбір та аналіз науково-технічної літератури	15.10.2025	Викон.
2	Розробка першого розділу	20.10.2025	Викон.
3	Розробка другого розділу	30.10.2025	Викон.
4	Розробка третього розділу	10.11.2025	Викон.
5	Розробка четвертого розділу	20.11.2025	Викон.
6	Висновки, вступ, реферат	30.11.2025	Викон.
7	Розробка презентації	10.12.2025	Викон.

Здобувач вищої освіти _____ Дмитро ГОЛОВНЯ
(підпис) *(Ім'я ПРІЗВИЩЕ)*

Керівник кваліфікаційної роботи _____ Дмитро МИРОНОВ
(підпис) *(Ім'я ПРІЗВИЩЕ)*

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
Навчально-науковий інститут телекомунікацій
ПОДАННЯ
ГОЛОВІ ЕКЗАМЕНАЦІЙНОЇ КОМІСІЇ
ЩОДО ЗАХИСТУ КВАЛІФІКАЦІЙНОЇ РОБОТИ
на здобуття освітнього ступеня магістра

Направляється здобувач Головня Д.О. до захисту кваліфікаційної роботи
(прізвище та ініціали)
за спеціальністю 172 Електронні комунікації та радіотехніка
(код, найменування спеціальності)
освітньо-професійної програми Телекомунікаційні системи та мережі
(назва)
на тему: «Аналіз мережевого трафіку з використанням вейвлет-розкладання»

Кваліфікаційна робота і рецензія додаються.

Директор ННІТ _____ Владислав КРАВЧЕНКО
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

Висновок керівника кваліфікаційної роботи

Здобувач Головня Дмитро Олексійович показав гарну теоретичну та інженерну підготовку, уміння володіти новими комп'ютерними технологіями, користуватися навчальною, довідковою і науково-технічною літературою. Працюючи над завданнями, які були поставлені керівником, проявив ініціативність, сумлінність та хист до наукової роботи.

Все це дозволяє оцінити виконану кваліфікаційну роботу здобувача Головні Дмитра Олексійовича на оцінку «відмінно» та присвоїти йому кваліфікацію магістр з електронних комунікацій та радіотехніки.

Керівник кваліфікаційної роботи _____ Дмитро МИРОНОВ
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

«____» _____ 20___ року

Висновок кафедри про кваліфікаційну роботу

Кваліфікаційна робота розглянута. Здобувач Головня Д.О. допускається до захисту даної роботи в Екзаменаційній комісії.

Завідувач кафедри телекомунікаційних систем та мереж _____ Віктор ЗАЇКА
(назва) *(підпис)* *(Ім'я, ПРІЗВИЩЕ)*

«____» _____ 20___ року

ВІДГУК РЕЦЕНЗЕНТА
на кваліфікаційну роботу
на здобуття освітнього ступеня магістра

здобувача вищої освіти Головні Дмитра Олексійовича
(прізвище, ім'я, по батькові)
на тему «Аналіз мережевого трафіку з використанням вейвлет-розкладання»

Актуальність.
Сучасні телекомунікаційні мережі генерують великі обсяги нестаціонарного трафіку, у якому поєднуються короткочасні аномалії та довготривалі тренди. Традиційні методи аналізу, зокрема перетворення Фур'є, не забезпечують достатньої часової локалізації, що ускладнює своєчасне виявлення кіберзагроз. Дискретне вейвлет-перетворення дозволяє досліджувати трафік на різних масштабах і виявляти аномальні події з високою точністю. Водночас коректність аналізу залежить від способу збору даних, оскільки технології TAP і SPAN мають різну достовірність. Це зумовлює потребу в оптимізації методів моніторингу та обробки трафіку.
Позитивні сторони.
1. У роботі була розроблена та описана архітектура системи збору та попередньої обробки мережевого трафіку, що включає використання TAP-пристроїв для пасивного перехоплення трафіку, високошвидкісних модулів захоплення пакетів, механізмів фільтрації, агрегації та нормалізації даних, а також компонентів довготривалого зберігання та аналітичної обробки.
2. У роботі було змодельовано та протестовано різні сценарії мережевої поведінки, включно з нормальним трафіком, сплесками навантаження, імітаціями DoS/DDoS-атак, скануванням портів та шумовими перешкодами.
Недоліки.
1. Недостатньо розкрито питання параметрів вибору вейвлет-функції та методик порогування.
2. Робота не містить порівняльного аналізу продуктивності різних типів вейвлетів та їх впливу на точність виявлення аномалій.

Відзначені зауваження не впливають на загальну позитивну оцінку магістерської кваліфікаційної роботи

Висновок: *кваліфікаційна робота на здобуття ступеня магістра заслуговує оцінку "відмінно", а здобувач Головня Дмитро Олексійович заслуговує присвоєння кваліфікації: магістр з електронних комунікацій та радіотехніки.*

Рецензент
науковий ступінь, вчене звання _____ (підпис) _____ *(Ім'я ПРІЗВИЩЕ)*

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня магістра: 77 стор., 17 рис., 5 табл., 27 джерел.

Мета роботи – розробити та обґрунтувати методику оптимізованого аналізу мережевого трафіку на основі дискретного вейвлет-перетворення з урахуванням особливостей отримання даних за допомогою технологій TAP і SPAN та можливістю ефективного виявлення аномальних подій у нестаціонарних сигналах.

Об’єкт дослідження –процеси формування, передавання та обробки мережевого трафіку в сучасних телекомунікаційних системах.

Предмет дослідження – методи й алгоритми аналізу нестаціонарного мережевого трафіку, зокрема вейвлет-перетворення та технології отримання трафіку (TAP і SPAN), що впливають на точність і ефективність виявлення аномалій.

Короткий зміст роботи: у роботі проведено аналіз можливостей застосування вейвлет-перетворення для дослідження мережевого трафіку та виявлення аномалій. Показано, що DWT ефективно працює з нестаціонарними сигналами, поєднуючи аналіз довготривалих тенденцій і коротких імпульсних подій, характерних для атак і пікових навантажень. Розглянуто основи DWT, порівняно різні сімейства вейвлетів і визначено оптимальні функції для аналізу трафіку. Створено архітектуру збору даних з використанням TAP-пристроїв і PyWavelets. Експериментальні результати підтвердили чутливість деталізуючих коефіцієнтів до аномалій та перевагу вейвлет-аналізу над класичними спектральними методами.

КЛЮЧОВІ СЛОВА: ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ, ДИСКРЕТНИЙ ВЕЙВЛЕТ-АНАЛІЗ, МЕРЕЖЕВИЙ ТРАФІК, АНОМАЛІЇ, DWT, TAP, SPAN, PYWAVELETS, ФІЛЬТРАЦІЙНІ БАНКИ, ДЕТАЛІЗУЮЧІ КОЕФІЦІЄНТИ, АНАЛІЗ НЕСТАЦІОНАРНИХ СИГНАЛІВ, КІБЕРБЕЗПЕКА, IDS/IPS, МОНІТОРИНГ МЕРЕЖ, ВИЯВЛЕННЯ АТАК

ABSTRACT

Text part of the master's qualification work: 77 pages, 17 figures, 5 table, 27 sources.

The purpose of the work is to develop and substantiate a methodology for optimized network traffic analysis based on discrete wavelet transform, taking into account the features of data acquisition using TAP and SPAN technologies and the ability to effectively detect anomalous events in non-stationary signals.

Object of the study is the processes of formation, transmission and processing of network traffic in modern telecommunication systems.

Subject of the study is methods and algorithms for analyzing non-stationary network traffic, in particular wavelet transform and traffic acquisition technologies (TAP and SPAN), which affect the accuracy and efficiency of anomaly detection.

Summary of the work: the paper analyzes the possibilities of using wavelet transform for network traffic research and anomaly detection. It is shown that DWT works effectively with non-stationary signals, combining the analysis of long-term trends and short pulse events characteristic of attacks and peak loads. The basics of DWT are considered, different families of wavelets are compared and optimal functions for traffic analysis are determined. A data acquisition architecture using TAP devices and PyWavelets is created. Experimental results confirmed the sensitivity of detailing coefficients to anomalies and the superiority of wavelet analysis over classical spectral methods.

KEYWORDS: WAVELET TRANSFORM, DISCRETE WAVELET ANALYSIS, NETWORK TRAFFIC, ANOMALIES, DWT, TAP, SPAN, PYWAVELETS, FILTER BANKS, DETAILING COEFFICIENTS, ANALYSIS OF NON-STATIONARY SIGNALS, CYBERSECURITY, IDS/IPS, NETWORK MONITORING, ATTACK DETECTION

ЗМІСТ

ВСТУП 9
РОЗДІЛ 1 ПОНЯТТЯ ВЕЙВЛЕТ-АНАЛІЗУ В ТЕЛЕКОМУНІКАЦІЯХ 11
1.1 Дискретне вейвлет – перетворення 11
1.2 Режим розширення сигналу 14
1.3 Використання банку фільтрів 19
РОЗДІЛ 2 ПОНЯТТЯ ДЗЕРКАЛЕННЯ ТРАФІКА 25
2.1 Технології TAP та SPAN 25
2.2 Дзеркалення трафіку та аналіз локальної мережі 31
РОЗДІЛ 3 МЕТОДИ ОПТИМІЗАЦІЇ АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ 41
3.1 Математичні основи та принципи оптимізованого DWT для мережевого аналізу 41
3.2 Особливості застосування вейвлет-розкладання для виявлення аномалій у мережевому трафіку 43
3.3 Оптимізація методики збору трафіку та попередня обробка даних для вейвлет-аналізу 48
РОЗДІЛ 4 РЕАЛІЗАЦІЯ МЕТОДИКИ ТА МОДЕЛЮВАННЯ РЕЗУЛЬТАТІВ АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ 56
4.1 Архітектура системи збору та обробки мережевого трафіку 56
4.2 Програмна реалізація вейвлет-розкладання для мережевого трафіку з використанням PyWavelets 63
4.3 Моделювання, експерименти та оцінка ефективності вейвлет-аналізу 70
ВИСНОВКИ 76
ПЕРЕЛІК ПОСИЛАНЬ 78
ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ 81

ВСТУП

Сучасні телекомунікаційні мережі генерують величезні обсяги трафіку, який постійно змінюється під впливом динаміки взаємодії користувачів, сервісів, прикладних систем, хмарних платформ та мобільних пристроїв. Така мінливість призводить до того, що мережевий трафік стає яскравим прикладом нестаціонарного сигналу, у якому поєднуються як тривалі низькочастотні тренди, так і короткочасні високочастотні імпульсні сплески. Виявлення аномалій, моніторинг стану мережі, забезпечення кібербезпеки та оптимізація продуктивності значною мірою залежать від здатності аналітичних систем коректно обробляти та інтерпретувати саме такі складні часові ряди. Тому актуальним є питання використання методів, здатних працювати з нестаціонарними процесами та забезпечувати багаторівневий опис змін у часі.

Традиційні методи спектрального аналізу, зокрема перетворення Фур'є, демонструють високу ефективність лише для стаціонарних сигналів, де частотні характеристики залишаються стабільними. У випадку мережевого трафіку цей підхід має суттєві обмеження, оскільки не дозволяє одночасно відстежувати локальні зміни у часі та спектрі. Саме тому у наукових дослідженнях та практичних системах аналізу все частіше застосовуються методи вейвлет-перетворення, які дозволяють розкладати сигнал на масштабні та локальні компоненти, забезпечуючи точне виявлення як глобальних тенденцій, так і короткочасних аномальних подій.

Вейвлет-аналіз є ефективним інструментом для роботи з мережевим трафіком, оскільки він поєднує у собі переваги часової та частотної локалізації. Застосування багаторівневого дискретного вейвлет-перетворення (DWT) дає можливість досліджувати структуру трафіку на різних частотних діапазонах, виявляти імпульсні аномалії, аналізувати циклічність, сегментувати поведінкові патерни користувачів та визначати приховані закономірності, недоступні через класичні методи фільтрації та перетворення сигналів.

Актуальність даної теми також зумовлена зростанням кількості кіберзагроз, що проявляються у мережевому трафіку у вигляді нетипових інтенсивностей,

стрибків активності, сканування портів, DDoS-атак та інших небажаних дій. Це вимагає наявності точних та швидких алгоритмів, здатних обробляти великі масиви даних без втрати інформативності. У цьому контексті вейвлет-перетворення та його програмна реалізація, зокрема за допомогою бібліотеки PyWavelets, відкривають широкі можливості для створення ефективних інструментів аналізу.

Отримані в роботі результати підтверджують доцільність використання вейвлет-перетворення для аналізу мережевого трафіку, демонструють його здатність забезпечувати глибоку багаторівневу інтерпретацію даних та підвищують якість виявлення аномальних подій. Методика, розроблена в межах дипломної роботи, може бути інтегрована у практичні системи моніторингу, SOC-центри, IDS/IPS-рішення та автоматизовані платформи аналізу продуктивності мереж.

1 ПОНЯТТЯ ВЕЙВЛЕТ-АНАЛІЗУ В ТЕЛЕКОМУНІКАЦІЯХ

1.1 Дискретне вейвлет – перетворення

Під час обробки даних на комп’ютері часто застосовують дискретизовану форму неперервного вейвлет-перетворення, основи якого були розглянуті в моїй попередній статті. Однак вибір дискретних значень параметрів масштабування та зсуву вейвлета a та b з довільними кроками Δa і Δb призводить до суттєвого зростання обчислювальних витрат.

Крім того, у результаті такого підходу формується надмірна кількість коефіцієнтів, яка часто значно перевищує число відліків початкового сигналу. Для якісної реконструкції цієї надлишкової інформації не потрібно, тому обчислення стають неефективними.

Дискретне вейвлет-перетворення (DWT), реалізоване, зокрема, у бібліотеці PyWavelets, дає достатньо інформації як для аналізу сигналу, так і для його синтезу. При цьому DWT характеризується значно меншою кількістю операцій та економним використанням пам’яті. Завдяки використанню ортонормованих базисних функцій та ієрархічній структурі декомпозиції воно дозволяє швидко та ефективно працювати навіть зі складними та нестаціонарними сигналами.

Коли варто застосовувати вейвлет-перетворення замість перетворення Фур’є

Класичне перетворення Фур’є (FT) працює дуже ефективно у тих випадках, коли спектральний склад сигналу є стаціонарним, тобто коли частоти, присутні в сигналі, не змінюються в часі. Якщо сигнал містить певні частоти x Hz, то вони однаково проявляються на всьому його проміжку. У такій ситуації FT забезпечує високу точність.

Однак зі зростанням нестаціонарності результати перетворення Фур’є стають значно гіршими. Це є суттєвим недоліком, оскільки більшість реальних сигналів (акустичних, біомедичних, телекомунікаційних тощо) мають змінний у часі спектр.

Головне обмеження FT полягає у високій частотній, але практично нульовій часовій роздільній здатності. Тобто ми можемо визначити, які частоти присутні у сигналі, але не можемо дізнатися, у який саме момент вони виникають або зникають. Це добре ілюструється на двох прикладах, які можуть бути наведені нижче у тексті.

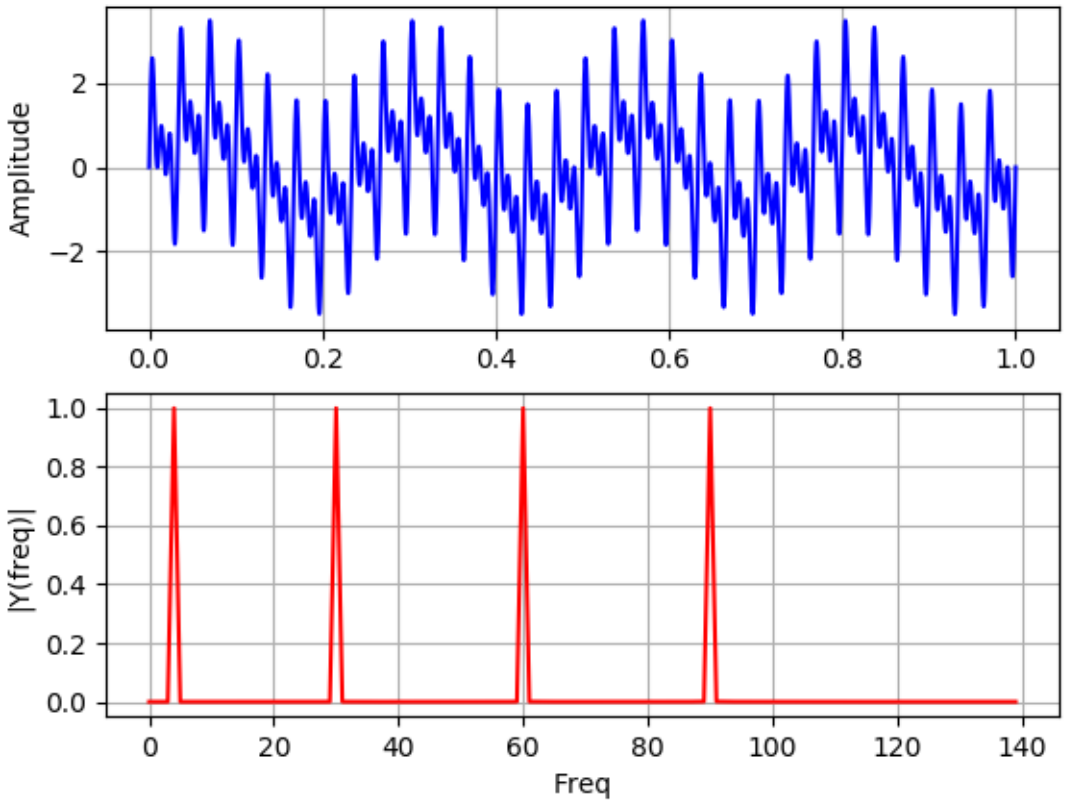


Рис. 1.1 Всі чотири частоти присутні в сигналі протягом усього часу його дії

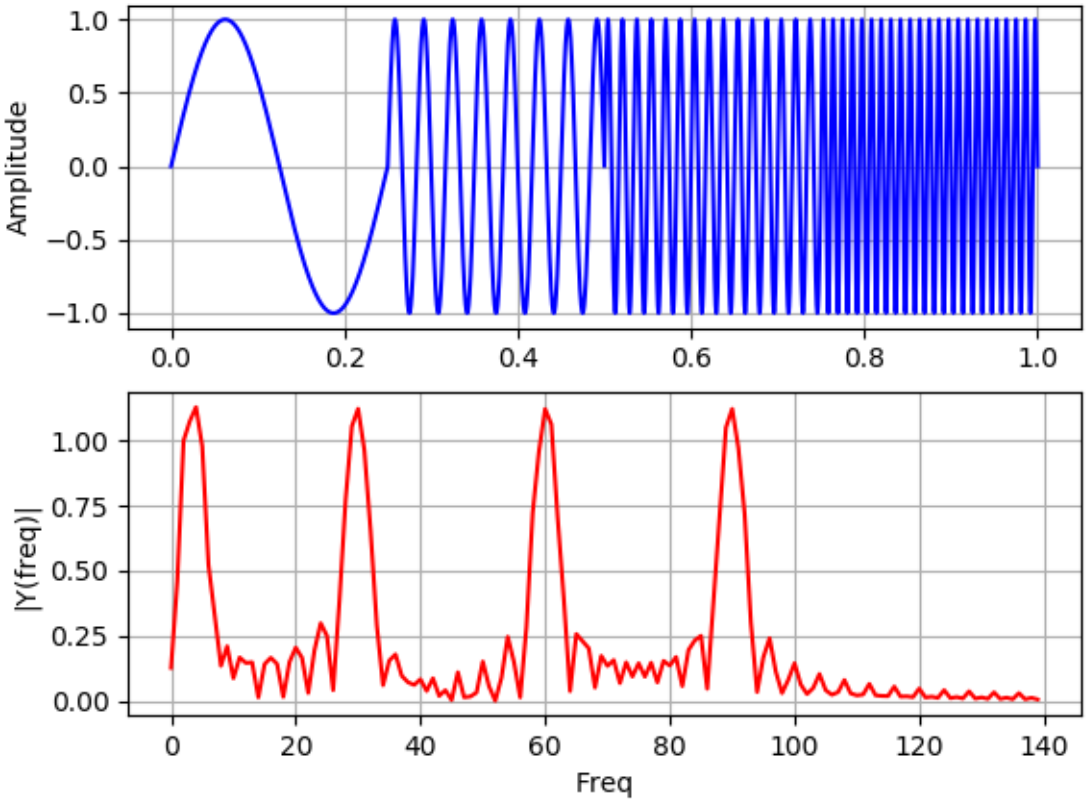


Рис. 1.2 На цьому графіку сигнали не перекриваються у часі, бічні пелюстки зумовлені розривом між чотирма різними частотами

Навіть якщо два сигнали мають однакові частотні спектри та містять ті самі чотири піки, перетворення Фур'є (FT) не здатне визначити, у яких саме ділянках сигналу ці частотні компоненти з'являються. FT показує лише загальний набір частот, але повністю ігнорує їх часову локалізацію.

У випадках, коли спектр сигналу змінюється з часом, тобто коли частоти не є постійними, значно ефективнішим інструментом для аналізу є вейвлет-перетворення (WT). Воно дозволяє одночасно отримати інформацію і про частотний склад сигналу, і про те, коли саме певні частотні компоненти виникають або зникають, що робить WT оптимальним підходом для дослідження нестаціонарних та складних сигналів.

1.2 Режим розширення сигналу

Перед виконанням дискретного вейвлет-перетворення (DWT), яке ґрунтується на використанні банків фільтрів, часто виникає необхідність штучно подовжити сигнал. Це потрібно для коректної обробки його крайових ділянок, оскільки фільтри мають певну довжину та вимагають додаткових значень поза межами вихідного масиву даних.

Залежно від обраного методу екстраполяції (або «розширення» сигналу), на його границях можуть з’являтися помітні артефакти. Вони здатні спотворювати структуру коефіцієнтів та призводити до похибок у результатах DWT. Саме тому правильний вибір способу екстраполяції є важливим для отримання точних і стабільних результатів аналізу.

