УДК 681.325

І.І. Лагун[1], А.Е. Лагун[2]
Національний університет "Львівська політехніка",
[1]кафедра комп'ютеризованих систем автоматики,
[2]кафедра безпеки інформаційних технологій

ВИКОРИСТАННЯ ДИСКРЕТНОГО МАЛОХВИЛЬОВОГО ПЕРЕТВОРЕННЯ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ МЕРЕЖЕВОГО ТРАФІКУ

Проаналізовано основні методи виявлення аномалій мережевого трафіку, досліджено механізми виявлення аномалій, встановлено особливості застосування властивостей порогу виявлення аномалій при статистичному аналізі, досліджено аномалії мережевого трафіку після завершення нападу та в режимі реального часу за допомогою дискретного малохвильового перетворення.

Ключові слова: дискретне малохвильове перетворення, аномалія мережевого трафіку, статистичний аналіз.

The article analyzes the main methods of detecting network traffic anomalies, a study anomaly detection mechanisms, the characteristics of the application properties of threshold detection of anomalies in the statistical analysis, and also the anomalies of network traffic after the attack and in real time using the discrete wavelet transform.

Key words: discrete wavelet transform, anomaly network traffic, statistical analysis.

Вступ. Розуміння природи аномалій трафіку у мережі є важливою задачею. Незалежно від того, шкідливими чи ні є аномалії, важливо проаналізувати їх з двох причин:

- аномалії можуть створювати перевантаження в мережі і підвищити використання ресурсів маршрутизаторів, що робить їх виявлення вкрай важливим.
- деякі аномалії не обов'язково впливають на мережу, але вони можуть мати серйозний вплив на клієнта або кінцевого користувача.

Істотна проблема діагностування аномалій полягає в тому, що їхні форми можуть змінюватися, залежно від причини: від DoS-атак (відмови в обслуговуванні) до неправильних конфігурацій маршрутизатора.

Незважаючи на велику кількість літератури, в якій описуються основні характеристики трафіку, аномалії трафіку залишаються недостатньо вивченими. Є багато причин цього. Однією з них є необхідність використання складної інфраструктури моніторингу для ідентифікації аномалій, а також інструментів для обробки вимірювань, які достатньо швидкі, щоб виявити аномалії в режимі реального часу. Інша причина полягає в тому, що характер мережевого трафіку є багатовимірним і зашумленим, що заважає отримати корисну інформацію про аномалії від статистичних даних трафіку.

Основні методи виявлення аномалій мережевого трафіку. Трафік в маршрутизаторах, особливо маршрутизаторах магістральних мереж, є дуже великим і змінюється безперервно, тоді як аномальний трафік є маленьким порівняно з нормальним трафіком і змінами нормального трафіку. Основна мета виявлення аномалії полягає в тому, щоб виявити відносно маленький трафік аномалії у відносно великому фоновому трафіку. Тому швидке і точне виявлення аномалій трафіку є однією з умов безпечної ефективної роботи мережі.

Методи виявлення аномалій поділяються на дві категорії: виявлення на основі сигнатур й виявлення на основі статистики. Сигнатурний метод виявляє аномалію за відомими ознаками – «сигнатурами». Недоліком цього методу є властивість виявляти відомі заздалегідь типи аномалій. Тому цей метод не можна застосувати для ідентифікації невідомих аномалій. За статистичними методами виявлення аномалій трафіку використовують інший підхід, який визначає "нормальну" мережеву активність, і виділяє все, що виходить за межі норми як аномальність. Перевага цього підходу полягає в тому, що він не вимагає попередніх знань про властивості аномалій і тому може бути ефективний з невідомими аномаліями і навіть зміною існуючих відомих аномалій. Одним із статистичних методів виявлення є метод, що ґрунтується на основі малохвильового аналізу.

Цей метод є одним із останніх сучасних інструментів моделювання з використанням як нестаціонарних та довгострокових залежностей, так і для аналізу властивостей рядів даних. Він використовує малохвильове перетворення для аналізу трафіку та може виявляти масштабні властивості одночасно часової та частотної динаміки на відміну від перетворення Фур'є.

Одним з методів, за якими використовують малохвильове перетворення для виявлення аномалій трафіку, є метод з кореляцією адрес для певної кількості точок відбору. Сигнал може бути визначений на відповідному часовому масштабі та на певних позиціях часового масштабу, крім того, існує можливість відображення частотних та часових компонент одночасно. Також малохвильове перетворення відіграє роль узгоджувального фільтру для синхронізації між інформаційним сигналом та сигналом шуму з максимальним відношенням сигнал-шум (SNR) у системі зв'язку.

Для кращого розуміння наведеного в статті дослідження проведемо короткий аналіз дискретного малохвильового перетворення (ДМП). ДМП складається з розкладу (аналізу) та реконструкції (синтезу) сигналу. Рис. 1 ілюструє принцип роботи багаторівневого одновимірного малохвильового аналізу з використанням спеціальних фільтрів розкладу Lo_D (низькочастотний) та реконструкції Hi_D (високочастотний) вихідного сигналу [5].