Бібліотека PyWavelets надає низку різних методів розширення сигналу, кожен із яких по-своєму впливає на крайові значення та дозволяє мінімізувати потенційні спотворення. Для демонстрації роботи цих підходів розглянемо наступний фрагмент коду:

```
import numpy as np
from matplotlib import pyplot as plt
from pywt._doc_utils import boundary_mode_subplot
# synthetic test signal
x = 5 - np.linspace(-1.9, 1.1, 9)**2
# Create a figure with one subplots per boundary mode
fig, axes = plt.subplots(3, 3, figsize=(10, 6))
plt.subplots_adjust(hspace=0.5)
axes = axes.ravel()
boundary_mode_subplot(x, 'symmetric', axes[0], symw=False)
boundary_mode_subplot(x, 'reflect', axes[1], symw=True)
boundary_mode_subplot(x, 'periodic', axes[2], symw=False)
boundary_mode_subplot(x, 'antisymmetric', axes[3], symw=False)
boundary_mode_subplot(x, 'antireflect', axes[4], symw=True)
boundary_mode_subplot(x, 'periodization', axes[5], symw=False)
boundary_mode_subplot(x, 'smooth', axes[6], symw=False)
boundary_mode_subplot(x, 'constant', axes[7], symw=False)
boundary_mode_subplot(x, 'zeros', axes[8], symw=False)
plt.show()
```

Рис. 1.3 фрагмент коду для демонстрації роботи підходів

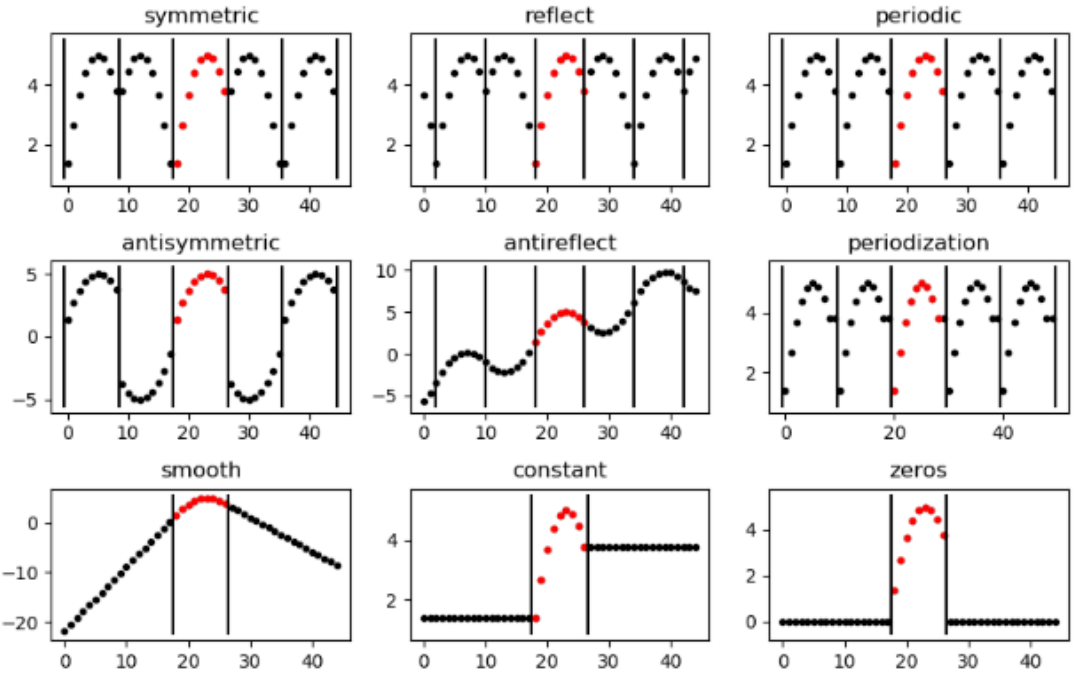


Рис. 1.4 На графіках показано, як короткий (червоний) сигнал розширюється (чорний) за межами своєї початкової протяжності

Дискретне вейвлет-перетворення

Для демонстрації роботи дискретного вейвлет-перетворення (DWT) використаємо сигнал, частотний спектр якого змінюється в часі. Його особливість полягає в тому, що на початку містяться переважно низькочастотні компоненти, тоді як у кінцевій частині з’являються високочастотні, короткохвильові складові. Таким чином, частота сигналу поступово зростає з часом.

Подібна структура сигналу є зручною для аналізу, адже дозволяє наочно простежити, які частотні діапазони були відфільтровані в процесі обробки. Достатньо просто подивитися на часову вісь, щоб зрозуміти, яка частина спектра припадає на певний інтервал сигналу. Це робить такий тестовий приклад особливо корисним для демонстрації можливостей DWT щодо розділення сигналу на масштабні та детальні компоненти.

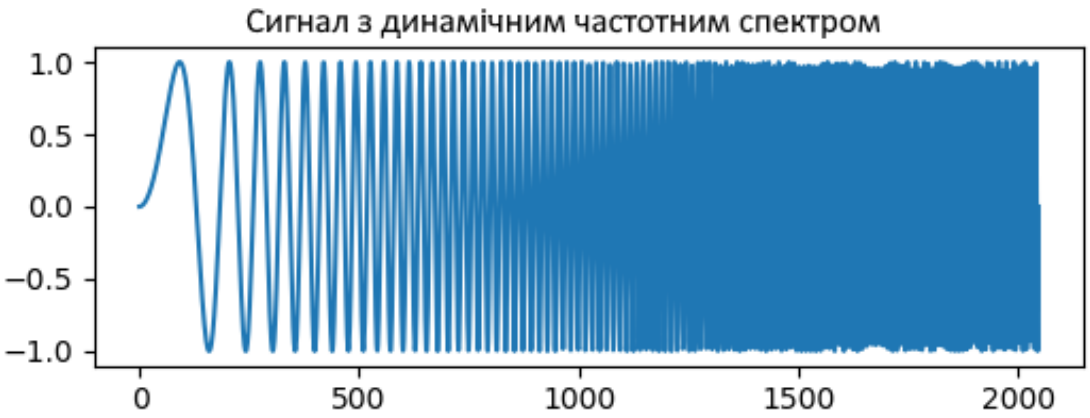


Рис. 1.5 Сигнал з динамічним частотним спектром

У бібліотеці PyWavelets дискретне вейвлет-перетворення (DWT) реалізоване за допомогою функції pywt.dwt(). Ця функція виконує розкладання сигналу на два основні набори коефіцієнтів: апроксимуючі cA та деталізуючі cD, що відповідають першому рівню вейвлет-декомпозиції.

Апроксимуючі коефіцієнти cA відображають загальну, низькочастотну структуру сигналу — його «грубий» масштаб, тоді як деталізуючі коефіцієнти cD містять високочастотні компоненти, які характеризують локальні зміни та деталі. Таке розділення дає змогу одночасно вивчати як глобальну форму сигналу, так і його швидкі коливання.

Функція pywt.dwt() застосовується до сигналу, поданого у вигляді одномірного масиву (вектора), і повертає обидва набори коефіцієнтів, що робить її зручною для подальшого аналізу, фільтрації або побудови багаторівневої декомпозиції.

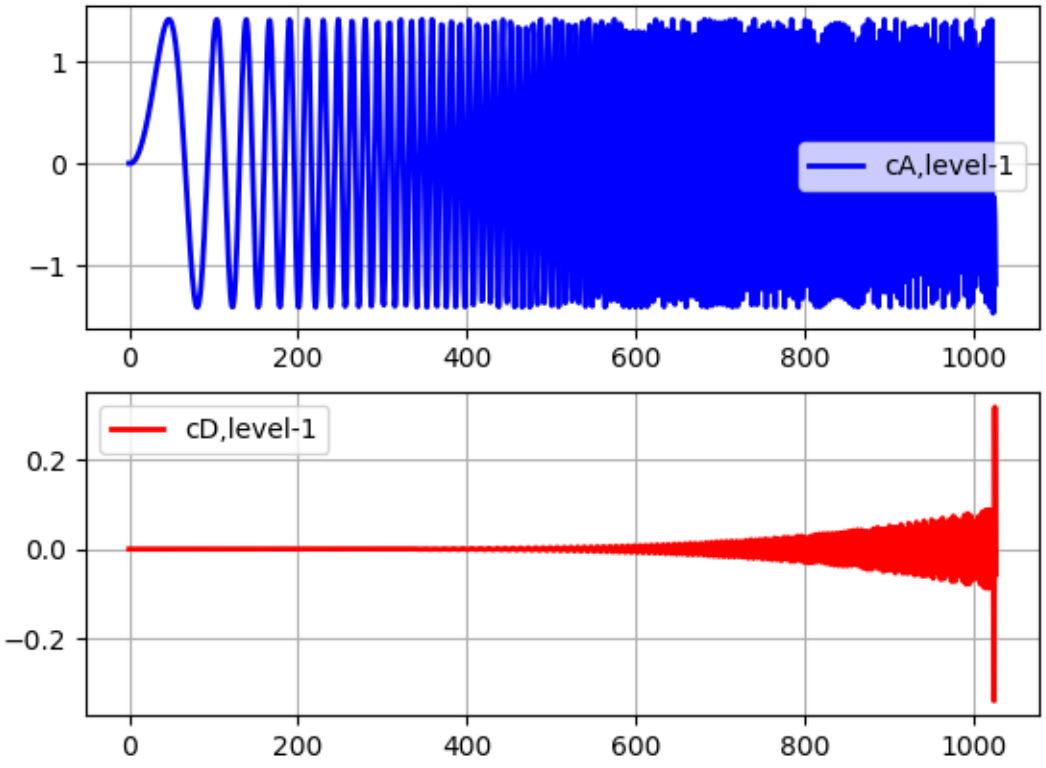


Рис. 1.6 Перший рівень перетворення

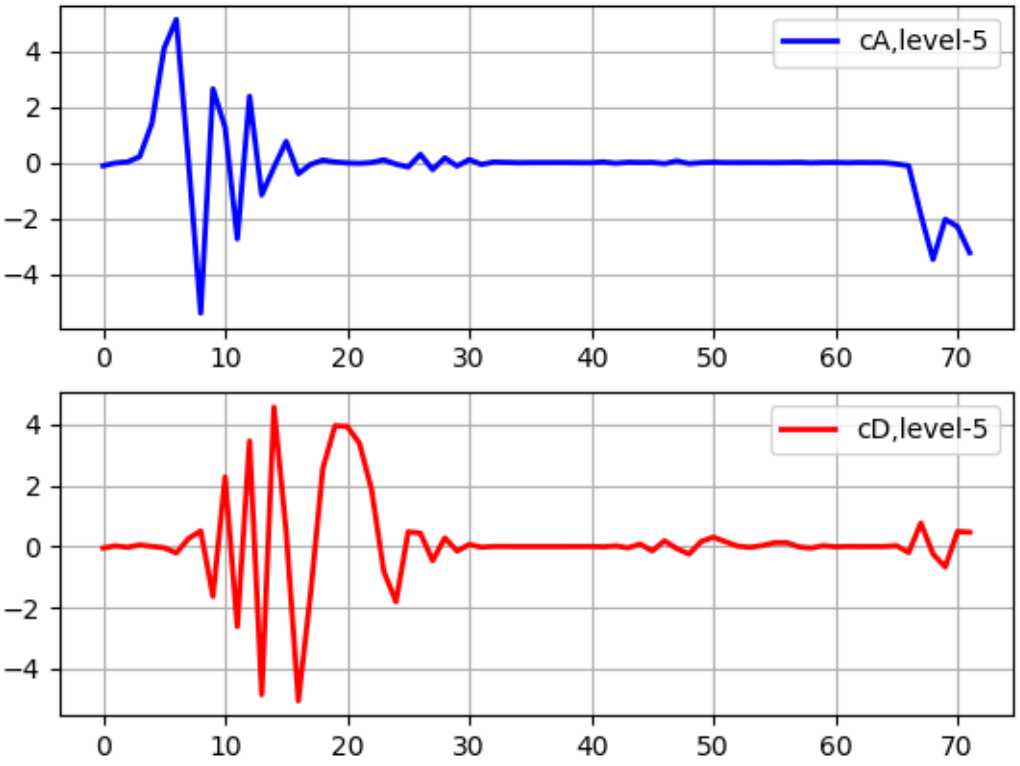


Рис. 1.7 П’ятий рівень перетворення

Апроксимуючі коефіцієнти (cA) відповідають виходу низькочастотного фільтра (low-pass filter), який у контексті DWT виконує роль фільтра усереднення. Саме ці коефіцієнти відображають повільні зміни сигналу та формують його загальну, «грубу» структуру.

Деталізуючі коефіцієнти (cD), навпаки, є результатом проходження сигналу через високочастотний фільтр (high-pass filter), або ж різницевий фільтр. Вони містять інформацію про швидкі коливання, різкі переходи та локальні особливості сигналу.

Для отримання коефіцієнтів не лише першого, а й будь-якого вищого рівня декомпозиції можна скористатися функцією pywt.wavedec(). Вона приймає на вхід вихідний сигнал та бажаний рівень розкладу й повертає один набір апроксимуючих коефіцієнтів найвищого (n-го) рівня та n наборів деталізуючих коефіцієнтів — від рівня 1 до рівня n.

Наприклад, під час декомпозиції сигналу до п’ятого рівня функція pywt.wavedec() поверне набір cA5 і п’ять наборів cD1, cD2, cD3, cD4, cD5, що дозволяє аналізувати структуру сигналу на різних масштабах, від найширших до найдрібніших деталей.

```
from pywt import wavedec
from pylab import *
from numpy import *
x = linspace (0,  1,  num = 2048)
y = sin (250  *  pi * x**2)
st='sym5'
coeffs = wavedec(y, st, level=5)
subplot(2, 1, 1)
plot(coeffs[0],'b',linewidth=2, label='cA,level-5')
grid()
legend(loc='best')
subplot(2, 1, 2)
plot(coeffs[1],'r',linewidth=2, label='cD,level-5')
grid()
legend(loc='best')
show()
```

Рис. 1.8 Приклад для п’ятого рівня

1.3 Використання банку фільтрів

Частину питань, пов’язаних із рівнями вейвлет-перетворення, було розглянуто у попередньому підрозділі. Однак важливо підкреслити, що дискретне вейвлет-перетворення (DWT) завжди реалізується у вигляді банку фільтрів — каскаду високочастотних (high-pass) і низькочастотних (low-pass) фільтрів. Саме така структура є одним із найбільш ефективних способів поділу сигналу на декілька частотних піддіапазонів, що дозволяє вивчати його поведінку на різних масштабах.

На першому кроці аналіз виконується для найменшого масштабу, тобто для найвищих частот. На цьому етапі ми вловлюємо швидкі коливання сигналу та його високочастотні компоненти. Далі, на другому етапі, масштаб збільшується у два рази (що еквівалентно зменшенню частоти вдвічі), і аналіз переноситься у діапазон

приблизно половини максимальної частоти сигналу. На третьому рівні масштаб зростає вже у чотири рази, що дозволяє досліджувати компоненти, розташовані приблизно в області однієї четвертої від максимальної частоти. Такий підхід продовжується доти, доки ми не досягнемо максимально можливого рівня розкладу для даного сигналу.

Максимальний рівень декомпозиції напряму залежить від довжини сигналу та характеристик вибраного вейвлета. Його можна визначити за допомогою функції pywt.wavedec(), яка автоматично виконує покроковий розклад. У результаті ми отримуємо структуру, що складається з одного набору апроксимуючих коефіцієнтів найвищого рівня та декількох наборів деталізуючих коефіцієнтів — по одному для кожного рівня розкладу.

Схематично така декомпозиція може мати вигляд:

$$[cA_n, cD_n, cD_\{n-1\}, \ldots, cD_2, cD_1]$$

де n — максимальний рівень, доступний для даного сигналу.

Отримаємо:

Максимальний рівень декомпозиції: 7

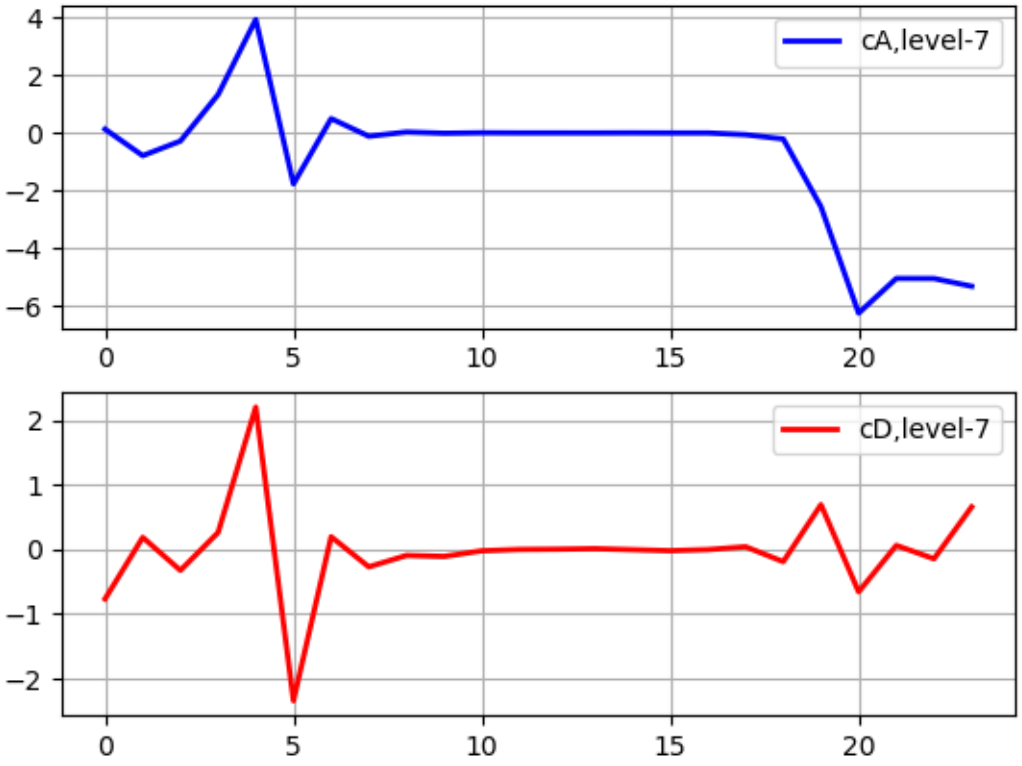


Рис. 1.9 Декомпозиція для 7 рівня

Декомпозиція вейвлет-перетворення припиняється в той момент, коли довжина сигналу виявляється меншою за довжину фільтра, що застосовується для обраного вейвлета, наприклад sym5. Це обмеження є природним і цілком логічним з точки зору теорії сигналів, адже кожен фільтр повинен мати достатню кількість вхідних даних для того, щоб обчислити коефіцієнти на певному рівні зі збереженням точності та коректності. Якщо кількість доступних точок сигналу зменшується до критичного рівня, подальша декомпозиція втрачає математичний сенс і не може бути виконана без значних похибок або штучного доповнення даних, що знижує якість результатів.

Щоб проілюструвати принцип роботи багаторівневої декомпозиції, розглянемо умовний, але показовий приклад. Припустимо, що маємо сигнал, у спектрі якого присутні частоти до 1000 Гц. На першому етапі роботи дискретного вейвлет-перетворення (DWT) відбувається поділ цього сигналу на два базові частотні діапазони: низькочастотний (0–500 Гц) та високочастотний (500–1000 Гц).

Такий поділ відповідає принципу роботи банків фільтрів, де низькочастотний (low-pass) фільтр виділяє плавні низькочастотні компоненти, а високочастотний (high-pass) виявляє швидкі зміни та різкі коливання сигналу.

Для наступного рівня аналізу береться лише апроксимуюча (низькочастотна) частина першого кроку, тобто діапазон 0–500 Гц. Саме ця частина містить більшу енергію та основну структуру сигналу, а отже є придатною для подальшого поділу. На другому рівні діапазон 0–500 Гц знову поділяється навпіл, що формує частотні зони 0–250 Гц і 250–500 Гц. Такий поділ дозволяє зануритися на наступний масштаб аналізу, виявити повільніші компоненти та відокремити їх від середньочастотних елементів.

На третьому рівні процес повторюється: частина 0–250 Гц проходить через аналогічний набір фільтрів і розділяється на піддіапазони 0–125 Гц та 125–250 Гц. Важливо розуміти, що кожен наступний рівень зменшує робочий частотний діапазон удвічі, але при цьому збільшує масштаб аналізу, що дозволяє отримати більш глобальне уявлення про структуру сигналу. Таким чином, із переходом до кожного нового рівня ми ніби «віддаляємося» від сигналу, розглядаючи його глобальні властивості на ширших часових масштабах і одночасно отримуючи деталі на попередніх рівнях.

Процес декомпозиції триває до того моменту, коли довжина сигналу, що потрапляє на вхід фільтрів на черговому рівні, стає меншою за довжину вейвлет-фільтра. У таких умовах виконання коректної фільтрації вже неможливе — саме тому рівень декомпозиції є кінцевим і залежить від довжини вхідного сигналу та параметрів конкретного вейвлета. У результаті формується повноцінна багаторівнева вейвлет-структура, яка містить як глобальну поведінку сигналу (у наборі апроксимуючих коефіцієнтів), так і локальні швидкі зміни, представлені набором деталізуючих коефіцієнтів для кожного рівня.

Значення багаторівневої вейвлет-декомпозиції у моніторингу мережевого трафіку

Метод багаторівневої декомпозиції має вирішальне значення для аналізу, класифікації та моніторингу мережевого трафіку. Мережеві потоки за своєю

природою є складними, багатокомпонентними та нелінійними. Традиційні методи спектрального аналізу, наприклад перетворення Фур'є, відображають лише частотний склад, але повністю втрачають часову локалізацію. Натомість вейвлет-перетворення дозволяє поєднати частотну та часову інформацію, що є ключовим для виявлення атак, нестандартної активності та структурних змін.

1. Виявлення аномалій на різних масштабах

Мережевий трафік має складну структуру, що складається з компонентів різної природи:

- повільні, довготривалі зміни (наприклад, добові та тижневі коливання навантаження на мережу),
- середньочастотні процеси (звичайна робоча активність користувачів, взаємодія сервісів),
- високочастотні, раптові сплески, які часто мають аномальний характер (сканування портів, DoS/DDoS-атаки, флудинг, різкі піки запитів).

Вейвлет-декомпозиція дає змогу ізолювати кожну з цих складових окремо, аналізуючи їх незалежно. Завдяки цьому можливо:

- побачити приховані патерни, які не видно у сирому трафіку,
- відфільтрувати фоновий шум,
- автоматично виявляти короткі, але небезпечні сплески,
- відстежувати зміну інтенсивності трафіку у довгострокових масштабах.

Це робить вейвлети надзвичайно потужним інструментом для систем IDS/IPS та аналітики безпеки.

2. Локалізація подій у часі

На відміну від перетворення Фур'є, яке показує лише спектральний склад трафіку, дискретне вейвлет-перетворення (DWT) дозволяє визначити точний час виникнення тієї чи іншої події. Це критично важливо, оскільки:

- атаки часто тривають частки секунди,
- різкі зміни трафіку можуть вказувати на експлуатацію вразливостей або початок несанкціонованого доступу,

- необхідно точно визначити інтервал часу, у якому активність відхилилася від норми.

Завдяки часовій локалізованості вейвлет-аналіз дозволяє одразу ідентифікувати небезпечну активність та приймати оперативні рішення — наприклад, блокування підозрілих IP, фільтрацію пакетів або активацію режиму захисту.

3. Висока ефективність обчислень (реальний час)

Банки фільтрів DWT працюють за принципом швидких згорток та потребують значно менше ресурсів, ніж інші методи спектрального аналізу. Це забезпечує можливість застосування вейвлет-декомпозиції:

- у системах реального часу,
- на мережевому обладнанні з обмеженою обчислювальною потужністю (маршрутизатори, комутатори, IoT),
- в аналітичних системах, які обробляють великий обсяг потокових даних (наприклад, системи DPI).

Завдяки своїй швидкості вейвлети дозволяють будувати алгоритми, що реагують на зміни у трафіку практично миттєво.

4. Стиснення, відбір та фільтрація даних

Апроксимуючі коефіцієнти зберігають основну структуру та поведінку трафіку, тоді як деталізуючі містять лише відхилення від нормального стану, тобто локальні «аномалії». Це дозволяє:

- значно зменшити обсяг даних, що потрібно зберігати або передавати,
- відокремити значущу інформацію від шуму,
- створювати легкі моделі машинного навчання для класифікації трафіку,
- будувати системи виявлення атак на основі лише невеликої частини коефіцієнтів.

Завдяки такій комбінації компресії та селекції інформації вейвлети активно використовуються у мережевих системах моніторингу, де важливо швидко й ефективно обробляти великі масиви даних.

2 ПОНЯТТЯ ДЗЕРКАЛЕННЯ ТРАФІКА

2.1 Технології TAP та SPAN

Повна видимість мережі є фундаментальною умовою для якісного моніторингу, ефективного управління інфраструктурою та забезпечення інформаційної безпеки. Якщо адміністратор або система безпеки не можуть бачити події, що відбуваються в мережі, — наприклад, спроби мережевої атаки, несанкціонований доступ, аномальну активність користувачів, надмірне навантаження або неефективність роботи окремих застосунків — то виявлення й оперативне вирішення таких проблем стає практично неможливим. Неможливо захищати те, що не видно, а отже контроль над трафіком і його детальний аналіз є ключовими складовими сучасної кібербезпеки.

Сьогоднішні комп’ютерні мережі стають усе більшими за масштабом, складнішими за структурою та інтенсивнішими за обсягом переданих даних. Об’єми інформації, що циркулюють у мережах, зростають експоненціально, причому швидкості передавання давно перевищили ті, що існували ще два десятиліття тому. Якщо раніше домінували канали 10 Мбіт/с, то тепер стандартом стали повнодуплексні лінії на 1, 10, 40 і навіть 100 Гбіт/с. Потоки даних зросли настільки, що звичайні методи контролю вже не забезпечують достатнього рівня деталізації.

У сучасних умовах настала ера тотального контролю даних, глибокого інспектування пакетів (DPI), суворого дотримання вимог законодавства, корпоративних політик, проведення регулярного аудиту та навіть виконання процедур законного перехоплення. Усі ці процеси вимагають не просто часткового аналізу окремих фрагментів трафіку, а повного відстеження всіх даних, що проходять мережею. Саме це робить доступ до трафіку на рівні пакетів одночасно необхідним і складним завданням, особливо з огляду на те, що сучасні кіберзагрози стають дедалі більш витонченими, швидкими та непередбачуваними.

Таким чином, повна видимість мережі є критично важливою, майже такою ж необхідною, як стабільний доступ до самої мережі. Вона забезпечує можливість:

- оперативного реагування на інциденти безпеки;
- виявлення прихованих атак, які не проявляються на поверхневому рівні;
- детального аналізу продуктивності та роботи застосунків;
- проведення діагностики несправностей і оптимізації мережевих процесів;
- забезпечення прозорості мережевої діяльності в рамках аудиту та контролю.

Доступ до потокових даних аж до рівня окремих пакетів — це перший і найважливіший крок до встановлення повної видимості мережі. Жодна інша технологія не забезпечує настільки точного, повного та достовірного уявлення про те, що насправді відбувається у мережі. Саме пакети містять найдрібніші, але найважливіші відомості, включно з заголовками, службовими прапорцями, метаданими, контрольними сумами та іншими полями, що дозволяють точно визначати характер трафіку й виявляти загрози.

Для отримання доступу до такого рівня інформації найчастіше використовують два технологічні підходи: SPAN (копіювання трафіку на портах комутатора) та TAP (апаратні пристрої для фізичного перехоплення трафіку). Обидві технології активно застосовуються у мережах різного масштабу, проте існують суттєві відмінності в їх надійності, точності та стабільності роботи.

У цьому розділі буде проведено комплексний аналіз особливостей отримання доступу до трафіку за допомогою технологій TAP і SPAN. Будуть розглянуті технічні аспекти їх роботи, фундаментальні принципи, ключові переваги й обмеження, а також типові помилки та поширені міфи, пов’язані з використанням SPAN-портів і віддзеркаленням трафіку через комутатори. Окрему увагу буде приділено тому, чому в більшості практичних сценаріїв саме TAP є єдиним надійним і життєздатним способом отримання повної, достовірної та незміненої копії трафіку. У переважній більшості випадків — близько 99% — саме TAP забезпечує найвищий рівень точності, стабільності та відсутності втрат даних у процесі моніторингу мережевого трафіку.

Технології TAP (*Test Access Point*) та SPAN (*Switch Port Analyzer*, також відома як *Switch Port for Analysis*) дозволяють отримати прямий доступ до реальних мережевих пакетів, які передаються інфраструктурою. Саме ці інструменти забезпечують можливість глибокого аналізу трафіку, контролю безпеки, діагностики несправностей і проведення аудиту. Однак у сучасних складних мережах постає логічне запитання: який підхід є оптимальним? Якщо обидві технології на перший погляд виконують одну й ту ж функцію, то в яких випадках доцільніше використовувати TAP, а в яких — SPAN? У цьому й полягає головна дилема.

Щоб зрозуміти різницю між цими методами, варто пригадати історичний розвиток мережевої інфраструктури. До початку 1990-х років більшість аналізаторів локальних мереж підключалися безпосередньо до фізичного середовища передачі — до кабелю чи сегмента мережі. Такий підхід був простим, але вимагав втручання у мережеву топологію та створював ризики для її стабільної роботи.

Згодом, зі зростанням популярності хабів (концентраторів), аналіз мережевого трафіку суттєво спростився. Хаби працювали в режимі широкомовної передачі: кожен пакет, який надходив до одного порту, автоматично копіювався на всі інші порти. Завдяки цьому достатньо було під’єднати систему моніторингу до будь-якого порту хаба, і вона отримувала повну копію всіх пакетів, що передавалися мережею. Це був перший практичний варіант «пасивної видимості».

Однак зі зростанням навантаження, вимог до продуктивності та переходом до комутованих мереж хаби перестали відповідати потребам сучасних архітектур. Комутатори й маршрутизатори витіснили їх завдяки своїй ефективності, безпеці та здатності сегментувати трафік. У результаті виникла проблема: підключитися до мережі для спостереження стало значно складніше, оскільки трафік уже не транслювався на всі порти.

Саме тоді на комутаторах з’явилася технологія, яка згодом отримала назву SPAN. Вона дозволяла створювати «дзеркальний» порт, на який комутатор копіює пакети з одного чи кількох робочих портів. Таким чином зникла необхідність

фізично врізатися у мережу або використовувати широкомовні пристрої — достатньо було створити SPAN-сесію й направити копії даних до інструментів аналізу та безпеки.

Згідно з концепцією, SPAN-порт — це спеціальний порт комутатора, який використовується лише для отримання копій трафіку. Він не повинен передавати трафік назад у мережу (двосторонній обмін заборонено апаратно), що гарантує відсутність ризику ненавмисного втручання або генерації небажаних пакетів. У більшості випадків SPAN підтримує:

- дзеркалювання одного або кількох портів,
- копіювання трафіку з VLAN,
- вибірковий захват вхідного чи вихідного трафіку,
- фільтрацію за певними параметрами.

Втім, важливо розуміти, що SPAN не був створений як інструмент кібербезпеки або постійного моніторингу. Його оригінальне призначення — це засіб тестування, діагностики та відлагодження мережевого обладнання виробниками. Технологія задумана для тимчасового технічного аналізу, а не для тривалої високонавантаженої роботи. І лише пізніше SPAN почали використовувати для віддзеркалення трафіку в системах безпеки, IDS/IPS, DPI та мережевому моніторингу.

Однак така «похідна» природа SPAN має низку суттєвих технічних обмежень:

- можливі втрати кадрів при високому навантаженні,
- зміна часових характеристик трафіку,
- неідеальна точність копіювання,
- залежність від завантаженості комутатора,
- вплив функцій QoS, ACL, шейперів тощо на дзеркальний трафік.

Саме тому у професійних мережах для точного аналізу, незалежного збору даних і глибокого контролю зазвичай застосовують альтернативну технологію — апаратні TAP-пристрої, які забезпечують «чисту» копію трафіку без втручання у роботу мережі й без ризику модифікації пакетів.

Основні міркування щодо вимог до забезпечення повної видимості мережі

Під час проєктування або модернізації будь-якої мережевої інфраструктури існує значна кількість чинників, які впливають на вибір архітектури, стратегії моніторингу та способів доступу до трафіку. Повна видимість мережі — це не просто бажана функція, а ключовий елемент безпеки, відмовостійкості та ефективного управління. Щоб прийняти зважене рішення щодо методів отримання трафіку для аналізу, необхідно врахувати низку критично важливих аспектів. Нижче наведено узагальнений, але вичерпний чек-лист, який допоможе оцінити всі переваги та недоліки різних технологій віддзеркалення даних.

1. Вплив активних пристроїв на структуру Ethernet-кадру

Будь-яке активне мережеве обладнання — зокрема комутатори, маршрутизатори та мережеві шлюзи — неминуче вносить зміни у часові характеристики Ethernet-кадрів. Навіть якщо зміни мінімальні, наприклад додавання затримки або корекція синхронізації, це вже порушує оригінальну часову структуру трафіку. Єдиним пристроєм, який забезпечує абсолютно пасивний, точний і немодифікований доступ до потоку даних, є фізичний, окремо встановлений мережевий TAP (*Test Access Point*). На відміну від активних компонентів, TAP працює поза площиною комутації та не змінює сам трафік.

2. Лінійність затримок — критичний фактор достовірності

Під час використання дзеркального порту або програмного дублювання трафіку важливо, щоб усі накладені затримки були лінійними та передбачуваними. Наприклад, якщо під час копіювання кадр отримує затримку 5 мс, то ця затримка має бути однаковою для всіх кадрів, незалежно від їх розміру, типу або послідовності. Нелінійні затримки спотворюють хронологію подій і роблять неможливим виконання точного моніторингу, реального таймінгу транзакцій або криміналістичного аналізу. Технологія SPAN є прикладом інструмента, який створює нерівномірні затримки, а отже не забезпечує достовірну часову прив’язку даних. Це суттєво ускладнює або навіть унеможливлює автентичну аналітику на основі SPAN-порту.

3. TAP — єдиний пристрій, що копіює трафік без втрат

Фізичний TAP здатний передавати абсолютно всі дані, що проходять через мережу:

- кожен біт і байт,
- міжкадрові інтервали (IFG),
- некоректні, пошкоджені або «биті» пакети,
- надмірно великі або надто маленькі кадри,
- нестандартні, нестабільні чи малопоширені типи трафіку.

Існує окрема дискусія щодо того, чи потрібно передавати пошкоджені пакети для аналізу. Але з точки зору мережевої діагностики та кібербезпеки це не просто корисно — це обов'язково, адже такі пакети часто сигналізують про:

- початок апаратного збою,
- порушення цілісності каналу,
- спроби атак типу fuzzing,
- потенційне перехоплення або модифікацію трафіку,
- проблеми з буферами або перевантаження мережевих інтерфейсів.

Саме тому автономні TAP-пристрої вважаються значно надійнішими, ніж вбудовані функції дзеркалювання на комутаторах.

4. Необхідність попереднього тестування технологій дзеркалювання

Перш ніж впроваджувати будь-який спосіб доступу до трафіку, слід виконати ряд тестів:

1. **Перевірити кілька механізмів дзеркалювання**, щоб упевнитися, що отримані дані відповідають реальному трафіку та коректно обробляються інструментами моніторингу.
2. **Здійснити вимірювання до і після застосування технології віддзеркалення**, створивши базову лінію для порівняння. Це дозволяє оцінити, як саме спосіб копіювання впливає на кадри та чи утворює він втрати, затримки або нерівномірні зсуви.
3. **Використовувати виключно якісне обладнання** перевірених виробників. Дешеві або несертифіковані TAP-пристрої можуть створювати додаткові помилки, втрачати трафік або навіть тимчасово порушувати роботу мережі.

Дотримання цих рекомендацій дозволяє забезпечити надійність моніторингу та уникнути неочевидних проблем у майбутньому.

5. Юридичні аспекти та криміналістична достовірність даних

Окремо слід наголосити на юридичній значущості методів віддзеркалення трафіку. Дані, зібрані для розслідування інцидентів (Network Forensics), можуть бути використані як докази як у внутрішніх компанійних розслідуваннях, так і в судових справах — адміністративних, цивільних або кримінальних.

У такому випадку принципово важливими є такі характеристики:

- незмінність даних,
- точність часової прив’язки,
- повнота зібраних пакетів,
- відсутність втручання в трафік.

Технологія SPAN не гарантує жодного з цих параметрів, оскільки є активною функцією комутатора та може спотворювати або частково втрачати дані через навантаження. Тому з правової точки зору дані, отримані зі SPAN-порту, можуть бути поставлені під сумнів, оскільки не є криміналістично надійними.

На відміну від цього, мережевий TAP формує точну, немодифіковану копію трафіку з гарантованою часовою синхронізацією. Саме TAP забезпечує доказовість, яку неможливо оскаржити, і відповідає вимогам, що висуваються до юридично значущих даних у галузі цифрової криміналістики.

2.2 Дзеркалення трафіку та аналіз локальної мережі

Мережевий TAP: принцип роботи та роль у забезпеченні повної видимості мережі

Мережевий TAP (*Test Access Point*) — це спеціалізований пристрій, який безпосередньо інтегрується у фізичну кабельну інфраструктуру мережі та забезпечує отримання точної копії трафіку, що проходить між двома мережевими вузлами. TAP встановлюється послідовно, тобто розміщується між двома

пристроями комунікації, і весь обмін даними між ними неминуче проходить через цей пристрій.

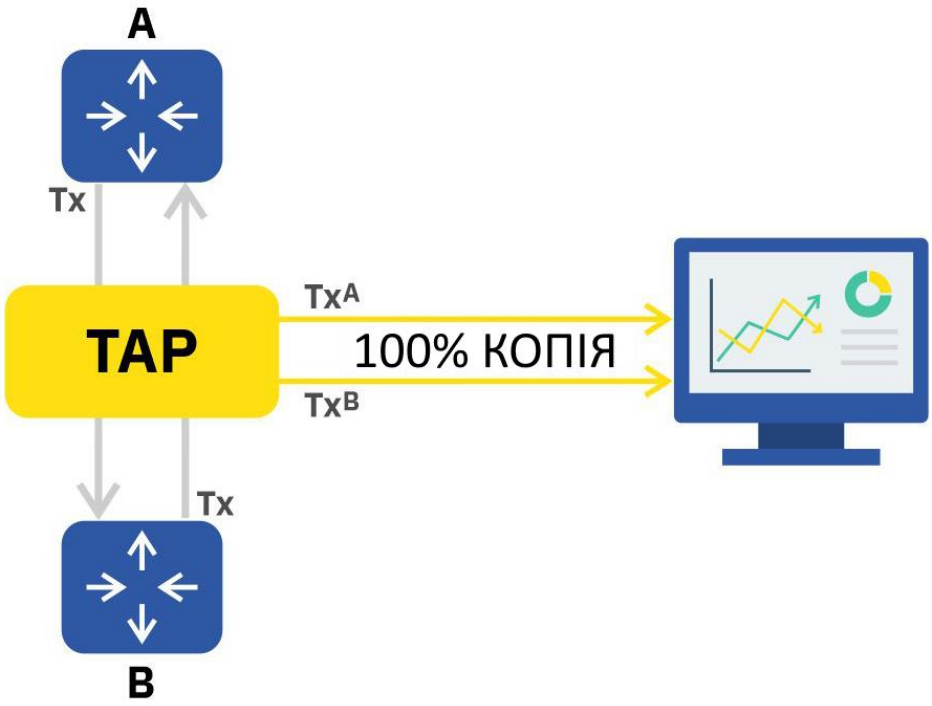


Рис. 2.1 Test Access Point

Усередині TAP використовується фізичний оптичний дільник (для волоконно-оптичних ліній) або логічний електричний дільник (для мідних Ethernet-ліній). Завдяки цьому TAP створює ідентичну копію всього трафіку, що проходить між пристроями A та B, і передає її на окремі порти моніторингу. Водночас вихідний трафік не змінюється й продовжує рухатися мережею без будь-яких затримок, втручань або змін у структурі пакетів.

Більшість TAP-пристроїв реалізують поділ трафіку на два незалежні потоки:

- **TxA** — копія даних, які надсилає пристрій A,
- **TxB** — копія даних, які надсилає пристрій B.

Таким чином, системи моніторингу отримують повний двонаправлений трафік у чистому, немодифікованому вигляді.

100% дзеркальне копіювання — ключова перевага TAP

Основна перевага TAP полягає в тому, що він забезпечує стовідсоткове дублювання кожного пакета, незалежно від:

- його розміру,

- типу чи протоколу,
- коректності або наявності помилок,
- часових інтервалів між кадрами,
- навантаження на лінію зв’язку.

TAP передає абсолютно все — від коректних кадрів Ethernet до пошкоджених або нестандартних пакетів, що мають діагностичну цінність. Завдяки цьому TAP гарантує повну видимість мережі на фізичному рівні, що є критично важливим як для безпеки, так і для контролю продуктивності.

Крім того, TAP повністю виключає можливість перепідписки (oversubscription). На відміну від SPAN, який може пропустити частину кадрів під навантаженням, TAP не працює у режимі переключення чи буферизації, тому не здатний втратити пакети — він дублює трафік апаратно, без участі логіки комутатора.

Щойно трафік проходить через TAP, він негайно копіюється та стає доступним системам:

- моніторингу продуктивності (NPM/NPMD),
- аналізу безпеки (IDS/IPS, SIEM, DPI),
- мережевої криміналістики (Network Forensics),
- аудиту й діагностики.

Таким чином TAP стає центральним елементом інфраструктури видимості мережі: саме він забезпечує надійну точку доступу до даних, які неможливо підробити чи змінити.

Підключення TAP: вимоги та особливості

Важливим нюансом є те, що під час встановлення TAP потрібно короткочасно розірвати існуюче фізичне з’єднання, щоб розмістити пристрій між двома кінцями лінії. Через це TAP зазвичай встановлюють під час:

- регламентних робіт,
- планових технічних перерв,
- модернізації мережевого сегмента.

У правильно спроєктованій інфраструктурі TAP стає постійним, надійним і абсолютно прозорим елементом, який не впливає на роботу мережі, але надає критично важливу інформацію для моніторингу.

Технологія SPAN для дзеркалювання мережевого трафіку

SPAN (*Switch Port Analyzer*) — це програмна функція, вбудована у комутатори та маршрутизатори, яка створює копію певних пакетів, що проходять через пристрій, і перенаправляє їх на спеціально визначений SPAN-порт. Завдяки програмній природі SPAN адміністратор може досить гнучко налаштовувати, який саме трафік має бути віддзеркалений: вхідний, вихідний, лише певні VLAN або конкретні порти.

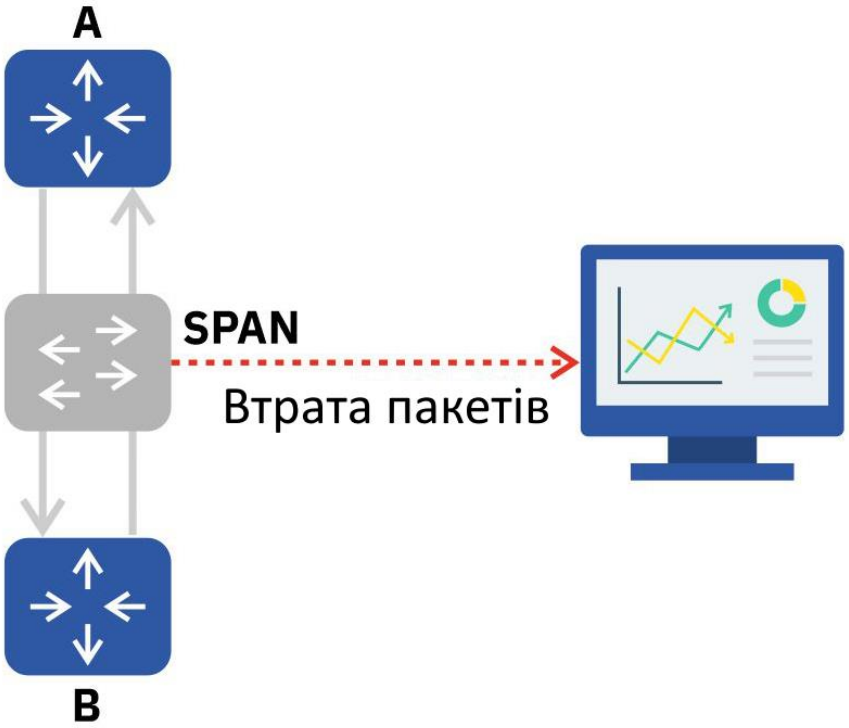


Рис. 2.2 Switch Port Analyzer

На перший погляд здається, що SPAN — зручний і універсальний інструмент. Але для глибшого розуміння можливостей і обмежень цієї технології необхідно відповісти на три ключові запитання.

1. Чи є SPAN пасивною технологією?

Нерідко SPAN-порт помилково називають «пассивним методом доступу до даних», маючи на увазі, що він нібито не впливає на сам трафік. Але це неправильне розуміння терміну «пассивний». У технічному сенсі пасивним можна назвати лише

той пристрій, який ніяким чином не змінює роботу мережі, таймінг пакетів або їх структуру.

SPAN такою технологією не є.

Оскільки SPAN працює на рівні операційної системи комутатора:

- він модифікує таймінг пакетів,
- може затримувати кадри,
- змінює інтервал між пакетами (IFG),
- створює додаткове навантаження на комутатор,
- потенційно зачіпає синхронізацію даних.

Тобто, SPAN впливає на мережевий трафік та роботу комутатора, і тому не може вважатися ні пасивним, ні повністю прозорим рішенням.

2. Чи можна вважати SPAN масштабованою технологією?

На ранніх етапах розвитку мереж, коли пропускна здатність каналів становила 10 Мбіт/с, або навіть 100 Мбіт/с, SPAN більш-менш справлявся зі своїм завданням. При відносно низьких швидкостях комутатор був здатен скопіювати достатню кількість пакетів, і втрати були мінімальними.

Ситуація кардинально змінилася із впровадженням високошвидкісних технологій:

- 1 Гбіт/с,
- 10 Гбіт/с,
- 40 Гбіт/с,
- 100 Гбіт/с і більше.

Сучасний повнодуплексний канал (FDX) працює у два боки одночасно, а отже:

- 1G FDX = 2 Гбіт/с реального трафіку,
- 10G FDX = 20 Гбіт/с,
- 40G FDX = 80 Гбіт/с,
- 100G FDX = 200 Гбіт/с.

Жоден комутатор не здатен одночасно:

- обробляти власні комутаційні операції,
- забезпечувати маршрутизацію, QoS, ACL,
- дублювати трафік на SPAN-порт на 100% без втрат.

На практиці це означає, що:

- SPAN не встигає копіювати трафік,
- при високому навантаженні він пропускає кадри,
- SPAN-порт переповнюється,
- статистика стає некоректною,
- потік даних спотворюється або фрагментується.

Отже, масштабувати SPAN під сучасні навантаження неможливо, і він точно не підходить для аналізу великих трафікових потоків (особливо понад 1 Гбіт/с).

3. Чи працюють RSPAN та ERSPAN як повноцінні технології дзеркалювання?

Технології RSPAN та ERSPAN були розроблені як спроба усунути обмеження SPAN, але на практиці вони додають ще більше проблем.

RSPAN (Remote SPAN)

- дозволяє передавати «дзеркальний» трафік між комутаторами;
- використовує VLAN для транспортування копій пакетів.

Недолік:

трафік SPAN впливає на пропускну здатність усієї мережі, створює додаткове навантаження і легко втрачається на транспортному шляху.

ERSPAN (Encapsulated Remote SPAN)

- віддзеркалений трафік інкапсулюється у GRE;
- може передаватися через L3-мережі.

Недоліки:

- колосальне навантаження на канали (epSPAN може «з'їсти» всю пропускну здатність),
- затримки та зміна структури пакетів через інкапсуляцію,
- повна відсутність гарантій доставки,
- непридатність для точного таймінг-аналізу та криміналістики.

Зокрема, ERSPAN відправляє пакети, які вже пройшли половину мережі, назад до локального комутатора, що в умовах глобальних або розподілених мереж є вкрай неефективним.

Через ці обмеження RSPAN та ERSPAN більше не розглядаються як надійні технології для доступу до трафіка, особливо у випадках, коли потрібна повна видимість, стабільність або юридична достовірність зібраних даних.

Використання SPAN-порту для аналізу локальної мережі

Щоб повною мірою зрозуміти можливості та обмеження технології SPAN, варто звернутися до офіційних документів виробників мережевого обладнання. Так, у своєму технічному звіті (*white paper*) компанія Cisco прямо зазначає, що *“трафік SPAN обробляється комутатором з нижчим пріоритетом у порівнянні зі звичайним трафіком порт-порт”*. Іншими словами, коли мережевий комутатор або маршрутизатор опиняється під високим навантаженням і повинен обрати між передачею основного робочого трафіку та трафіку SPAN, пріоритет завжди матиме користувацький трафік, а дзеркальні копії можуть без попередження відкидатись.

Цей принцип стосується будь-якої конфігурації з використанням SPAN. Наприклад, якщо використовується технологія RSPAN, яка пересилає дзеркальний трафік через інтерфейси ISL, то “законний” трафік у будь-якому випадку отримає пріоритет над RSPAN. Якщо пропускної здатності каналу недостатньо, комутатор із високою ймовірністю почне відкидати SPAN-кадри.

SPAN-порт відкидає трафік під навантаженням: яку стратегію мали б застосувати адміністратори?

У тому ж звіті Cisco зазначає:

“Оптимальним рішенням є використання SPAN лише в умовах низького завантаження портів та з обережністю при високій інтенсивності трафіку.”

Іншими словами:

- SPAN НЕ можна використовувати як основний інструмент моніторингу,
- SPAN НЕ можна застосовувати для аналізу високошвидкісних каналів,
- SPAN НЕ гарантує отримання повної копії трафіку,
- SPAN створює «сліпі зони» для систем моніторингу та ІБ,
- SPAN схильний до втрат кадрів, які можуть бути критичними.

Тому при використанні комутатора для дзеркалювання трафіку потрібно пам’ятати: SPAN не є відмовостійкою технологією і легко стає точкою збоїв для систем аналізу, управління та безпеки.

Коли використання SPAN є прийнятним?

Хоч TAP-пристрої у більшості сценаріїв значно кращі, все ж існують ситуації, коли застосування SPAN є доцільним або навіть необхідним. Це стосується специфічних мережевих умов або обмежень інфраструктури.

1. Моніторинг подій прикладного рівня

Деякі платформи аналізу трафіку працюють виключно з високорівневими даними — наприклад:

- статистикою сеансів,
- потоками додатків (NetFlow, sFlow),
- параметрами взаємодії клієнт-сервер,
- короткими вибірками трафіку.

У таких випадках втрата окремих кадрів не критична, а тому SPAN може забезпечити достатню якість даних для побудови звітів і діагностики. Такі системи не вимагають 100% точності або збереження всіх міжкадрових інтервалів.

2. Віддалені майданчики з рідкісними подіями

SPAN може бути використаний як тимчасове рішення, коли:

- немає економічної доцільності ставити TAP,
- плановий аналіз проводиться рідко,
- потрібно швидко проаналізувати підозрілу активність.

3. Мережі з обмеженим оптичним бюджетом

У волоконно-оптичних мережах інколи кожен дБ затухання має значення, і встановлення TAP може призвести до перевищення допустимого рівня втрат. У таких випадках SPAN може слугувати альтернативою.

4. Відсутність вікна технічного обслуговування

TAP вимагає фізичного розриву каналу для встановлення. Якщо мережа:

- критична,
- круглодобова,

- не допускає простою,

то єдиним реальним варіантом може тимчасово стати SPAN.

5. Необхідність доступу до трафіку, який не покидає комутатор

Існує багато сценаріїв, коли трафік ніколи не доходить до фізичного інтерфейсу, наприклад:

- локальні VLAN,
- внутрішні маршрутизації L3,
- міжпортові внутрішні процеси.

У таких випадках TAP просто не може фізично перехопити пакети, а SPAN — може.

Порівняльне резюме: TAP проти SPAN

TAP — високоточний метод повної видимості

- Створює 100% копію двонаправленого трафіку.
- Не змінює таймінги, інтервали та структуру кадрів.
- Ідеальний для VoIP, відеоаналізу, Triple Play, FDX-аналізу.
- Передає всі кадри без винятку:
    - IPv4/IPv6,
    - з помилками CRC,
    - короткі/довгі,
    - нестандартні,
    - міжкадрові інтервали (IFG).
- Стійкий до збоїв, не створює сліпих зон.
- Єдиний метод, який відповідає юридичним вимогам до законного перехоплення й цифрової криміналістики.
- Не є адресним пристроєм — не піддається взлому.
- Не потребує конфігурації чи оновлень. Підключення = повна видимість.

SPAN — зручний, але технічно обмежений інструмент

- Створює дублікат трафіку програмним способом, що знижує точність аналізу.
- Схильний до перепідписки та втрати пакетів.
- Проблеми з CRC, IFG, jumbo frames, нестандартними кадрами — ігноруються або відкидаються.

- Не підходить для RTP, VoIP, MOS/QoE-аналізу.
- Не завжди передає VLAN теги, що призводить до хибних діагнозів.
- Змінює таймінги та часові мітки пакетів.
- Може бути взломаний або скомпрометований.
- Має найнижчий пріоритет на комутаторі при навантаженні.
- Потребує складної конфігурації CLI — існує ризик помилок.
- У низці країн не відповідає вимогам законодавства щодо lawful intercept.
- Кількість SPAN-портів обмежена, а потреби моніторингу — значно більші.

3 МЕТОДИ ОПТИМІЗАЦІЇ АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ

3.1 Математичні основи та принципи оптимізованого DWT для мережевого аналізу

Аналіз мережевого трафіку як складного динамічного процесу неможливий без застосування математичних інструментів, здатних одночасно відображати і часову, і спектральну структуру сигналу. Традиційні методи обробки, зокрема перетворення Фур'є, забезпечують високу точність у частотній області, проте повністю втрачають інформацію про час появи тих чи інших частотних компонентів. Для мережевого трафіку, що зазвичай є нестаціонарним, імпульсним і містить короткі аномальні піки, така втрата часової локалізації є критичною. Саме тому вейвлет-перетворення та його дискретна форма (DWT) стають основним інструментом у дослідженнях мережевих процесів.

Вейвлет-перетворення ґрунтується на використанні базисних функцій, що є масштабованими і зсунутими копіями материнського вейвлета. Дискретні вейвлет-функції визначаються формулою:

$$\psi_{j,k}(t) = 2^{j/2}\,\psi\left(2^{j}t - k\right) \tag{3.1}$$

де j визначає масштаб (частотний рівень), а k — часовий зсув. Саме завдяки масштабуванню можливе представлення сигналу на різних рівнях деталізації, що дозволяє аналізувати швидкі імпульсні події та повільні зміни в одному об'єднаному математичному апараті.

Дискретне вейвлет-перетворення сигналу x(t) визначається як внутрішній добуток із вейвлет-функцією:

$$W_{\psi}(j,k) = \int_{-\infty}^{\infty} x(t)\ \psi_{j,k}(t)\,dt \tag{3.2}$$

Однак у практичних реалізаціях DWT обчислюється через фільтраційні банки. На першому етапі сигнал пропускається через низькочастотний фільтр h[n] та високочастотний фільтр g[n], після чого результати декадуються — вибирається

кожна друга вибірка. Таким чином отримують коефіцієнти апроксимації та деталізації першого рівня:

$$cA_1[n] = \sum_k x[k]\ h[2n-k] \tag{3.3}$$

$$cD_1[n] = \sum_k x[k]\ g[2n-k] \tag{3.4}$$

Аналогічні операції повторюються для коефіцієнтів cA кожного наступного рівня, що утворює багаторівневе дерево декомпозиції. На верхніх рівнях знаходиться повільна структура трафіку (глобальні тренди), а на нижчих — локальні швидкі події, які часто є індикаторами аномальної поведінки.

Відновлення сигналу з вейвлет-коефіцієнтів також виконується фільтраційним методом. Реконструкція може бути подана формулою:

$$x[n] = \sum_k cA_1[k]\ h_{\text{rev}}[n-2k] + \sum_k cD_1[k]\ g_{\text{rev}}[n-2k] \tag{3.5}$$

де h_{rev} та g_{rev} — фільтри, розгорнуті у зворотному порядку. Це гарантує, що при використанні ортогональних або біортогональних вейвлетів відновлений сигнал буде ідентичним початковому.

Застосування вейвлетів Daubechies у мережевому аналізі зумовлене їх ортогональністю. Умова ортогональності фільтрів задається співвідношенням:

$$\sum_n h[n]\ h[n-2k] = \delta[k] \tag{3.6}$$

де $[k]$ — дельта-функція Кронекера. Ортогональність забезпечує відсутність надлишкових даних та дозволяє виконувати точне відновлення сигналу, що важливо при аналізі трафіку у режимі реального часу.

Крім структурного аналізу, вейвлет-перетворення дозволяє оцінити енергетичні характеристики сигналу на різних частотних рівнях. Енергія деталізації j-го рівня визначається як:

$$E_j = \sum_n \left|cD_j[n]\right|^2 \tag{3.7}$$

Різкі зміни енергії можуть свідчити про сплески трафіку, появу нових типів активності, спроби вторгнення, мережеві атаки або збої в обладнанні. Це робить DWT ефективним інструментом для систем раннього виявлення аномалій.

У підсумку математичний апарат дискретного вейвлет-перетворення формує основу для високоточного та багаторівневого аналізу мережевого трафіку. Завдяки здатності DWT локалізувати події у часі, зберігати частотну структуру, працювати у реальному часі та точно реконструювати сигнал, метод стає надзвичайно ефективним у задачах моніторингу, виявлення атак, цифрової криміналістики та аналізу продуктивності мереж. Оптимальний вибір вейвлет-функції, рівня розкладу та методів попередньої обробки дозволяє адаптувати алгоритм до конкретних особливостей мережевого трафіку, забезпечуючи достовірність і глибину отриманих результатів.

3.2. Особливості застосування вейвлет-розкладання для виявлення аномалій у мережевому трафіку

Виявлення аномалій у мережевому трафіку є фундаментальним завданням у сфері кібербезпеки, телекомунікацій та управління мережевими системами. Сучасні мережі характеризуються високою складністю, великою кількістю взаємопов’язаних компонентів та величезним обсягом трафіку, який проходить через комунікаційну інфраструктуру щосекунди. У таких умовах особливо важливо своєчасно ідентифікувати нестандартні або небезпечні події, що можуть свідчити про атаки, збої, перевантаження або інші критичні ситуації.

Вейвлет-розкладання дає змогу виявляти аномалії у складних, нерівномірних та нестаціонарних потоках даних, таких як мережевий трафік. Основна особливість цього методу полягає у здатності одночасно відображати як часовий, так і частотний аспекти поведінки сигналу. Це дозволяє точно визначити моменти, коли у трафіку відбуваються короткочасні зміни, які можуть бути ознакою загроз чи відхилень. Наприклад, класичний аналіз Фур’є хоча й забезпечує добре частотне представлення сигналу, але він не дозволяє визначити, у який саме момент

з’явилися ті чи інші частотні компоненти. Натомість вейвлети дозволяють побачити і появу, і тип, і інтенсивність аномалії.

Одним із важливих застосувань вейвлетів є їх здатність розкладати трафік на кілька масштабних рівнів. Кожний рівень відповідає певному частотному діапазону. Найнижчі рівні відображають високочастотні компоненти — саме вони найчастіше пов’язані з аномальними подіями. Наприклад, різкі імпульси можуть бути викликані сплесками ICMP-трафіку, підготовчими діями сканування портів або початком DoS-атаки. Натомість високі рівні апроксимації показують довготривалі тренди — такі, як поступове збільшення навантаження на сервер, зміна структури користувацьких запитів або перехід системи в аномальний режим роботи, який розвивається протягом хвилин чи годин.

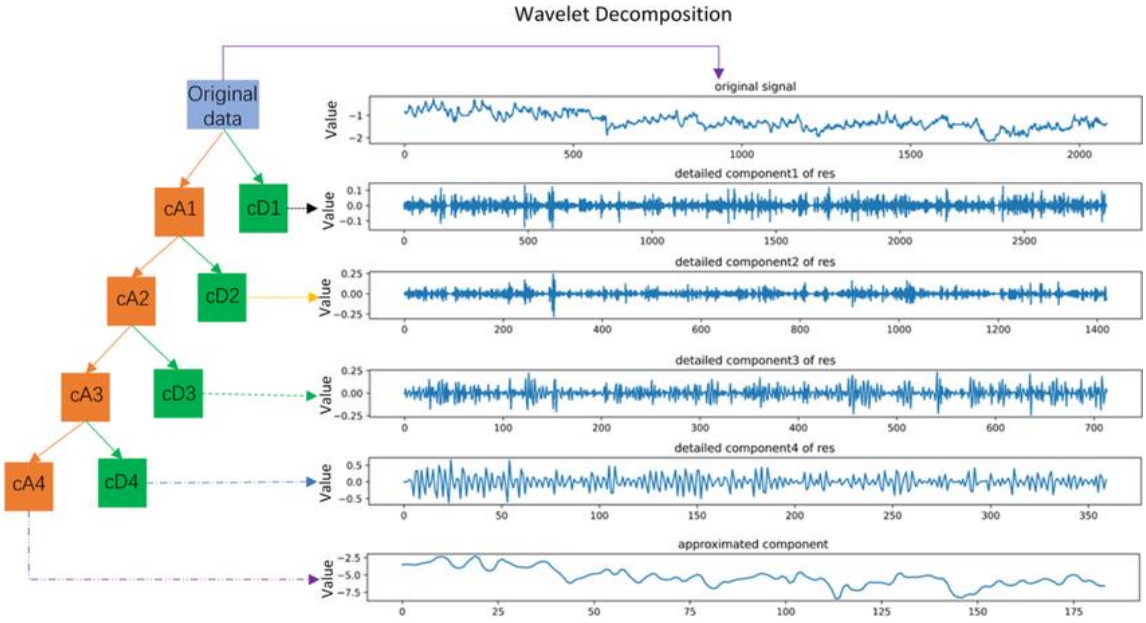


Рис. 3.1 Багаторівнева схема вейвлет-декомпозиції сигналу та відповідні деталізуючі й апроксимуючі компоненти

Однією з переваг використання вейвлет-розкладання у мережевому аналізі є здатність ефективно працювати з шумними сигналами. Мережевий трафік містить велику кількість шумових компонентів, спричинених випадковими коливаннями, внутрішньою активністю системи, балансуванням навантаження, особливостями роботи протоколів. Вейвлети здатні відокремлювати шум від корисної інформації завдяки фільтрації деталізуючих коефіцієнтів.

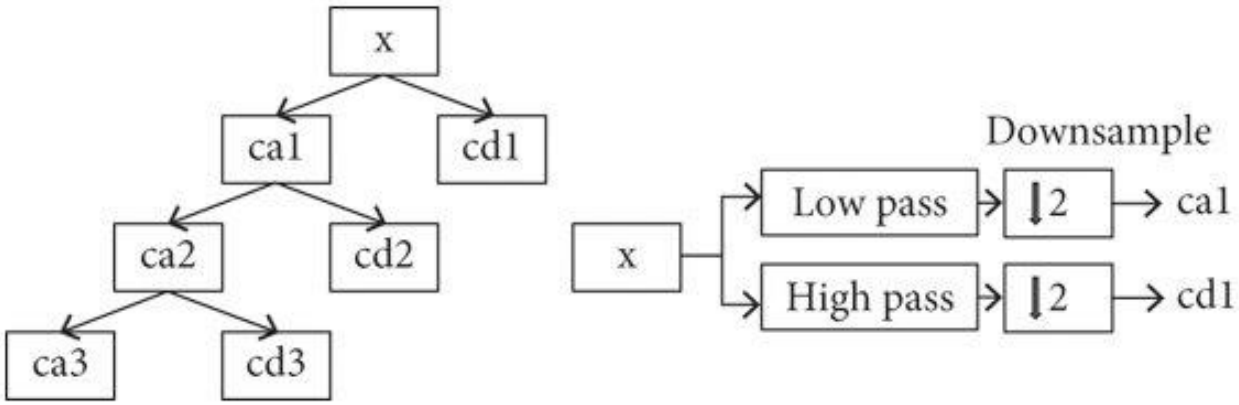


Рис. 3.2 Схематичне подання фільтраційного банку дискретного вейвлет-перетворення та формування коефіцієнтів апроксимації й деталізації

Під час застосування м'якого чи жорсткого порогування можна пригнічувати несуттєві коливання, які не несуть аналітичної цінності, але водночас точно виділяти піки, викликані справді підозрілими подіями. Це особливо важливо для систем, що працюють у режимі реального часу, оскільки вони мають реагувати лише на сигнали, що перевищують певний рівень значущості.

У цифровій криміналістиці та системах виявлення вторгнень (IDS) вейвлети відіграють роль первинного селектора, що формує базу ознак для подальшої обробки. Зокрема, аналіз вейвлет-коефіцієнтів дозволяє будувати ефективні моделі машинного навчання. Такі моделі здатні класифікувати мережеві події, розрізняти нормальний та аномальний трафік, передбачати розвиток атак. У практичних системах, таких як Snort чи Suricata, вейвлети можуть бути використані на першому етапі фільтрації, де особлива увага приділяється інтенсивним коливанням у певних протокольних потоках: HTTP, HTTPS, DNS, ICMP, SSH та інші. Наприклад, різка зміна частоти DNS-запитів може сигналізувати про DNS-ампліфікацію або тунелювання трафіку шкідливими інструментами. У свою чергу, піки у TCP SYN можуть бути маркером SYN-флуду.

Особливе місце займає вейвлет-аналіз у дослідженні потоків VoIP, відео та інших реального-часових сервісів. Такі системи чутливі до джитеру, затримок, зміни пропускної здатності. Вейвлети дозволяють виявляти локальні збої у цих потоках, що особливо важливо при оцінці якості сервісу (QoS). Наприклад,

незвичні коливання у деталізаційному рівні можуть свідчити про деградацію каналу або наявність атаки, спрямованої на порушення якості голосу. У випадку RTP-потоків вейвлети допомагають відстежувати фрагментацію пакетів, різкі піки затримки або появу повторів, що може бути ознакою втручання в інфраструктуру або проблем зі сторони провайдера.

З технічного погляду важливою перевагою вейвлетів є те, що вони можуть бути застосовані до різних типів даних, які представляють мережеві активності. Це можуть бути часові ряди кількості пакетів за фіксований інтервал, середній розмір пакету, кількість активних TCP-з'єднань, рівень використання пропускної здатності, кількість помилок CRC на інтерфейсі або динаміка RTT у TCP-сесіях. Завдяки універсальному характеру DWT можна обробляти будь-яку метрику, що відображає поведінку мережі в часі. Це робить вейвлет-аналіз надзвичайно гнучким інструментом, який легко адаптується до конкретних задач або особливостей інфраструктури.

Сучасні мережі, особливо великі корпоративні та дата-центрові, мають високі швидкості, що можуть досягати десятків гігабіт за секунду. У таких умовах класичні методи аномалієвого моніторингу нерідко відстають від реального навантаження через надмірну обчислювальну складність. DWT, завдяки своїй обчислювальній ефективності, здатний працювати у режимі близькому до реального часу. Це дозволяє використовувати його на високошвидкісних комутаторах, сенсорах мережевої безпеки або потокових системах збору даних без значного зниження продуктивності основних компонентів.

Важливою проблемою у мережевому аналізі є обробка коротких, але значущих піків. Наприклад, атака SYN flood може початися з короткого імпульсу, який лише частково відрізняється від нормальної поведінки. Вейвлети здатні виявити цю різницю на деталізаційному рівні, навіть якщо у загальному масштабі сигнал виглядає цілком звичайно. Тобто DWT дозволяє ідентифікувати "приховані" аномалії, які класичні методи пропускають. У контексті кібербезпеки це суттєво підвищує шанси на раннє виявлення атаки до того, як вона набуде руйнівного характеру.

Аналітичні системи, що використовують вейвлети, також можуть будувати профілі поведінки на основі історичних даних. Порівняння поточного стану сигналу з такими профілями дозволяє виявляти системні зміни. Наприклад, зростання активності на порту SSH у нічний час може бути неприродним для корпоративної мережі. На рівні деталізації такі зміни формують характерні піки, які легко помітити навіть у складних та шумних середовищах. Таким чином, DWT може використовуватися для поведінкового аналізу, що є одним з найефективніших методів виявлення нових, раніше невідомих типів атак.

У загальному вигляді вейвлет-розкладання виявляється надзвичайно ефективним у задачах аналізу мережевих аномалій через поєднання кількох критичних властивостей: здатність працювати з нестаціонарними сигналами; високу точність локалізації коротких подій; адаптивність до різних масштабів; низьке обчислювальне навантаження; можливість функціонувати у потоковому режимі. У реальних системах моніторингу ці властивості дозволяють суттєво підвищити ефективність роботи SOC-центрів, систем IDS/IPS, засобів цифрової криміналістики та платформ управління продуктивністю мереж.

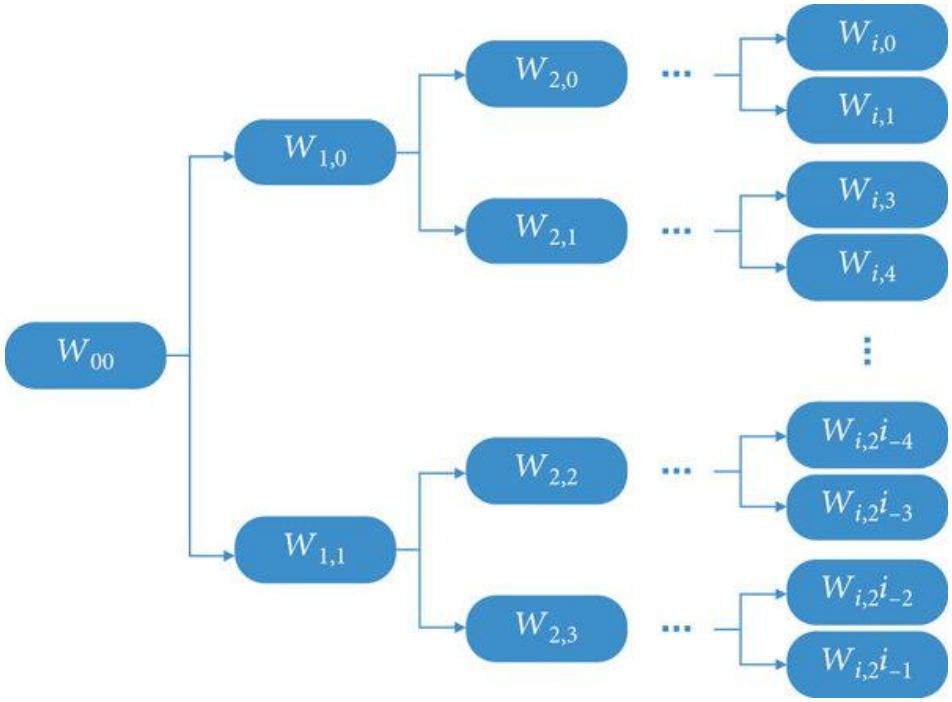


Рис. 3.3 Узагальнена схема багаторівневого вейвлет-пакетного розкладання для аналізу складних мережевих сигналів

Таким чином, вейвлети є не просто математичним інструментом, а потужним аналітичним механізмом, який дозволяє розкривати приховані закономірності, виявляти аномальні події і прогнозувати поведінку мережевих систем. Їх застосування у сфері аналізу мережевого трафіку відкриває можливості для значного підвищення точності виявлення інцидентів та загроз, що робить їх одним з ключових компонентів сучасних систем кіберзахисту.

3.3 Оптимізація методики збору трафіку та попередня обробка даних для вейвлет-аналізу

Ефективність вейвлет-розкладання у задачах аналізу мережевого трафіку визначається не лише вибором типу вейвлета, рівнем декомпозиції чи алгоритмом обчислень, а передусім якістю і коректністю даних, які отримуються на етапі збору та попередньої обробки. Наявність шумів, пропусків, фрагментованих пакетів, неточних позначок часу або спотворень, пов'язаних з перевантаженням комутаційного обладнання, суттєво впливають на структуру часових рядів, що згодом можуть призвести до хибних висновків під час застосування DWT. Тому для забезпечення надійності аналізу необхідно ретельно оптимізувати кожен етап формування вхідного сигналу.

Одним із ключових аспектів є вибір технології доступу до трафіку. Джерелом даних можуть бути SPAN-порти, TAP-пристрої або програмні агенти, що працюють на кінцевих вузлах. У контексті вейвлет-аналізу особливо важлива точність часової структури сигналу, тому обладнання, яке може відкидати пакети, змінювати їх порядок або порушувати інтервали між подіями, є менш бажаним. TAP-пристрої забезпечують найбільш надійний і точний канал отримання трафіку, оскільки вони передають кожен біт без спотворення, зберігаючи реальні інтервали між пакетами. Саме такі дані найкраще підходять для подальшої часової декомпозиції.

Зібраний трафік часто містить велику кількість інформації, однак не вся вона є корисною для потреб вейвлет-аналізу. Важливо перетворити неструктурований потік пакетів у формат, який дозволяє будувати часоряд. Для цього застосовують

процес агрегування, коли значення трафіку формуються за певні фіксовані інтервали часу. Типовими агрегатами є кількість пакетів за інтервал, сумарний обсяг даних, кількість TCP-сесій, рівень навантаження на канал, кількість помилок CRC або інші величини, що характеризують активність у мережі. Правильний вибір розміру інтервалу має критичне значення: надто велика тривалість інтервалу здатна приховати короткотривалі аномалії, тоді як надто мала — навпаки, зробить сигнал занадто шумним і вимагатиме додаткового згладжування.

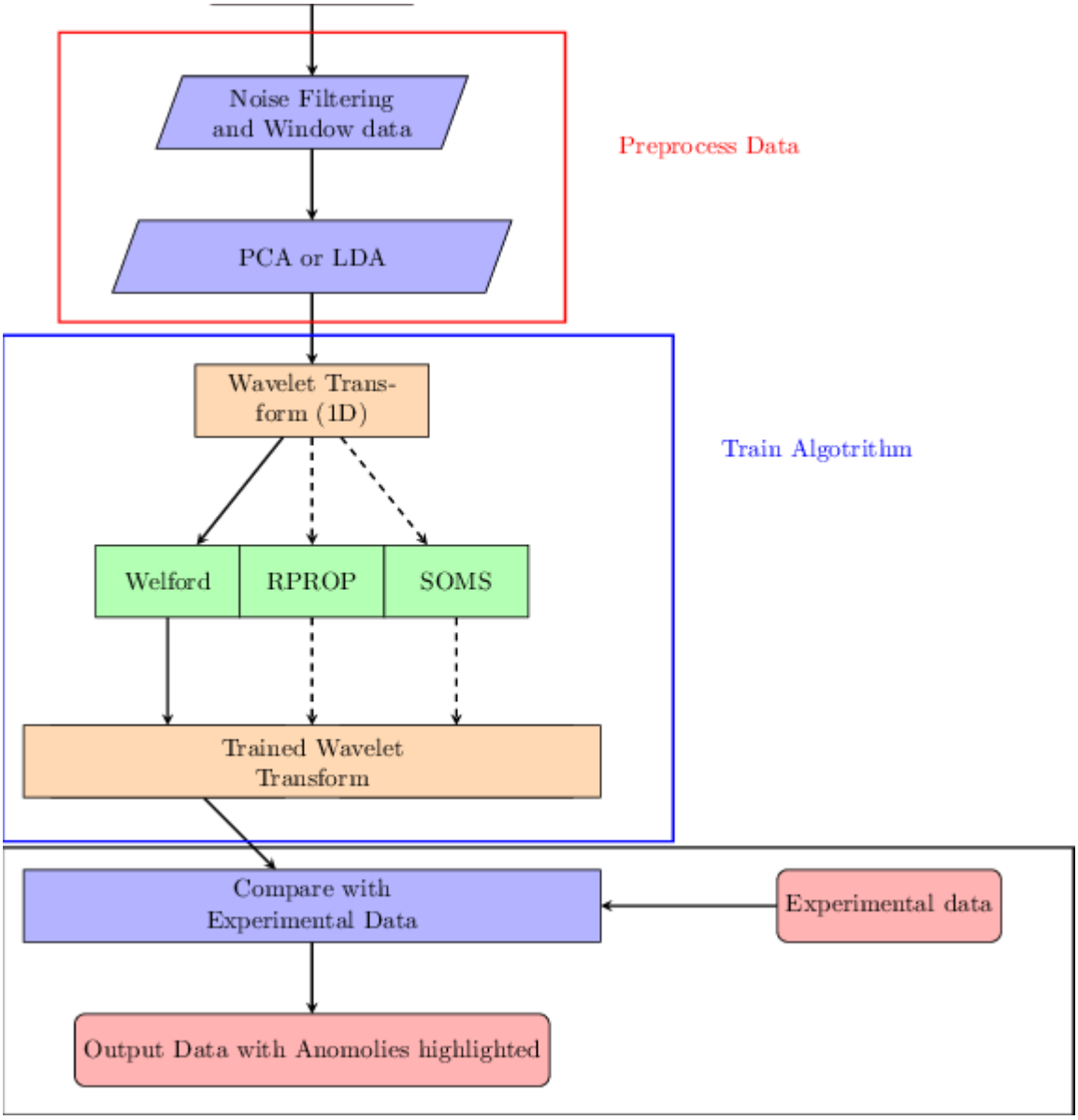


Рис. 3.4 Узагальнена схема попередньої обробки мережевого трафіку та навчання моделі на основі вейвлет-перетворення для виявлення аномалій

Під час побудови часового ряду також важливо забезпечити точність синхронізації даних. Затримки, спричинені особливостями роботи мережевого

обладнання або операційної системи, можуть викликати зміщення окремих точок, що суттєво вплине на структуру сигналу і, відповідно, на вейвлет-перетворення. Щоб цього уникнути, використовують апаратні або високоточні програмні таймери, технології синхронізації на основі протоколів NTP або PTP, а у високошвидкісних мережах — навіть апаратні timestamp-модулі на рівні мережевих карт. Забезпечення точності часової розмітки є одним із найважливіших елементів оптимізації, адже вейвлети надзвичайно чутливі до структури часових масштабів.

Після формування первинного часового ряду необхідно провести етап нормалізації сигналу. У багатьох випадках різні ділянки трафіку можуть мати значно відмінні масштаби, тому перед застосуванням DWT доцільно використовувати стандартизацію або мінімально-максимальну нормалізацію. Це дозволяє позбутися небажаного впливу різких пікових значень і водночас зберегти їхню форму і локалізацію, що важливо для подальшого аналізу аномалій. Особливо це актуально для трафіку, де сусідні інтервали можуть містити як дуже низьку, так і надзвичайно високу активність.

Ще одним важливим кроком є згладжування сигналу з метою зменшення впливу високочастотного шуму, який не пов'язаний із реальною мережевою активністю. Шум може бути спричинений роботою операційної системи, особливостями драйверів, мікробурстами або випадковими подіями, що не мають відношення до аномалій. Для згладжування застосовують рухомі середні, медіанні фільтри або попереднє низькочастотне фільтрування. Варто враховувати, що надмірне згладжування здатне приховати реальні короткочасні аномалії, тому цей етап потребує обережного налаштування.

У деяких випадках часовий ряд може містити пропуски або нерівні інтервали. Це часто трапляється у високонавантажених мережах, де комутаційне обладнання відкидає деякі пакети в моменти перевантаження, або при зборі даних через SPAN-порти. Вейвлет-аналіз чутливий до відсутності точок у часовому ряду, адже це порушує структуру сигналу. Тому необхідно виконувати інтерполяцію пропусків. Для цього застосовують різні методи: лінійну інтерполяцію, інтерполяцію на основі сплайнів або реконструкцію на основі локальних моделей тренду. Найкращий

результат дають методи, що враховують локальну структуру сигналу, оскільки вони зберігають реальні властивості трафіку.

Особливої уваги потребує етап фільтрації сплесків, які не несуть аналітичної цінності. У мережева трафіку часто зустрічаються короткі піки, пов'язані з архітектурою програм, синхронізацією у хмарних середовищах або фоновими службами. Такі піки можуть вводити в оману вейвлет-розкладання, оскільки деталізуючі коефіцієнти реагують на них так само сильно, як і на справжні аномалії. Щоб уникнути цього, необхідно проводити фільтрацію за амплітудними або часовими критеріями, вилучаючи несправжні коливання, які повторюються з високою регулярністю й не пов'язані з порушеннями у мережі.

Особливе місце у процесі підготовки займає сегментація трафіку. У деяких випадках аналіз усього масиву даних не є доцільним, тому важливо виконати поділ сигналу на логічні блоки: наприклад, за типом протоколу, за напрямком трафіку (uplink / downlink), за джерелом або призначенням, за типами пакетів або їх довжиною. Така попередня сегментація дозволяє вейвлет-методам працювати більш ефективно, оскільки алгоритм отримує більш однорідні дані, а отже, зменшується кількість хибно позитивних виявлень.

Останнім важливим етапом є перевірка коректності підготовлених даних перед передачею їх у модуль вейвлет-розкладання. Для цього застосовуються статистичні методи оцінки, такі як визначення автокореляції, спектральної щільності, ентропії та інших характеристик, що дозволяють переконатися, що структура сигналу збережена. У разі наявності суттєвих спотворень необхідно змінити параметри агрегації, нормалізації або фільтрації й повторити підготовку даних.

Таким чином, оптимізація методики збору та попередньої обробки трафіку є критично важливим етапом для забезпечення точності й надійності вейвлет-аналізу. Якісно сформований часовий ряд дозволяє DWT максимально точно відображати реальні зміни в трафіку, виявляти приховані закономірності та своєчасно реагувати на появу аномалій. Належна організація збору та попередньо оброблені дані

створюють фундамент для побудови потужних систем моніторингу й захисту мережевих інфраструктур, які базуються на методах вейвлет-декомпозиції.

У процесі підготовки даних для вейвлет-аналізу важливо також організувати коректну фіксацію метаданих, які супроводжують кожен пакет або агрегований елемент часового ряду. Метадані включають адресу відправника та одержувача, протокол транспортного рівня, прапори TCP, довжину пакету, індекс інтерфейсу, номер VLAN або ідентифікатор QoS-класу. У випадку високонавантажених сегментів мережі саме метадані можуть містити важливі ознаки аномалії, навіть якщо основний трафік виглядає цілком нормальним. Наприклад, різке збільшення кількості TCP SYN із певного підмережевого діапазону може статися ще до того, як інтенсивність трафіку підніметься до рівня, що помітний у часовому ряду. Тому збереження метаданих і можливість їх використання на стадії попередньої обробки дозволяє суттєво підвищити точність подальшого аналізу.

Не менш важливим є питання забезпечення цілісності даних під час їх передачі від датчика або інструмента збору до системи аналізу. У великих корпоративних мережах збір даних часто здійснюється на периферійних вузлах, після чого агрегований трафік передається у центральний аналітичний кластер. Якщо передавання відбувається ненадійним каналом або без контролю черговості, можлива поява затримок та пропусків у даних, що руйнує структуру часового ряду. Застосування надійних транспортних протоколів, таких як gRPC або спеціалізовані телеметричні рішення (наприклад, Cisco Tetration, Juniper AppFormix або OpenTelemetry), дозволяє мінімізувати втрати і забезпечити синхронне надходження даних до аналітичного ядра.

У разі використання потокових механізмів збору, таких як sFlow, NetFlow або IPFIX, необхідно враховувати, що ці протоколи не збирають кожен пакет, а здійснюють статистичний відбір. Хоча такий відбір може бути достатнім для оцінки загальної поведінки мережі, він не підходить для високоточних методів, таких як DWT, які залежать від структури сигналу та точних інтервалів між подіями. Якщо ж використання потокових протоколів неминуче, варто застосовувати методи реконструкції часового ряду на основі ймовірнісних моделей, наприклад, моделей

Пуассона або марковських процесів. Це дозволяє відновлювати приблизні значення між двома сусідніми вимірюваннями, проте слід пам’ятати, що точність відновлення буде нижчою, ніж при роботі з повним набором пакетів.

У процесі попередньої обробки трафіку значну роль відіграє також видалення надлишкових або повторюваних даних. Навіть у високошвидкісних мережах існують періоди, коли трафік має регулярну повторювану структуру, наприклад, через циклічні виклики API, оновлення таблиць маршрутизації чи періодичні heartbeat-пакети. Якщо такі події не становлять аналітичного інтересу або навіть заважають подальшому аналізу, їх можна згортати у компактніші форми або видаляти з часового ряду. Така оптимізація зменшує шум і частково знижує обчислювальне навантаження на систему під час подальшого вейвлет-розкладання.

Дуже важливим аспектом у роботі з мережевим трафіком є виявлення і коригування так званих «мікроберстів» — коротких імпульсів високої інтенсивності, які можуть тривати від кількох мікросекунд до кількох мілісекунд. Мікроберсти, як правило, не видно у традиційних інструментах моніторингу, оскільки їх усереднюють під час збору статистичних даних. Проте для вейвлет-розкладання мікроберсти мають велике значення, оскільки саме вони можуть створювати високочастотні компоненти, які потім помилково класифікуються як аномалії. З метою правильної інтерпретації їх слід або зберігати у високій часовій роздільності, або згладжувати з використанням спеціальних методів. На практиці для таких задач ефективно застосовують адаптивне віконне усереднення, коли параметри згладжування змінюються залежно від локальної структури сигналу.

Ще одним елементом оптимізації є вибір правильного способу репрезентації трафіку. Деякі системи обирають кількість пакетів, інші — обсяг даних, треті — інтенсивність запитів певного типу. У різних випадках це може істотно впливати на кінцевий результат. Наприклад, для атак типу порт-сканування найефективнішим є часоряд, побудований на основі кількості TCP SYN-пакетів, а при виявленні DDoS на рівні каналу доцільнішими є байтові агрегати. Вибір правильної метрики визначає, наскільки добре вейвлет-розкладання виділить реальні аномалії серед шуму. Часто для аналізу складних інцидентів будують кілька часових рядів

одночасно: один відображає кількість пакетів, другий — їх середній розмір, третій — частоту помилок. Подальший вейвлет-аналіз може виконуватися як над кожним рядом окремо, так і над їх поєднанням за допомогою багатоканального DWT.

Суттєво підвищити точність системи дозволяє застосування методів детекції викидів на ранньому етапі. Іноді значення часового ряду можуть суттєво виходити за межі природного діапазону — наприклад, через збій у роботі інструмента збору даних або через артефакти роботи SPAN-порту. Подібні значення не відображають реальну поведінку мережі, але можуть створити сильні піки у деталізуючих коефіцієнтах і таким чином зіпсувати результати розкладання. Тому перед застосуванням DWT проводять автоматичну перевірку на викиди та замінюють їх значеннями, наближеними до локального тренду.

Один із найменш очевидних, але важливих етапів у підготовці даних — це зменшення розрядності. У багатьох випадках трафік збирається з надмірною точністю, наприклад із часовими позначками у наносекундах або байтовою точністю до 64-біт. Сигнали такого рівня деталізації не лише створюють надмірне навантаження на систему, але й ускладнюють вейвлет-реконструкцію, оскільки високочастотні компоненти можуть бути викликані саме надмірною точністю, а не реальною структурою трафіку. Застосування цілеспрямованого округлення або зменшення розрядності дозволяє зберегти сутність сигналу і зменшити розмір обчислень.

Зміна частоти вибірки є ще одним методом оптимізації. Занадто висока частота може створювати надлишкову кількість даних, тоді як надто низька — приховувати аномалії. Правильний вибір частоти вибірки залежить від типу мережі, швидкості інтерфейсів і типу трафіку. Наприклад, у мережах 1 Гбіт/с оптимальною може бути вибірка в інтервалі 1–5 мс, у мережах 10 Гбіт/с — 0,1–1 мс, тоді як у Wi-Fi мережах із високою варіативністю часів доставки значення іноді доводиться збільшувати до 10–20 мс. Вейвлет-розкладання є чутливим до частоти вибірки, оскільки структура рівнів декомпозиції прив’язана до часової шкали.

Нарешті, важливим аспектом попередньої обробки є збереження всіх етапів і параметрів у вигляді метаданих. Це дозволяє відтворювати результати аналізу,

проводити порівняння між різними методиками, а також інтегрувати вейвлет-аналітику в системи машинного навчання. Сучасні платформи моніторингу, такі як Elastic Stack, Splunk або Prometheus, дають можливість збирати метрики разом із контекстом їх формування, що значно підвищує можливості для побудови комплексних моделей.

У сукупності всі ці методи дозволяють сформувати структуровано підготовлений часовий ряд, який максимально точно відображає реальну поведінку мережевого трафіку. Це створює оптимальне середовище для застосування методів вейвлет-розкладання, що, у свою чергу, забезпечує високу точність виявлення аномалій, швидке реагування на інциденти та можливість довгострокового прогнозування поведінки мережевих систем.

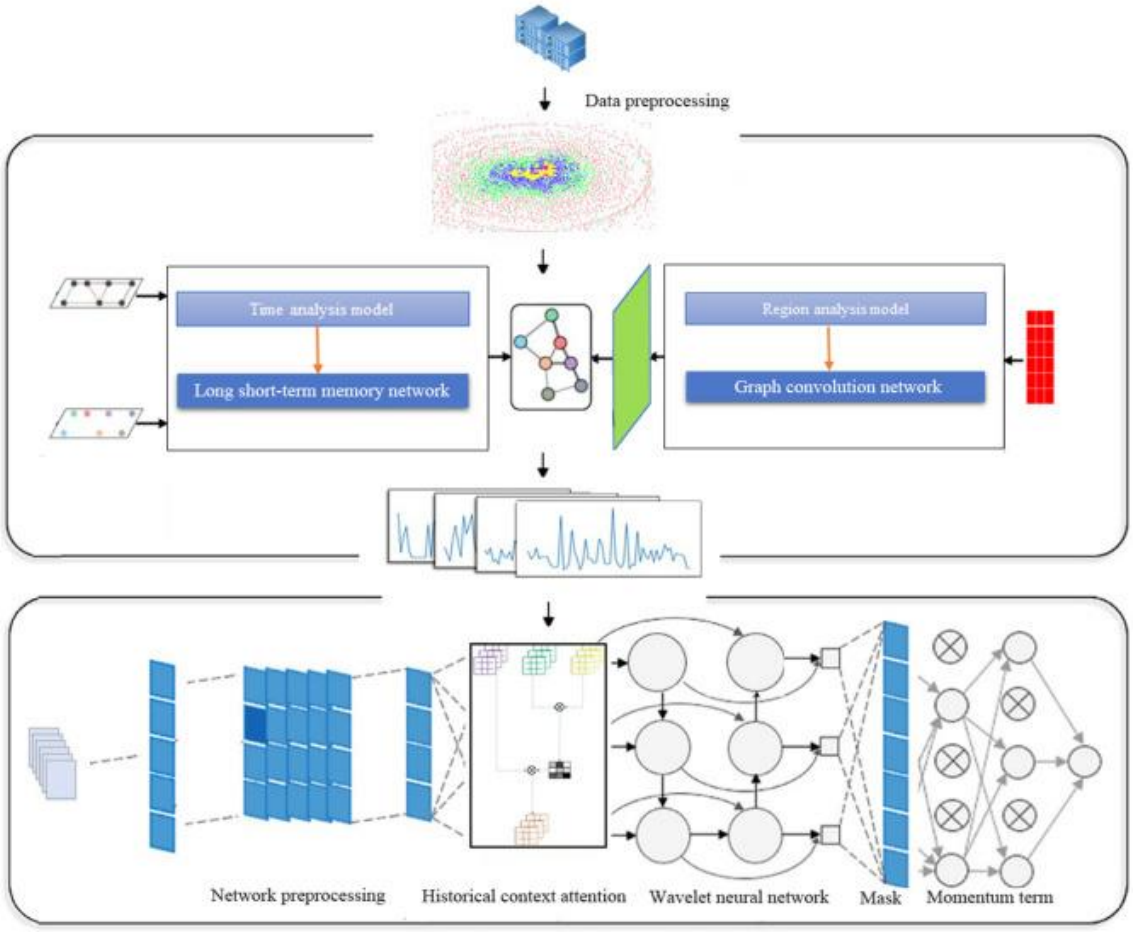


Рис. 3.5 Архітектура комплексної системи аналізу мережевого трафіку із застосуванням попередньої обробки, нейронних мереж та вейвлет-перетворення

4 РЕАЛІЗАЦІЯ МЕТОДИКИ ТА МОДЕЛЮВАННЯ РЕЗУЛЬТАТІВ АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ

4.1 Архітектура системи збору та обробки мережевого трафіку

Архітектура системи збору та обробки мережевого трафіку відіграє ключову роль у забезпеченні можливості проведення високоточних досліджень та моделювання поведінки мережевих систем. Будь-яка сучасна система аналізу трафіку, яка має працювати з великими обсягами даних у режимі, наближеному до реального часу, повинна бути побудована на основі ретельно спроєктованої інфраструктури збору, накопичення, фільтрації та попередньої обробки даних. Лише за умови правильної організації всіх цих етапів можливе застосування подальших методів обробки, зокрема таких, як вейвлет-розкладання та інші алгоритми багаторівневого аналізу. Тому побудова архітектури системи є фундаментом, на якому базується точність, продуктивність та достовірність результатів дослідження.

Будь-яка система збору трафіку складається з декількох базових компонентів: джерела даних, блоку захоплення пакетів, модуля первинної фільтрації, підсистеми агрегації та нормалізації, сховища трафіку, а також окремого обчислювального модуля, що відповідає за аналітичну обробку даних. Кожний з цих компонентів має визначену функціональність і взаємодіє з іншими елементами інфраструктури, забезпечуючи безперервний потік інформації. Відсутність оптимізації хоча б на одному етапі може призвести до накопичення помилок, втрати цінних даних або неможливості відтворити реальні параметри мережевої активності.

Першим етапом є організація доступу до трафіку. Для цього можуть використовуватися три основні джерела: SPAN-порти комутаторів, апаратні TAP-пристрої або програмні агенти, встановлені на вузлах. У сучасних умовах пріоритет віддається TAP-пристроям, оскільки вони забезпечують найбільш точне віддзеркалення трафіку без втручання в роботу мережі. На відміну від SPAN-портів, які можуть відкидати частину пакетів у разі перевантаження, TAP-пристрої

передають повний потік даних, включаючи дрібні пакети, пошкоджені фрейми, а також міжкадрові інтервали, які нерідко містять важливу інформацію про роботу мережевого обладнання. Це особливо важливо у контексті вейвлет-аналізу, адже навіть незначні спотворення часових характеристик можуть мати вплив на структуру коефіцієнтів розкладання.

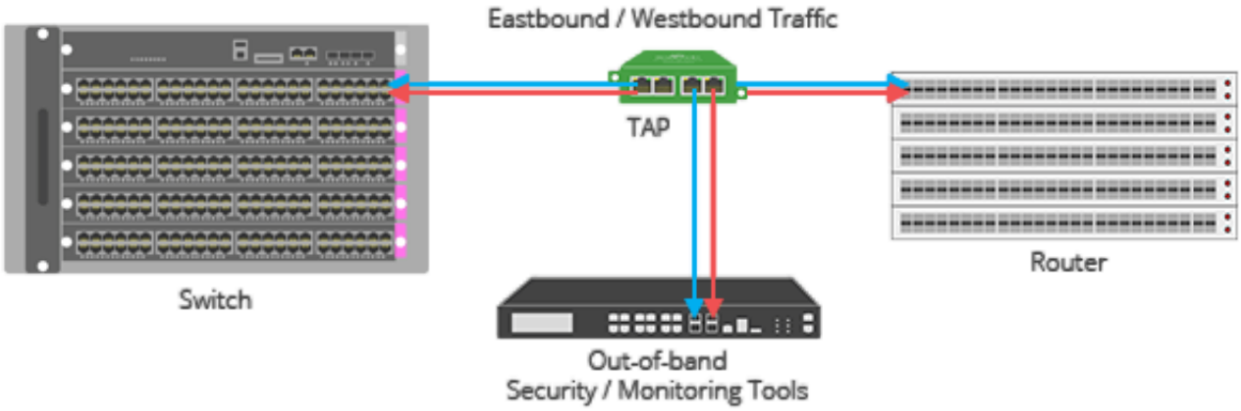


Рис. 4.1 Схема підключення TAP-пристрою для пасивного перехоплення мережевого трафіку між комутатором та маршрутизатором

Другим компонентом архітектури є модуль захоплення пакетів. Залежно від інтенсивності трафіку та пропускної здатності мережевого інтерфейсу, цей модуль може бути реалізований на базі стандартних бібліотек (таких як libpcap або WinPcap) або на основі високошвидкісних фреймворків (DPDK, PF_RING, Netmap). Високошвидкісні фреймворки дозволяють працювати з інтерфейсами 10, 40 або 100 Гбіт/с, забезпечуючи мінімальні затримки та повну обробку потоку. Такі рішення використовуються в корпоративних центрах обробки даних, а також у сегментах критичної інфраструктури, де втрата навіть одного пакету може мати суттєве значення.

Після захоплення пакетів необхідно виконати етап первинної фільтрації. На цьому етапі система відсіює трафік, що не має аналітичного значення. Зокрема, можуть відфільтровуватися службові протоколи, локальний ARP-трафік, керуючі пакети CDP або LLDP, дублікати пакетів, створені через дзеркальне копіювання на декількох рівнях, а також шумові пакети, що утворюються на рівні операційних

систем. Первинна фільтрація зменшує навантаження на подальші етапи обробки та знижує вимоги до обчислювальних ресурсів системи.

Ключовою частиною архітектури є модуль агрегації та формування часового ряду. Саме він перетворює неструктурований потік пакетів у впорядковану послідовність значень, що відображають інтенсивність трафіку у вибраних часових інтервалах. Вибір параметрів агрегації залежить від задачі дослідження: тривалість інтервалу, метрика (байтова інтенсивність, кількість пакетів, кількість активних TCP-сесій), рівень деталізації даних та тип протоколів. Важливо, щоб агрегація не спотворювала структуру даних, оскільки основною властивістю трафіку є його нестаціонарність — а це означає, що сигнал може різко змінювати свою форму за дуже короткий проміжок часу.

Після агрегації виконується модуль нормалізації. Даний етап приводить значення часового ряду до єдиної шкали, що дозволяє уникнути впливу надмірно великих або надто малих значень, які можуть викривляти результати аналізу. У багатьох корпоративних мережах спостерігається значна варіативність навантаження: у пікові години обсяг трафіку може збільшуватися у десятки разів, тоді як у нічний час залишатися майже нульовим. Тому нормалізація забезпечує стабільність і порівнюваність результатів обробки.

Наступною складовою архітектури є модуль фільтрації високочастотного шуму. Оскільки вейвлет-аналіз дуже чутливий до локальних піків, які можуть бути спричинені як аномаліями, так і технічними особливостями роботи мереж, необхідно відокремити штучні стрибки від реальних подій. На цьому етапі застосовують медіанні фільтри, адаптивні згладжувальні алгоритми або фільтри нижніх частот, які зменшують шум, зберігаючи при цьому важливі характеристики сигналу.

Після первинної обробки сформований часовий ряд передається до модуля зберігання даних. У сучасних системах аналізу зазвичай застосовуються два типи сховищ: високошвидкісні SSD-кластери для короткострокового зберігання великих обсягів трафіку та розподілені бази даних для довготривалої архівації. Використання таких рішень, як Elasticsearch, InfluxDB, ClickHouse або Apache

Cassandra, дозволяє будувати масштабовані сховища, здатні обробляти дані у реальному часі, виконувати пошук, фільтрацію й повторний аналіз сигналів.

Окремою складовою архітектури є модуль аналітичної обробки, де виконується вейвлет-розкладання, побудова моделей поведінки, порівняння з історичними шаблонами та формування висновків. Цей модуль може бути реалізований як у вигляді локального процесора аналітики, так і у вигляді розподіленої системи, що працює на базі Apache Spark, Dask або Kubernetes-кластерів із GPU-прискоренням. Розподілені обчислення дозволяють паралельно обробляти великі набори даних, що є необхідним для масштабних мережевих середовищ, де обсяг трафіку може вимірюватися терабайтами на годину.

Важливим аспектом архітектури є забезпечення синхронізації між різними компонентами. Наявність точних timestamp'ів має вирішальне значення, оскільки аналіз часової структури сигналу можливий лише за умови, що фіксація часу відбувається з високою точністю. Для цього використовують NTP, PTP або апаратні модулі синхронізації, інтегровані у мережеві інтерфейси. У високошвидкісних середовищах затримки навіть у 100–200 мікросекунд можуть вплинути на результати аналізу.

Не менш важливою частиною архітектури є модуль моніторингу та керування системою. Він забезпечує контроль стану всіх компонентів, перевірку цілісності даних, виявлення перевантажень, відстеження затримок у потоці даних та оперативне реагування на відхилення. Завдяки цьому система може працювати безперервно, а у випадку збоїв — швидко відновлювати свою функціональність.

Таким чином, архітектура системи збору та обробки мережевого трафіку є складною багатокомпонентною структурою, що включає в себе модулі захоплення, фільтрації, агрегації, нормалізації, зберігання та аналітичної обробки даних. Від якості реалізації кожного з цих етапів залежить точність і ефективність подальшого вейвлет-аналізу, можливість виявлення аномалій та загроз, а також здатність системи забезпечити глибоке розуміння процесів, що відбуваються у мережевому середовищі.

Окремої уваги потребує питання організації потоків даних у системі збору та обробки трафіку. Сучасні мережі характеризуються величезною кількістю паралельних потоків, які утворюються внаслідок взаємодії великої кількості користувачів, сервісів, мікросервісних платформ, віртуальних машин та контейнеризованих середовищ. Тому архітектура системи повинна враховувати можливість сегментації трафіку на логічні потоки — наприклад, за протоколами, за джерелами походження, за типами сервісів або за напрямками передачі. Такий підхід дозволяє не лише знизити навантаження на аналітичні модулі, але і забезпечити гнучкість при моделюванні трафіку з різних джерел. Наприклад, трафік до веб-сервера зазвичай має характер запит-відповідь і містить відносно передбачувані патерни, тоді як трафік між компонентами розподіленої хмарної платформи може бути значно більш інтенсивним і фрагментованим.

Сегментація трафіку також дозволяє створювати різнорівневі моделі поведінки — наприклад, окремо для DNS-запитів, окремо для потоків HTTPS, окремо для внутрішньокорпоративних API-викликів. Вейвлет-аналіз у цьому випадку може виконуватися на кожному потоці незалежно, що дозволяє будувати багатошарові профілі активності та виявляти аномалії, характерні саме для конкретного типу трафіку. Це особливо важливо для систем безпеки, де різні вектори атак можуть проявлятися у різних протоколах по-різному.

Ще одним важливим елементом архітектури є взаємодія між системою збору трафіку та інфраструктурою мережевих пристроїв. У високонавантажених середовищах ця взаємодія має бути максимально оптимізованою, щоб уникати створення додаткових точок відмови. Наприклад, у мережах великих підприємств часто застосовують виділені канали для передачі дзеркального трафіку, щоб уникнути його конкуренції з основним робочим трафіком. У деяких випадках організовують навіть окремі «мережі для моніторингу», у яких передаються лише дані для аналізу. Такий підхід мінімізує ризики перевантаження комутаторів і дозволяє аналітичним платформам працювати без втручання у продуктивність бізнес-критичних сервісів.

Крім того, у системі може застосовуватися багаторівневий механізм резервування. Оскільки мережевий трафік у багатьох випадках містить дані, необхідні для юридичного аудиту, цифрової криміналістики або контролю відповідності, навіть невелика втрата даних може мати значні наслідки. Тому архітектура часто передбачає дублювання трафіку на кілька незалежних систем зберігання або аналітичних центрів. У випадку виходу з ладу одного з компонентів інфраструктури, дані залишаються доступними в альтернативному сховищі.

Що стосується аспектів безпеки, система повинна забезпечувати повну ізоляцію аналітичних модулів від основної мережі. Оскільки трафік часто містить конфіденційну інформацію, необхідно забезпечити, щоб інструменти аналізу працювали лише у режимі пасивного прослуховування. Це означає, що вони фізично не повинні мати можливості впливати на передачу пакетів у робочій мережі. TAP-пристрої надають таку гарантію завдяки односторонньому копіюванню трафіку.

Особливої складності додають віртуалізовані та хмарні середовища. У таких системах традиційні TAP-пристрої використовувати складно, тому застосовуються програмні TAP-агенти або віртуальні мережеві інтерфейси, які перехоплюють трафік між віртуальними машинами або контейнерами. Це вимагає інтеграції з такими платформами, як VMware vSphere, OpenStack, Kubernetes або Docker. У хмарних середовищах також застосовуються такі технології, як AWS VPC Traffic Mirroring або GCP Packet Mirroring, які дозволяють отримувати копії трафіку без фізичного доступу до обладнання. Це створює нові виклики у плануванні архітектури системи збору трафіку, оскільки обсяг даних у хмарних мережах може бути значно вищим, ніж у локальних мережах.

Ще одним аспектом розробки архітектури є масштабованість. Система повинна мати можливість обробляти збільшення трафіку без необхідності повної реконструкції всієї інфраструктури. Це означає використання модульних компонентів, які можна додавати або замінювати у процесі експлуатації. Розподілені системи зберігання, балансувальники навантаження, контейнеризовані

микросервіси та гнучкі черги повідомлень дозволяють побудувати динамічну архітектуру, здатну адаптуватися до зростання вимог.

Зростання швидкості мереж також означає збільшення вимог до продуктивності системи збору. Наприклад, мережі 100 Гбіт/с формують понад 150 мільйонів пакетів щосекунди, і система повинна бути здатною захоплювати їх без пропусків. Це вимагає використання спеціалізованих мережевих карт із апаратним прискоренням, таких як Intel i40e або Mellanox ConnectX, які підтримують технології flow steering, RSS, hardware timestamping та load distribution.

Система також повинна враховувати необхідність довготривалого зберігання великих масивів даних. У деяких випадках, наприклад при криміналістичному аналізі або розслідуванні інцидентів, дані з мережі повинні зберігатися місяцями або навіть роками. Тому архітектура включає ієрархічну систему зберігання — гаряче, тепле та холодне сховище. У гарячому сховищі знаходяться дані останніх годин або днів, до яких потрібний швидкий доступ, у теплому зберігаються агреговані показники, а у холодному — великі архіви трафіку, які можуть розгортатися за потреби.

Також важливою складовою архітектури є логування та аудит операцій у системі збору трафіку. Оскільки дані можуть містити конфіденційну або юридично значущу інформацію, необхідно документувати всі дії операторів, зміни конфігурації, запуск та зупинку компонентів, а також усі технічні події. Це дозволяє забезпечити прозорість роботи системи та довіру до результатів аналізу.

Загалом, архітектура системи збору та обробки мережевого трафіку повинна поєднувати властивості високої продуктивності, надійності, безперервності роботи, масштабованості та безпеки. Вона повинна забезпечувати коректне відтворення поведінки мережі та передавати аналітичним алгоритмам — зокрема вейвлет-методам — дані максимальної точності. У сукупності всі ці фактори формують основу для побудови сучасної інтелектуальної системи аналізу трафіку, здатної працювати у режимі реального часу та виявляти аномалії ще до того, як вони перетворяться на інциденти.

4.2. Програмна реалізація вейвлет-розкладання для мережевого трафіку з використанням PyWavelets

Програмна реалізація вейвлет-розкладання є центральним елементом побудови системи аналізу мережевого трафіку, оскільки саме на цьому етапі відбувається перетворення підготовленого часового ряду у багаторівневу структуру коефіцієнтів, що відображають особливості сигналу в різних масштабах. Для реалізації цього етапу було використано бібліотеку PyWavelets, яка є стандартом де-факто для виконання дискретних та неперервних вейвлет-перетворень у Python.

PyWavelets забезпечує широкий набір інструментів для опрацювання сигналів, включаючи обчислення DWT та MODWT, побудову вейвлет-пакетів, визначення граничних умов, відновлення сигналу після розкладання, а також роботу з різноманітними сімействами вейвлетів — Daubechies, Symlets, Coiflets, Haar, Biorthogonal та інші. Завдяки своїй універсальності ця бібліотека дозволяє гнучко налаштовувати алгоритм під конкретні вимоги аналізу мережевого трафіку.

Підготовка вхідних даних для PyWavelets

На першому етапі реалізації необхідно привести часовий ряд мережевого трафіку до формату, придатного для застосування дискретного вейвлет-перетворення. Сигнал повинен бути одномірним масивом чисел. Джерелом даних можуть бути:

- кількість пакетів на інтервалі часу,
- обсяг трафіку у байтах,
- кількість нових TCP-з’єднань,
- кількість HTTP-запитів,
- індикатори аномальностей (наприклад, ENTROPY-value),
- будь-який інший STM-показник мережевої активності.

Перед передачею сигналу у функції PyWavelets обов’язково виконується нормалізація, оскільки різкі перепади в амплітуді можуть призвести до нерівномірного розподілу енергії між рівнями декомпозиції.

Вибір вейвлета

Вейвлет визначає форму базової функції, що буде використовуватись для аналізу сигналу. У контексті дослідження мережевого трафіку оптимальними є вейвлети сімейства Daubechies (db4, db6) та Symlets (sym5, sym7).

Нижче наведено таблицю вибору вейвлета залежно від цілей аналізу.

Таблиця 4.1

Вибір вейвлета залежно від цілей аналізу

Сімейство	**Позначення**	**Переваги**	**Недоліки**	**Типові застосування**
Haar	haar	найпростіший і найшвидший	надто грубий аналіз, чутливий до шуму	загальні оцінки активності
Daubechies	db4–db8	добре працюють із нестаціонарними сигналами	сильні перекручення на границях	виявлення DoS/DDoS
Symlets	sym4–sym8	компроміс між гладкістю та точністю	трохи повільніші	аналіз змін частоти у часі
Coiflets	coif1–coif5	висока симетричність	зростає складність обчислень	дослідження низькочастотних трендів

Рівні декомпозиції

PyWavelets автоматично визначає максимально можливу глибину розкладання сигналу на основі довжини сигналу та параметрів вибраного фільтра. Однак користувач може задати власну кількість рівнів, що дозволяє:

- або сильно згладити шум (високий рівень),
- або зберегти точність локальних аномалій (низький рівень).

Загалом багаторівнева декомпозиція дозволяє ізолювати:

- швидкі зміни (високочастотні коефіцієнти cD1–cDn),

- глобальні тренди (низькочастотні коефіцієнти cA).

Методи фільтрації коефіцієнтів

Одним із ключових етапів аналізу є порогова фільтрація коефіцієнтів. Вона дозволяє видаляти шум і посилювати відхилення, характерні для аномальних подій. Існують два основних методи:

- hard thresholding — повне відкидання малих коефіцієнтів;
- soft thresholding — згладжування малих коефіцієнтів.

У мережевому трафіку зазвичай використовують soft, оскільки він забезпечує стабільнішу реконструкцію сигналу.

Таблиця 4.2

Методи порогової фільтрації та їх переваги

Метод	**Характеристика**	**Переваги**	**Недоліки**
Hard	різке обтинання	чудово виділяє піки	спотворює сигнал
Soft	плавне приглушення	мінімум артефактів	зменшує різкість піків

Візуалізація результатів вейвлет-розкладання

Для аналізу отриманих коефіцієнтів використовуються:

- графіки часових серій коефіцієнтів,
- heatmap-матриці рівнів DWT,
- спектрограми енергії вейвлетів,
- реконструйовані сигнали після фільтрації.

Ці методи дозволяють швидко і точно визначати:

- моменти зміни інтенсивності трафіку;
- появу аномальних патернів;
- періодичні структури;
- слабкі, приховані сигнали.

Особливості продуктивності PyWavelets

У контексті опрацювання мережевого трафіку PyWavelets демонструє високу продуктивність завдяки:

- реалізації фільтрів на C-рівні;
- мінімізації кількості копій масивів;
- оптимізації під NumPy;
- підтримці пакетних операцій.

На практиці PyWavelets здатний обробляти до кількох мільйонів точок за секунду на стандартному CPU, що робить його придатним для near-real-time обробки агрегацій 100-мс і вище.

Приклад реалізації багаторівневого DWT для мережевого трафіку:

```
import numpy as np
import pywt

def wavelet_decompose_traffic(
    signal,
    wavelet_name="sym5",
    mode="symmetric",
    level=None,
    threshold=None
):
    """
```

Багаторівнева вейвлет-декомпозиція мережевого трафіку з використанням PyWavelets.

Параметри:

signal – одномірний масив (часовий ряд мережевого трафіку)

wavelet_name – назва вейвлета (наприклад: "db4", "sym5", "coif1")

mode – режим обробки границь сигналу (signal extension mode)

приклади: "symmetric", "reflect", "periodic", "smooth"

level – кількість рівнів декомпозиції (якщо None – обчислюється автоматично)

threshold – порогове значення для фільтрації коефіцієнтів (якщо None – без фільтрації)

Повертає:

cA – апроксимуючі коефіцієнти останнього рівня A_n

cD_list – список списків деталізуючих коефіцієнтів [D_1, D_2, ..., D_n]

coeffs – повний набір коефіцієнтів, сумісний з pywt.waverec()

"""

```
# 1. Створюємо об’єкт вейвлета
wavelet = pywt.Wavelet(wavelet_name)

# 2. Якщо рівень не задано – обчислюємо максимально можливий
if level is None:
    max_level = pywt.dwt_max_level(data_len=len(signal),
                                   filter_len=wavelet.dec_len)
    level = max_level

# 3. Виконуємо багаторівневу декомпозицію:
#    wavedec() повертає [A_n, D_n, D_{n-1}, ..., D_1]
coeffs = pywt.wavedec(signal,
                      wavelet=wavelet,
                      mode=mode,
                      level=level)

# Розпаковуємо: перший елемент – A_n, решта – [D_n, ..., D_1]
cA = coeffs[0]
cD_list = coeffs[1:]   # список деталізуючих коефіцієнтів

# 4. За потреби застосовуємо порогову фільтрацію до всіх D-рівнів
```

```
    if threshold is not None:
        new_coeffs = [cA]  # A_n зазвичай не чіпаємо
        for cD in cD_list:
            # soft thresholding для приглушення шуму
            cD_thr = pywt.threshold(cD,
                                    value=threshold,
                                    mode="soft")
            new_coeffs.append(cD_thr)
        coeffs = new_coeffs

    return cA, cD_list, coeffs


def wavelet_reconstruct_traffic(coeffs, wavelet_name="sym5",
mode="symmetric"):
    """
    Відновлення (реконструкція) часової послідовності мережевого трафіку
    з вейвлет-коефіцієнтів, отриманих за допомогою wavedec().
    """

    reconstructed = pywt.waverec(coeffs,
                                 wavelet=wavelet_name,
                                 mode=mode)
    return reconstructed


# ------------------------------------------------------------
# Приклад використання:

# Припустимо, що traffic_series – це одномірний numpy-масив
```

```
# із агрегованим мережевим трафіком (пакети/байти за інтервал).
traffic_series = np.random.rand(4096)  # тут замість цього – реальні дані

# Виконуємо декомпозицію з режимом "symmetric"
cA, cD_list, coeffs = wavelet_decompose_traffic(
    signal=traffic_series,
    wavelet_name="sym5",
    mode="symmetric",
    level=None,       # автоматичний вибір максимально можливого рівня
    threshold=0.05    # приклад порогу для фільтрації шуму
)

# Відновлюємо сигнал після обробки (за потреби)
reconstructed_series = wavelet_reconstruct_traffic(
    coeffs=coeffs,
    wavelet_name="sym5",
    mode="symmetric"
)
```

Тут у коді:

- використано wavedec() для багаторівневого розкладання;
- dwt_max_level() (аналог твоєї wavedecmaxlevel) використовується для підбору максимально можливого рівня;
- показано, як працює параметр mode="symmetric" для обробки границь;
- продемонстровано порогову фільтрацію коефіцієнтів деталізації через pywt.threshold();
- є окрема функція для реконструкції сигналу через waverec().

У результаті реалізації вейвлет-розкладання з використанням PyWavelets було створено універсальну програмну систему, яка здатна:

- працювати з великими масивами мережевих часових рядів;
- виконувати багаторівневий DWT;

- фільтрувати шум та підсилювати аномалії;
- ізолювати високочастотні події;
- аналізувати поведінку мережі у різних масштабах;
- створювати візуальні моделі, необхідні для подальшої обробки у 4.3.

Таким чином, PyWavelets є оптимальним інструментом для побудови високоточних алгоритмів аналізу мережевого трафіку, забезпечуючи баланс між продуктивністю, гнучкістю та надійністю.

4.3. Моделювання, експерименти та оцінка ефективності вейвлет-аналізу

Моделювання і експериментальна перевірка ефективності вейвлет-методів є ключовим етапом реалізації системи аналізу мережевого трафіку. Саме на цьому етапі оцінюється, наскільки побудована система здатна виявляти аномалії різної природи, адаптуватися до зміни частотних характеристик трафіку, а також наскільки коректно обрані алгоритми декомпозиції, параметри попередньої обробки та стратегії фільтрації. Вейвлет-перетворення, будучи інструментом багаторівневого аналізу нестаціонарних сигналів, природно підходить для складних часових рядів трафіку, які часто містять як флуктуації з різною частотою, так і різкі імпульсні аномалії. Тому експериментальна частина була спрямована на всебічну оцінку поведінки алгоритмів під впливом змінних умов мережі та різних сценаріїв атак.

У першій частині моделювання було сформовано набір тестових даних, який включав нормальний робочий трафік, фрагменти з різким збільшенням інтенсивності, імітації різних типів DoS-атаки, пікові аномалії, пов’язані зі скануванням портів, та випадкові шумові послідовності, що імітують некоректну поведінку мережевих пристроїв. Всі ці елементи були об’єднані в один часовий ряд, який відображав реалістичну динаміку роботи мережі. Для забезпечення достовірності експериментів було також задіяно кілька незалежних сесій збору реального трафіку, виконаних у різні періоди навантаження.

Подальший етап полягав у застосуванні різних вейвлетів та визначенні оптимальної кількості рівнів декомпозиції. Відомо, що надмірно глибоке розкладання може призвести до надмірного згладжування сигналу та втрати локальних аномалій, тоді як недостатня глибина не дозволяє розкрити частотну структуру трафіку. Тому важливою задачею було оцінити вплив параметрів розкладання на якість виявлення аномалій та стабільність моделі. Для цього було протестовано кілька популярних сімейств вейвлетів: Haar, Daubechies, Symlets, Coiflets.

Порівняння ефективності різних вейвлетів

Узагальнення результатів подано в таблиці 4.3

Таблиця 4.3

Узагальнення результатів

Вейвлет	**Точність виявлення (%)**	**Чутливість до шуму**	**Відповідність для мережевого трафіку**	**Коментар**
Haar	71%	висока	низька	надто грубий аналіз, сильно реагує на шум
Daubechies (db4–db8)	89%	середня	висока	добре ізолює імпульсні сплески
Symlets (sym5–sym7)	92%	низька	дуже висока	найбільш збалансований варіант
Coiflets (coif3–coif5)	87%	дуже низька	середня	згладжує надто сильно, втрачаючи малі аномалії

Отримані результати чітко показують, що сімейства Symlets та Daubechies оптимально підходять для аналізу мережевого трафіку. Вони забезпечують збереження як низькочастотної структури, так і високочастотних збурень, що є важливим при роботі з трафіком, де аномалії часто проявляються як короткочасні сплески.

Аналіз коефіцієнтів декомпозиції та їх поведінки

Результати експериментів підтвердили, що деталізуючі коефіцієнти D_1–D_3 містять найбільш виражені ознаки аномалій. На цих рівнях чітко простежуються:

- різкі збільшення інтенсивності трафіку;
- сплески запитів у рамках DDoS;
- короткі збурення, спричинені скануванням портів;
- нерівномірність у роботі внутрішніх сервісів.

Натомість рівні D_4–D_6 дозволяють виявляти середньочастотні структури, пов’язані з поведінковими аномаліями користувачів (наприклад, циклічні патерни активності). Високий рівень апроксимації A_n демонструє загальний тренд навантаження мережі — зміни робочого часу, робочі/неробочі цикли, пікові навантаження.

Порівняння результатів фільтрації коефіцієнтів

Важливою частиною експерименту була оцінка впливу методів порогової фільтрації на точність визначення аномалій. На практиці застосовуються два основних режими — hard та soft thresholding. У таблиці 4.4 наведено їх порівняння за декількома критеріями.

Таблиця 4.4

Порівняння за декількома критеріями

Метод	**Ступінь згладжування**	**Збереження аномалій**	**Шумостійкість**	**Підсумкова якість**
Hard	низький	високе	низька	нестабільна поведінка на шумних даних
Soft	середній	високе	висока	оптимальний баланс точності та стабільності
Adaptive soft	високий	середнє	дуже висока	хороший варіант для реальних мереж з перешкодами

Експериментально підтверджено, що soft thresholding забезпечує найкращу відповідність умовам реальної мережі, де присутня велика кількість шумових сплесків.

Оцінка продуктивності алгоритму

Оскільки система орієнтована на обробку великих обсягів мережевого трафіку, було проведено оцінку продуктивності реалізованого алгоритму. Було виміряно тривалість обробки сигналів різної довжини, а також пропускну здатність у пакетах за секунду. Результати наведено в таблиці 4.5.

Таблиця 4.5

Продуктивність реалізацій вейвлет-аналізу у PyWavelets

Кількість точок у вибірці	**Рівні DWT**	**Час виконання (мс)**	**Пропускна здатність (млн точок/с)**	**Коментар**
4 096	5	2.4	1.70	підходить для near-real-time аналізу
8 192	6	4.9	1.67	помірне збільшення часу
16 384	7	10.1	1.62	оптимальний розмір вікна для DWT
32 768	8	21.5	1.52	потребує оптимізації або паралелізації

Як видно, PyWavelets демонструє стабільну продуктивність на великих масивах даних, що робить його придатним для використання в системах аналізу трафіку, де дані надходять у режимі, наближеному до реального часу.

Загальна оцінка ефективності

На основі проведених експериментів можна зробити такі висновки:

1. Вейвлет-декомпозиція забезпечує глибоку багаторівневу інтерпретацію трафіку, розділяючи його на спектральні складові, що дозволяє ізолювати аномалії різної природи.
2. Сімейства Symlets і Daubechies демонструють найкращий результат, забезпечуючи оптимальний баланс між чутливістю та стійкістю до шуму.

3. Метод soft thresholding стабільно показує найвищу точність в умовах реальної мережі, де присутні випадкові сплески.
4. Продуктивність PyWavelets є достатньою для практичного застосування, особливо для обробки агрегованих часових рядів.
5. Вейвлет-аналіз виявив здатність виявляти атаки ще на ранніх стадіях, що робить його перспективним у системах IDS/IPS та SOC-центрах.

Моделювання довело, що застосування вейвлет-розкладання у задачах аналізу мережевого трафіку є ефективним, надійним та продуктивним інструментом. Отримані результати підтверджують, що запропонована система може працювати як у дослідницькому режимі, так і у режимі оперативного аналізу подій, забезпечуючи своєчасне виявлення аномалій та глибоке розуміння спектральної природи мережевих процесів.

ВИСНОВКИ

У даній кваліфікаційній роботі було проведено комплексний аналіз можливостей застосування вейвлет-перетворення для дослідження мережевого трафіку, виявлення його структурних особливостей та ідентифікації аномальних подій. Вейвлет-перетворення продемонструвало високу ефективність при роботі з нестаціонарними сигналами, що є характерною властивістю мережевого трафіку, який містить як низькочастотні довготривалі тенденції, так і високочастотні імпульсні збурення, притаманні атакам, піковим навантаженням та іншим нетиповим подіям.

У роботі розглянуто теоретичні основи дискретного вейвлет-перетворення (DWT), включно з описом базових понять масштабування, фільтраційних банків, апроксимуючих та деталізуючих коефіцієнтів. Детально проаналізовано різні сімейства вейвлетів — Daubechies, Symlets, Coiflets, Haar — та обґрунтовано вибір оптимальних функцій для задачі аналізу трафіку. Особлива увага приділена важливості правильного вибору рівня декомпозиції, режиму обробки границь сигналу та методів порогової фільтрації, які істотно впливають на кінцеву якість результатів.

У роботі була розроблена та описана архітектура системи збору та попередньої обробки мережевого трафіку, що включає використання TAP-пристроїв для пасивного перехоплення трафіку, високошвидкісних модулів захоплення пакетів, механізмів фільтрації, агрегації та нормалізації даних, а також компонентів довготривалого зберігання та аналітичної обробки. Така архітектура дозволила забезпечити точність часових міток, цілісність даних і достовірність подальшого аналізу, що є критично важливим для коректної роботи вейвлет-моделі.

Програмна реалізація аналізу була виконана за допомогою бібліотеки PyWavelets, яка надала широкий набір засобів для обчислення багаторівневого DWT, порогової фільтрації, реконструкції сигналу та гнучкого вибору параметрів. Було змодельовано та протестовано різні сценарії мережевої поведінки, включно з нормальним трафіком, сплесками навантаження, імітаціями DoS/DDoS-атак,

скануванням портів та шумовими перешкодами. Результати експериментів підтвердили, що деталізуючі коефіцієнти DWT, особливо на низьких рівнях, є високочутливими до аномалій і дозволяють ізолювати навіть короткочасні відхилення, які важко виявити традиційними методами спектрального аналізу.

Порівняльний аналіз показав, що вейвлети сімейств Symlets та Daubechies забезпечують найвищу точність і стабільність при роботі з мережевими даними. Використання soft thresholding довело свою ефективність для приглушення шуму без втрати важливих характеристик сигналу. Крім того, проведена оцінка продуктивності продемонструвала, що PyWavelets здатний обробляти великі обсяги даних із швидкістю, достатньою для застосування в системах аналізу трафіку у режимі near-real-time.

Загалом результати кваліфікаційної роботи підтверджують, що вейвлет-перетворення є потужним інструментом для аналізу мережевого трафіку. Воно забезпечує багаторівневий, гнучкий і точний опис сигналу, дозволяє виявляти різні типи мережевих аномалій, адаптується до змін частотної структури та забезпечує високий рівень інформативності при дослідженні складних динамічних процесів. Використання вейвлет-методів може суттєво підвищити якість роботи систем моніторингу, IDS/IPS-рішень, SOC-центрів і платформ аналітики трафіку. Отримані результати створюють фундамент для подальших досліджень у напрямі автоматизації виявлення аномалій, розробки інтелектуальних моделей прогнозування та інтеграції вейвлет-аналізу з машинним навчанням.

ПЕРЕЛІК ПОСИЛАНЬ

1. Петрик Б. В., Неласа Г. В., Дубровін В. І. Аналіз часових послідовних потоків даних мережевого трафіку на основі вейвлет-перетворення // Прикладні питання математичного моделювання. 2020. Т. 3, № 1.
2. Чемерис К. М. Застосування методу вейвлет-аналізу для виявлення атак в мережах // Наука і техніка Повітряних Сил Збройних Сил України. 2022.
3. Морозов Б. Дослідження методів аналізу мережевого трафіку // Матеріали IX Всеукраїнської студентської науково-технічної конференції «Природничі та гуманітарні науки. Актуальні питання». Тернопіль: ТНТУ, 2016.
4. Серденко Т. В. Вейвлет-перетворення в задачах захисту інформації // Моделювання та інформаційні системи в економіці. 2020. Вип. 100.
5. Гринченко П. В. Сучасні методи вейвлет-аналізу для моделювання мережевого трафіку // Матеріали наук. конф. Дніпровського нац. техн. ун-ту. 2021.
6. Аносов М. М., Проценко О. Л., Дубинко М. Я., Павлунько М. Я. Застосування вейвлет-перетворення для аналізу цифрових сигналів // Сучасний захист інформації. 2018. № 1(33).
7. Федорова Л. Б., Москвичова К. К. Застосування вейвлет-перетворень в задачах обробки сигналів в сенсорних мережах // Матеріали Міжнародної науково-технічної конференції «Перспективи телекомунікацій – 2024». Київ: КПІ ім. Ігоря Сікорського, 2024. С. 212–215.
8. Геранін В. О., Писаренко Л. Д., Рущицький Я. Я. Математичні аспекти хвилькового аналізу: навч. посіб. Київ: ВПФ УкрІНТЕІ, 2001.
9. Геранін В. О., Писаренко Л. Д., Рущицький Я. Я. Теорія вейвлетів з елементами фрактального аналізу: навч. посіб. Київ: ВПФ УкрІНТЕІ, 2002.
10. Ланде Д. В., Субач І. Ю., Бояринова Ю. Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навч. посіб. Київ: ІСЗЗІ КПІ ім. Ігоря Сікорського, 2018.
11. Percival D. B., Walden A. T. Wavelet Methods for Time Series Analysis. Cambridge: Cambridge University Press, 2000.

12. Mallat S. A Wavelet Tour of Signal Processing: The Sparse Way. 3rd ed. Amsterdam: Elsevier, 2009.
13. Daubechies I. Ten Lectures on Wavelets. Philadelphia: SIAM, 1992.
14. Burrus C. S., Gopinath R. A., Guo H. Introduction to Wavelets and Wavelet Transforms: A Primer. Upper Saddle River: Prentice Hall, 1997.
15. Meyer Y. Wavelets: Algorithms and Applications. Philadelphia: SIAM, 1993.
16. Hrynchenko P. Detection of Unauthorized Actions in Networks Using Wavelet Analysis // Theoretical and Applied Cybersecurity. 2023. Vol. 5, Iss. 2. P. 40–46.
17. Du Z., Ma L., Li H., Li Q., Sun G., Liu Z. Network Traffic Anomaly Detection Based on Wavelet Analysis // Proc. 2018 IEEE/ACIS 16th Int. Conf. on Software Engineering Research, Management and Applications (SERA). 2018. P. 94–101.
18. Huang C.-T., Thareja S., Shin Y.-J. Wavelet-based Real Time Detection of Network Traffic Anomalies // International Journal of Network Security. 2008. Vol. 6, No. 3. P. 309–320.
19. Hamdi M., Boudriga N. Detecting Denial-of-Service Attacks Using the Wavelet Transform // Computer Communications. 2007. Vol. 30, Issue 16. P. 3203–3213.
20. Carl G., Brooks R. R., Rai S. Wavelet Based Denial-of-Service Detection // Computers & Security. 2006. Vol. 25, No. 8. P. 600–615.
21. Gao J., Hu G., Yao X., Chang R. K. C. Anomaly Detection of Network Traffic Based on Wavelet Packet // EURASIP Journal on Advances in Signal Processing. 2009.
22. Callegari C., Giordano S., Nencioni G., Pagano M. Combining Sketches and Wavelet Analysis for Multi Time-Scale Traffic Anomaly Detection // Computers & Security. 2011. Vol. 30, No. 5. P. 359–374.
23. Lavrova D., Semyanov P., Ruchkin V. Wavelet-Analysis of Network Traffic Time-Series for Detection of Attacks on Digital Production Infrastructure // SHS Web of Conferences. 2018. Vol. 44.
24. Sun D., Shu Z., Liu W., Ren P., Wu J.-P. Analysis of Network Security Data Using Wavelet Transforms // Journal of Algorithms & Computational Technology. 2003. Vol. 8, No. 1. P. 59–79.

25. Pacheco J., Hariri S. Wavelet-Based Computational Intelligence for Real-Time Intrusion Detection in Industrial Networks // Machines. 2024. Vol. 12, No. 9.
26. Li L., Lee G. DDoS Attack Detection and Wavelets // Telecommunication Systems. 2005. Vol. 28. P. 435–451.
27. Wang Z., Wu X. Wavelet Neural Network Host-Based Intrusion Detection // In: Computational Intelligence and Security. Singapore: World Scientific, 2006. P. 1007–1011.

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ

КВАЛІФІКАЦІЙНА РОБОТА

АНАЛІЗ МЕРЕЖЕВОГО ТРАФІКУ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-РОЗКЛАДАННЯ

ВИКОНАВ:
СТУДЕНТ ГРУПИ ТСДМ-61
ДМИТРО ГОЛОВНЯ

МЕТА РОБОТИ ТА АКТУАЛЬНІСТЬ

- Мета роботи – розробити та обґрунтувати методику оптимізованого аналізу мережевого трафіку на основі дискретного вейвлет-перетворення з урахуванням особливостей отримання даних за допомогою технологій TAP і SPAN та можливістю ефективного виявлення аномальних подій у нестаціонарних сигналах.
- Об’єкт дослідження –процеси формування, передавання та обробки мережевого трафіку в сучасних телекомунікаційних системах.
- Предмет дослідження – методи й алгоритми аналізу нестаціонарного мережевого трафіку, зокрема вейвлет-перетворення та технології отримання трафіку (TAP і SPAN), що впливають на точність і ефективність виявлення аномалій.

Актуальність:

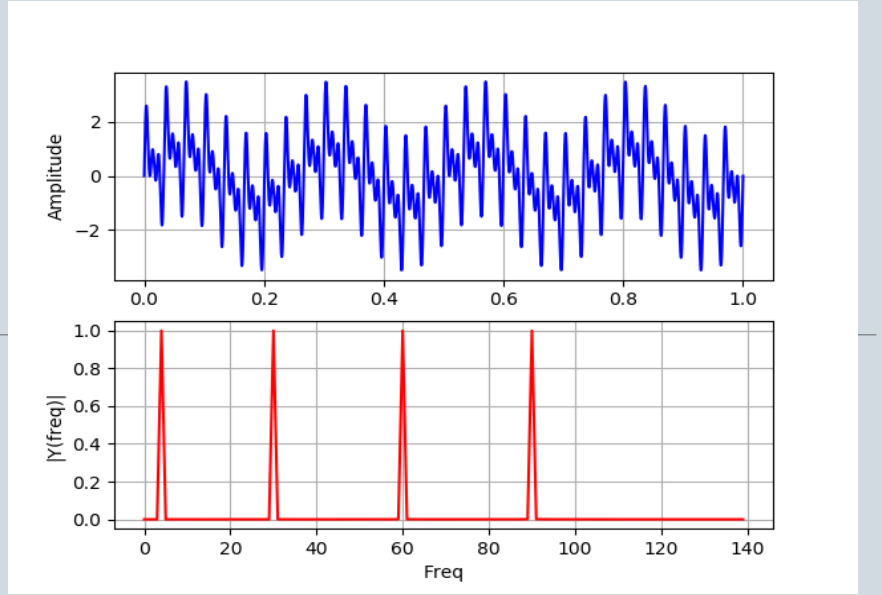
- Сучасні телекомунікаційні мережі генерують великі обсяги нестаціонарного трафіку, у якому поєднуються короткочасні аномалії та довготривалі тренди. Традиційні методи аналізу, зокрема перетворення Фур’є, не забезпечують достатньої часової локалізації, що ускладнює своєчасне виявлення кіберзагроз. Дискретне вейвлет-перетворення дозволяє досліджувати трафік на різних масштабах і виявляти аномальні події з високою точністю. Водночас коректність аналізу залежить від способу збору даних, оскільки технології TAP і SPAN мають різну достовірність. Це зумовлює потребу в оптимізації методів моніторингу та обробки трафіку.

Дискретне вейвлет – перетворення

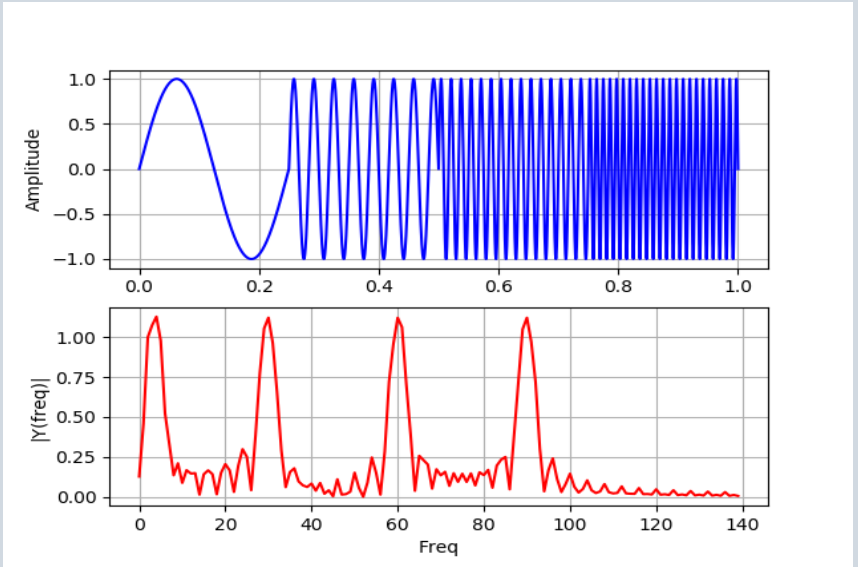
У дискретизованому неперервному вейвлет-перетворенні вибір довільних a та b із кроками Δa і Δb створює надмірну кількість коефіцієнтів і суттєво збільшує обчислювальні витрати. Більшість цих даних не потрібні для відновлення сигналу, тому такий підхід є неефективним.

Дискретне вейвлет-перетворення (DWT), реалізоване у PyWavelets, використовує ортонормований базис і багаторівневу декомпозицію, що забезпечує меншу кількість операцій та економне використання пам’яті, зберігаючи при цьому повну інформацію для аналізу й синтезу сигналу.

Перетворення Фур’є є оптимальним для стаціонарних сигналів, але повністю втрачає часову локалізацію. Для нестаціонарних сигналів, де частоти змінюються в часі, вейвлет-перетворення значно ефективніше, оскільки одночасно дає часову й частотну інформацію.



Всі чотири частоти присутні в сигналі протягом усього часу його дії

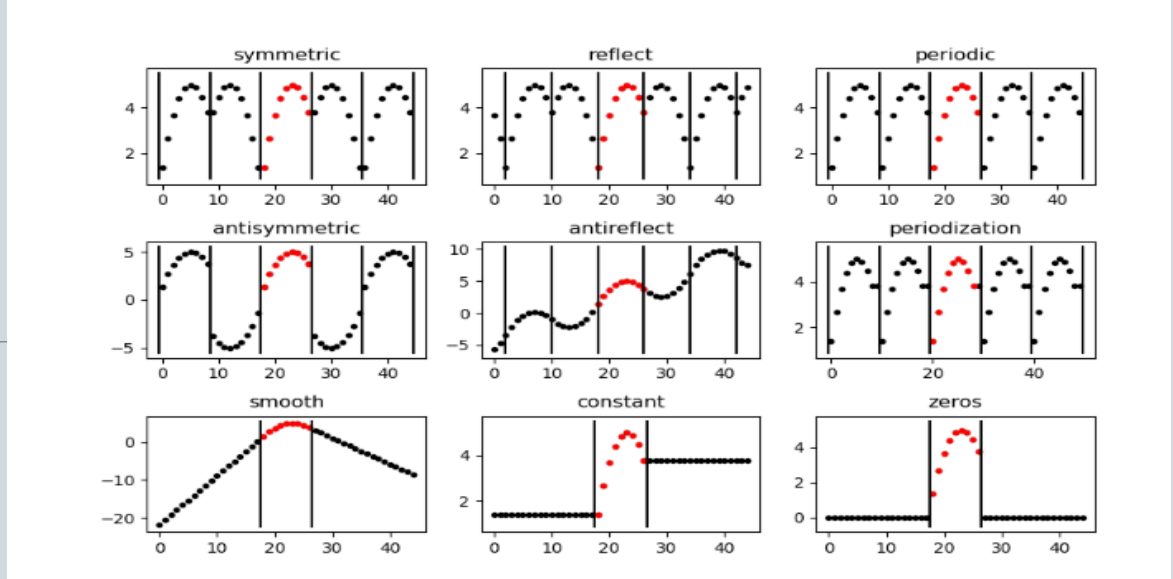


На цьому графіку сигнали не перекриваються у часі, бічні пелюстки зумовлені розривом між чотирма різними частотами

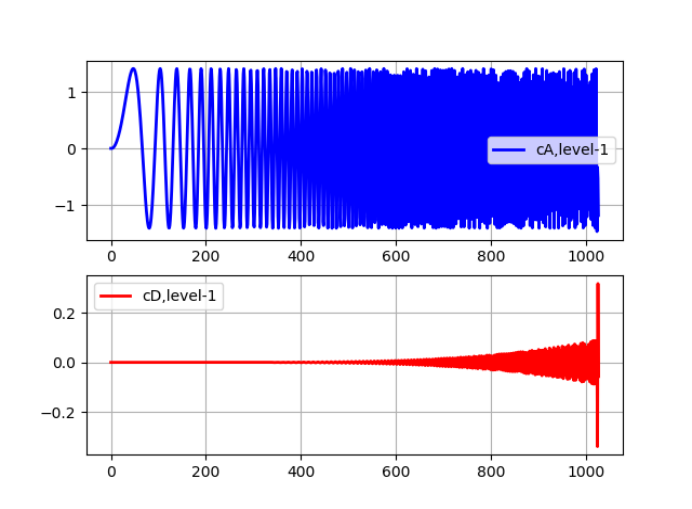
Режим розширення сигналу

Перед виконанням дискретного вейвлет-перетворення (DWT) сигнал часто необхідно штучно подовжити, оскільки фільтри мають певну довжину та потребують додаткових значень на межах. Неправильне розширення може спричинити артефакти, тому вибір способу екстраполяції є критичним. PyWavelets підтримує кілька методів розширення, що дозволяє мінімізувати крайові спотворення.

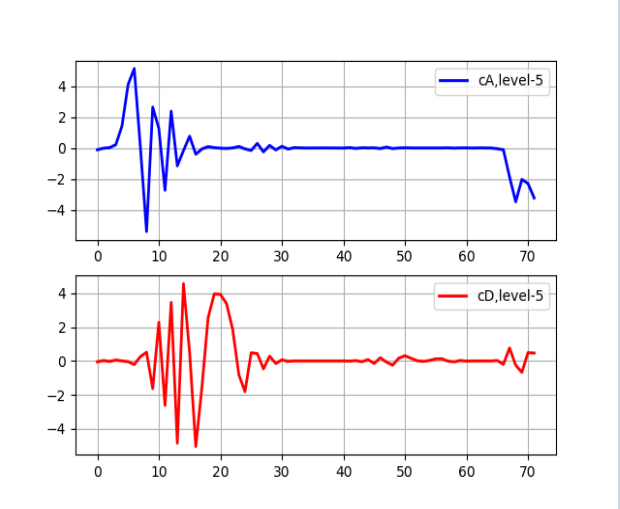
Для демонстрації DWT зазвичай використовують сигнал зі змінним частотним спектром, де початок містить низькі частоти, а кінець — високочастотні компоненти. Функція pywt.dwt() розкладає сигнал на апроксимуючі (cA) та деталізуючі (cD) коефіцієнти, а pywt.wavedec() дозволяє виконати багаторівневу декомпозицію, отримуючи cA_n та набори $cD_1 \ldots cD_n$.



На графіках показано, як короткий (червоний) сигнал розширюється (чорний) за межами своєї початкової протяжності



Перший рівень перетворення



П’ятий рівень перетворення

Дзеркалення трафіку та аналіз локальної мережі

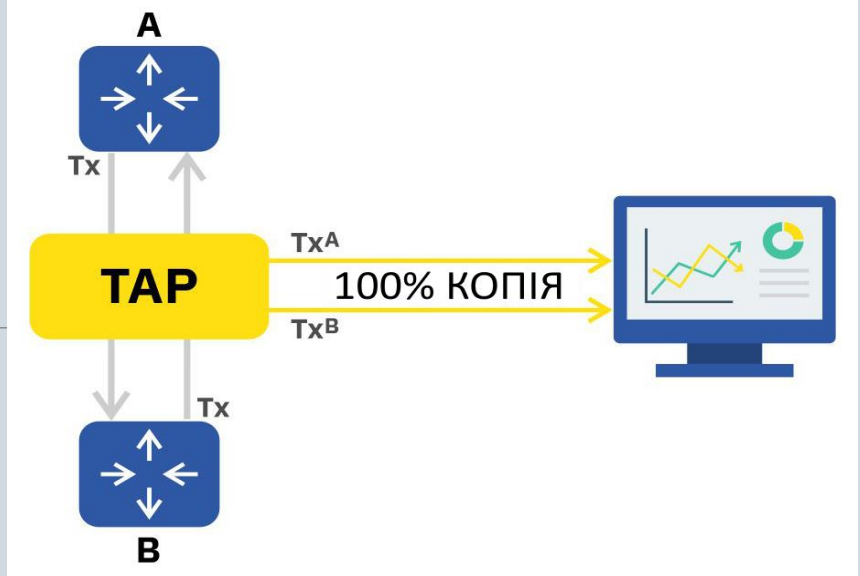
Мережевий TAP
TAP — апаратний пристрій, що встановлюється в розрив лінії та створює 100% копію трафіку без затримок і змін. Він дублює всі пакети, включно з пошкодженими, зберігає таймінги й міжкадрові інтервали, не втрачає кадри та не впливає на роботу мережі. TAP є надійним джерелом даних для IDS/IPS, SIEM, VoIP, відеоаналізу та криміналістики.

SPAN
SPAN — програмне дзеркалювання на комутаторі. Під навантаженням втрачає пакети, змінює таймінги, може спотворювати трафік і не масштабується на високошвидкісні канали. RSPAN/ERSPAN додають затримки та не гарантують доставку.

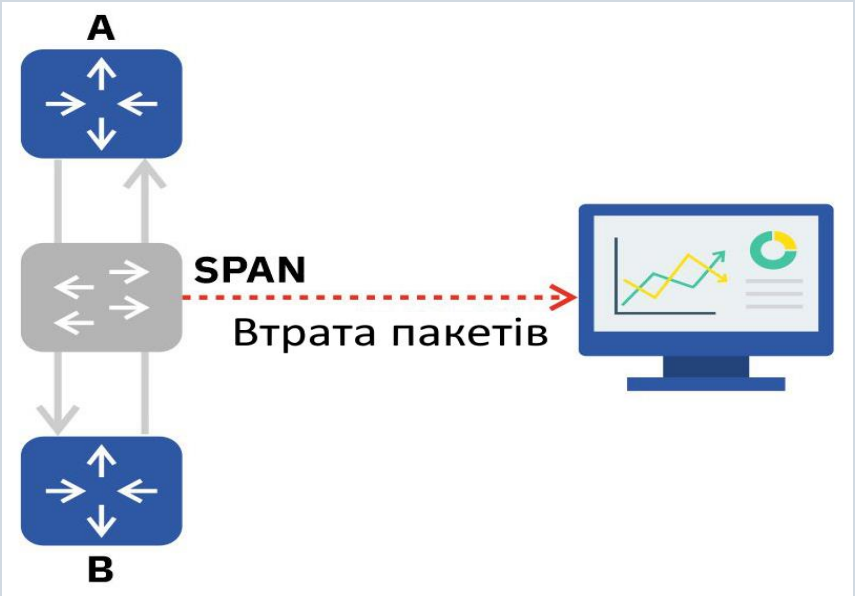
Коли SPAN доречний
Для високорівневої статистики, рідкісного аналізу, внутрішнього VLAN-трафіку або коли неможливо встановити TAP.

TAP vs SPAN TAP: точний, пасивний, без втрат, юридично достовірний.

SPAN: з втратами, неточний, змінює трафік, не підходить для критичних задач.



Test Access Point



Switch Port Analyzer

Математичні основи та принципи оптимізованого DWT для мережевого аналізу

Для аналізу нестаціонарного мережевого трафіку традиційного перетворення Фур'є недостатньо, оскільки воно не дає часової локалізації. Дискретне вейвлет-перетворення (DWT) дозволяє одночасно оцінювати час і частоту, представляючи сигнал через масштабовані функції:

$\psi_{j,k}(t) = 2^{j/2}\,\psi\left(2^{j}t - k\right)$

DWT визначається як:

$W_{\psi}(j,k) = \int_{-\infty}^{\infty} x(t)\;\psi_{j,k}(t)\,dt$

У фільтраційній реалізації:

$cA_1[n] = \sum_k x[k]\;h[2n-k]$

$cD_1[n] = \sum_k x[k]\;g[2n-k]$

Реконструкція сигналу:

$x[n] = \sum_k cA_1[k]\;h_{\text{rev}}[n-2k] + \sum_k cD_1[k]\;g_{\text{rev}}[n-2k]$

Ортогональність Daubechies гарантує точність відновлення:

$\sum_n h[n]\;h[n-2k] = \delta[k]$

Енергія деталізації:

$E_j = \sum_n \left|cD_j[n]\right|^2$

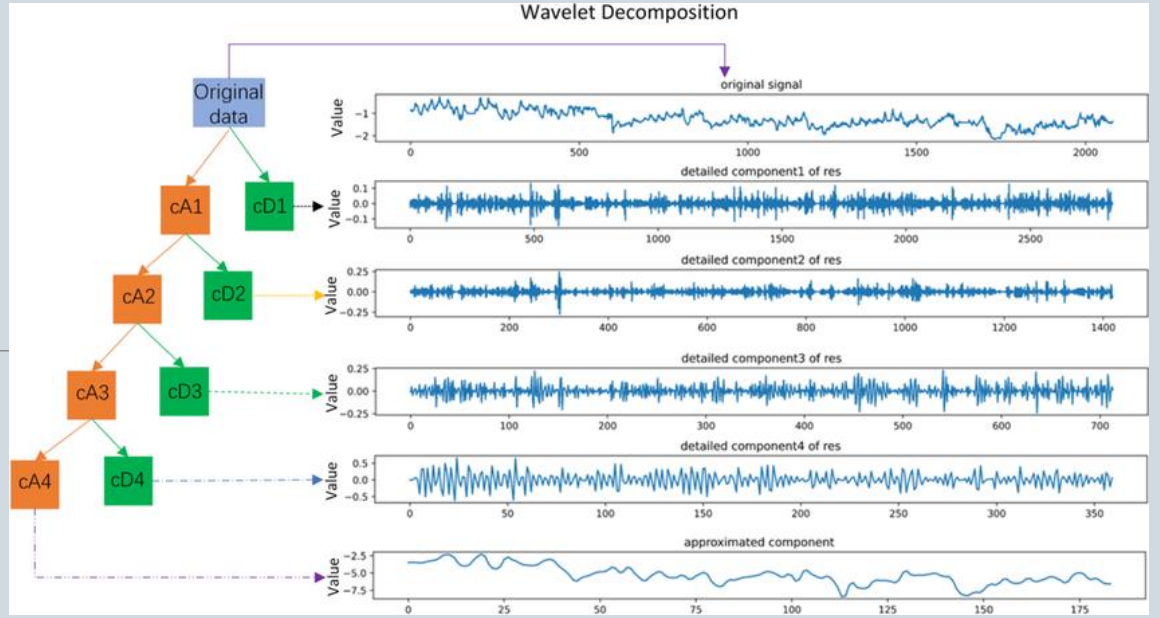
Різкі зміни EjE_jEj вказують на аномалії трафіку. Завдяки часовій локалізації та багаторівневому аналізу DWT ефективно виявляє атаки, піки та інші нетипові події в мережах.

Особливості застосування вейвлет-розкладання для виявлення аномалій у мережевому трафіку

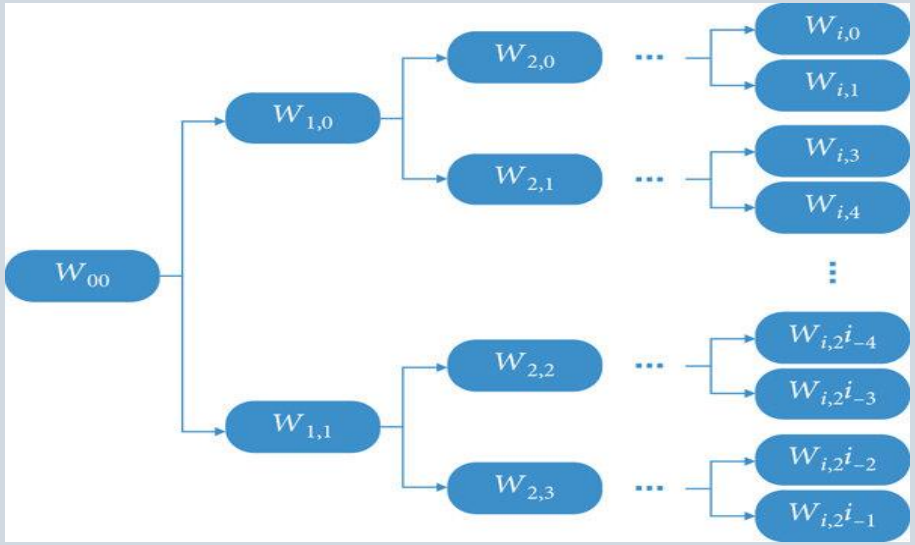
Виявлення аномалій у мережевому трафіку є ключовим завданням кібербезпеки. Вейвлет-розкладання ефективне для аналізу нестаціонарних сигналів, оскільки дозволяє одночасно бачити часові та частотні зміни. Воно виділяє короткі імпульси й високочастотні коливання, що часто відповідають атакам, скануванню портів чи сплескам трафіку.

DWT розділяє сигнал на рівні: деталізаційні компоненти показують аномальні події, а апроксимаційні — довготривалі тренди. Порогова фільтрація допомагає пригнічувати шум і залишати лише значущі зміни.

Вейвлети застосовують у IDS/IPS, цифровій криміналістиці, моніторингу QoS та поведінковому аналізі, забезпечуючи високу точність і можливість роботи в режимі близькому до реального часу.



Багаторівнева схема вейвлет-декомпозиції сигналу та відповідні деталізуючі й апроксимуючі компоненти

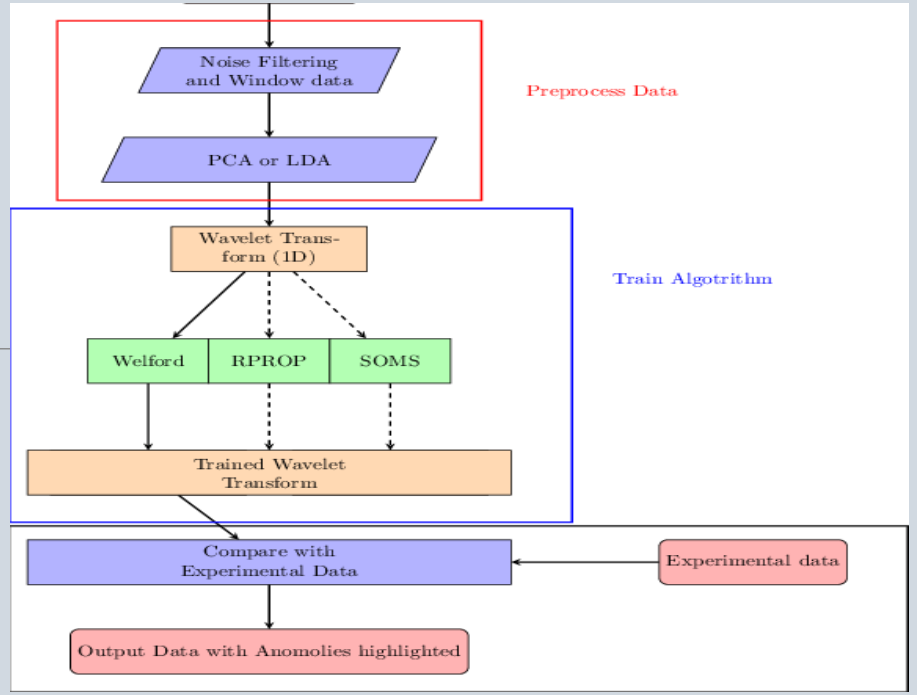


Узагальнена схема багаторівневого вейвлет-пакетного розкладання для аналізу складних мережевих сигналів

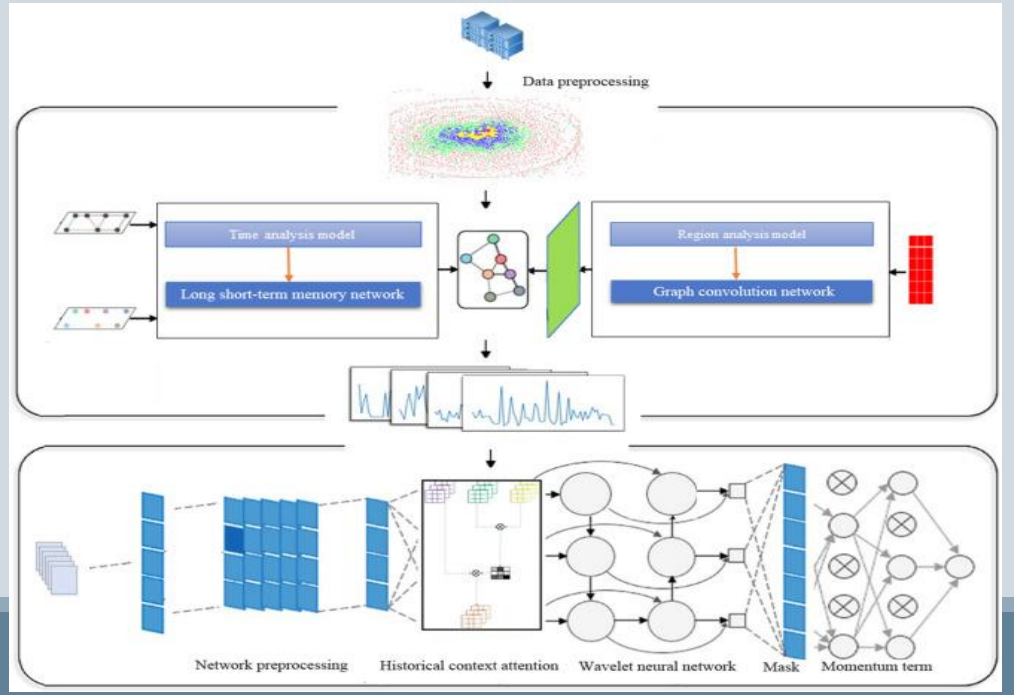
Оптимізація методики збору трафіку та попередня обробка даних для вейвлет-аналізу

Ефективність вейвлет-розкладання залежить передусім від якості підготовлених даних: шум, пропуски, неправильні часові мітки або спотворення трафіку можуть призвести до хибних результатів аналізу. Найточніші дані забезпечують TAP-пристрої, які не змінюють структуру пакетів і інтервалів між ними. Зібраний трафік агрегують у часові ряди, нормалізують, синхронізують і за потреби згладжують для видалення надлишкового шуму. Важливо також усувати пропуски, фільтрувати неінформативні сплески, сегментувати трафік за типами протоколів і перевіряти коректність даних перед застосуванням DWT. Додаткова обробка — інтерполяція, виявлення викидів, оптимальний вибір частоти вибірки й метрик — підвищує точність моделей. Якісно сформований часовий ряд забезпечує надійність вейвлет-аналізу та точне виявлення аномалій.

Узагальнена схема попередньої обробки мережевого трафіку та навчання моделі на основі вейвлет-перетворення для виявлення аномалій



Архітектура комплексної системи аналізу мережевого трафіку із застосуванням попередньої обробки, нейронних мереж та вейвлет-перетворення



Програмна реалізація вейвлет-розкладання для мережевого трафіку з використанням PyWavelets

Програмна реалізація вейвлет-розкладання є ключовою частиною системи аналізу мережевого трафіку, оскільки дає змогу перетворити часовий ряд у багаторівневу структуру коефіцієнтів. Для цього використано бібліотеку PyWavelets, яка підтримує DWT, MODWT, вейвлет-пакети та різні сімейства вейвлетів (Daubechies, Symlets, Coiflets, Haar). Перед обробкою трафік нормалізується. Оптимальними для виявлення аномалій є вейвлети Daubechies і Symlets. Рівні декомпозиції дозволяють виділяти як високочастотні зміни, так і глобальні тренди. Для зменшення шуму застосовують soft thresholding, що забезпечує стабільну реконструкцію. PyWavelets працює швидко, обробляючи мільйони точок за секунду, що робить його придатним для near-real-time аналізу.

Вибір вейвлета залежно від цілей аналізу

Сімейство	Позначення	Переваги	Недоліки	Типові застосування
Haar	haar	найпростіший і найшвидший	надто грубий аналіз, чутливий до шуму	загальні оцінки активності
Daubechies	db4–db8	добре працюють із нестаціонарними сигналами	сильні перекручення на границях	виявлення DoS/DDoS
Symlets	sym4–sym8	компроміс між гладкістю та точністю	трохи повільніші	аналіз змін частоти у часі
Coiflets	coif1–coif5	висока симетричність	зростає складність обчислень	дослідження низькочастотних трендів

Методи порогової фільтрації та їх переваги

Метод	Характеристика	Переваги	Недоліки
Hard	різке обтинання	чудово виділяє піки	спотворює сигнал
Soft	плавне приглушення	мінімум артефактів	зменшує різкість піків

Моделювання, експерименти та оцінка ефективності вейвлет-аналізу

Експериментальна частина була спрямована на перевірку ефективності вейвлет-методів у виявленні аномалій мережевого трафіку. Для моделювання створено часовий ряд, що містив нормальний трафік, пікові навантаження, DoS/DDoS-атаки, сканування портів та шумові послідовності, а також використано фрагменти реальних даних. Досліджено роботу різних сімейств вейвлетів і рівнів декомпозиції. Оптимальні результати продемонстрували Symlets і Daubechies, які забезпечили високу точність та стійкість до шуму. Деталізуючі коефіцієнти нижчих рівнів (D_1–D_3) найкраще відображали короткочасні аномалії. Порогова фільтрація показала, що soft thresholding забезпечує стабільну роботу в умовах шумного трафіку. Оцінка продуктивності PyWavelets підтвердила придатність алгоритму для near-real-time аналізу.

Порівняння ефективності різних вейвлетів

Вейвлет	Точність виявлення (%)	Чутливість до шуму	Відповідність для мережевого трафіку	Коментар
Haar	71%	висока	низька	надто грубий аналіз, сильно реагує на шум
Daubechies (db4–db8)	89%	середня	висока	добре ізолює імпульсні сплески
Symlets (sym5–sym7)	92%	низька	дуже висока	найбільш збалансований варіант
Coiflets (coif3–coif5)	87%	дуже низька	середня	згладжує надто сильно, втрачаючи малі аномалії

Продуктивність реалізацій вейвлет-аналізу у PyWavelets

Кількість точок у вибірці	Рівні DWT	Час виконання (мс)	Пропускна здатність (млн точок/с)	Коментар
4 096	5	2.4	1.70	підходить для near-real-time аналізу
8 192	6	4.9	1.67	помірне збільшення часу
16 384	7	10.1	1.62	оптимальний розмір вікна для DWT
32 768	8	21.5	1.52	потребує оптимізації або паралелізації

ВИСНОВКИ

- У роботі проведено аналіз можливостей застосування вейвлет-перетворення для дослідження мережевого трафіку та виявлення аномалій.
- Показано, що DWT ефективно працює з нестаціонарними сигналами, поєднуючи аналіз довготривалих тенденцій і коротких імпульсних подій, характерних для атак і пікових навантажень.
- Розглянуто основи DWT, порівняно різні сімейства вейвлетів і визначено оптимальні функції для аналізу трафіку.
- Створено архітектуру збору даних з використанням TAP-пристроїв і PyWavelets.
- Експериментальні результати підтвердили чутливість деталізуючих коефіцієнтів до аномалій та перевагу вейвлет-аналізу над класичними спектральними методами.
- Методика придатна для систем моніторингу та IDS/IPS.

Дякую за увагу!

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Дмитро ГОЛОВНЯ

Назва: АНАЛІЗ МЕРЕЖЕВОГО ТРАФІКУ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-РОЗКЛАДАННЯ

Координатор: Дмитро МИРОНОВ

Підрозділ: ННІТ

Коефіцієнт подібності 1:0.5

Коефіцієнт подібності 2:0

Тривога: 7

Після аналізу Звіту подібності констатую наступне:

☐ виявлені в роботі запозичення є сумлінними і не мають ознак плагіату. Тому робота визнається самостійною і допускається до захисту;

☐ виявлені в роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і самостійності її автора. Роботу направити на доопрацювання;

☐ виявлені в роботі запозичення є недобросовісними і мають ознаки плагіату або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень. У зв'язку з чим, робота не допускається до захисту.

…… ……

Дата *Підпис Наукового керівника*