ДМП сигналу S одержують застосуванням набору фільтрів. Спочатку сигнал пропускається через низькочастотний фільтр Lo_D із імпульсним відгуком g. Результатом буде згортка

$$S_{low}[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n-k]. \quad (1)$$

Одночасно сигнал розкладається за допомогою високочастотного фільтра Hi_D.

$$S_{high}[n] = (x * h)[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k]. \quad (2)$$

У результаті отримуються коефіцієнти деталізації (після ВЧ-фільтра) і коефіцієнти апроксимації (після НЧ-фільтра). Ці два фільтри зв'язані між собою й називаються квадратурними дзеркальними фільтрами (QMF).

Оскільки половину частотного діапазону сигналу було відфільтровано, то за теоремою Котельникова, відліки сигналів можна прорідити удвічі. Таке розкладання вдвічі зменшує розподіл за часом через проріджування сигналу. Однак кожний із вихідних сигналів представляє половину частотної смуги початкового сигналу, так що частотний розподіл подвоївся.

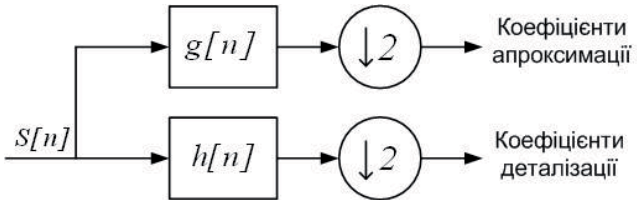

Рис. 1. Схема розкладання сигналу у ДМП

Це розкладання можна повторити кілька разів для подальшого збільшення частотного дозволу, з подальшим проріджуванням коефіцієнтів після НЧ- і ВЧ-фільтрації. Також можна

представити у вигляді двійкового дерева, де листки й вузли відповідають просторам з різною частотно-часовою локалізацією. Це дерево представляє структуру банку фільтрів.

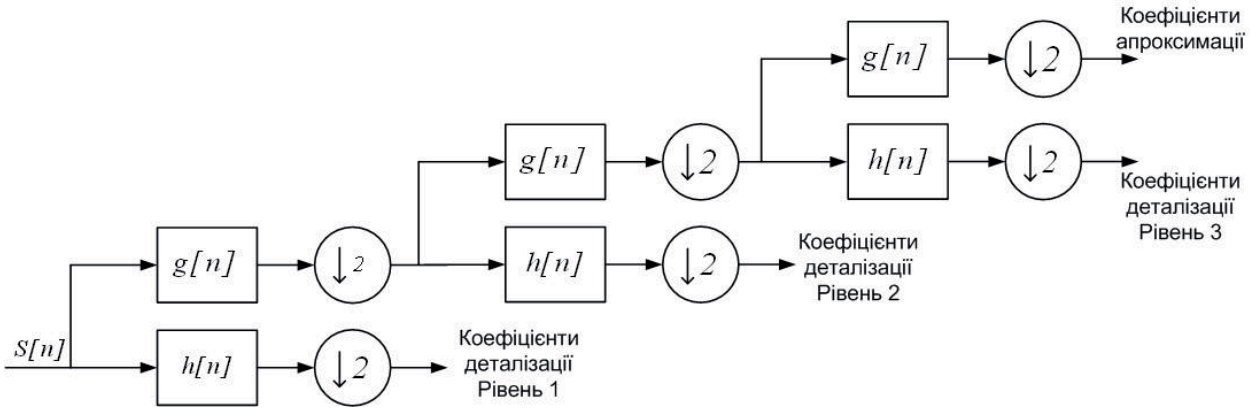

Рис. 2. Трирівневий банк фільтрів

На кожному рівні вищенаведеної діаграми сигнал розкладається на низькі й високі частоти. У силу дворазового проріджування, довжина сигналу повинна бути кратна 2n, де n — число рівнів розкладу.

Отже, коефіцієнти розкладу мають вигляд:

$$cD_{1,t} = \sum_{n=0}^{N-1} (H_i \_ D) S_{2t+l \;\; n \bmod T}, \quad cA_{l,t} = \sum_{n=0}^{N-1} (L_0 \_ D) S_{2t+l \;\; n \bmod T}$$

$$cD_{2,t} = \sum_{n=0}^{N-1} (H_i \_ D) cA_{1,2t+l \;\; n \bmod T}, \quad cA_{2,t} = \sum_{n=0}^{N-1} (L_0 \_ D) cA_{1,2t+l \;\; n \bmod T} \tag{3}$$

.....

$$cD_{l,t} = \sum_{n=0}^{N-1} (H_i \_ D) cA_{j \;\; 1,2t+l \;\; n \bmod T}, \quad cA_{j,t} = \sum_{n=0}^{N-1} (L_0 \_ D) cA_{j \;\; 1,2t+l \;\; n \bmod T}$$

Низькочастотна складова сигналу позначається буквою L, а високочастотна – H.

Для реконструкції, починаючи від двох наборів коефіцієнтів на рівні *j*, тобто коефіцієнтів апроксимації cA_j і коефіцієнтів деталізації cD_j, зворотне ДМП проводить інтерполяцію сигналів на кожному рівні удвічі, зворотну децимацію та згортку отриманих масивів з фільтрами реконструкції Lo_R і Hi_R. Нехай L-довжина cA і cD, і N-довжина фільтрів Lo_R і Hi_R, тоді довжина сигналу становить *2*L-N+2*. Для дискретного сигналу довжиною T ДМП може складатися не більше ніж з log_2T рівнів.

Дослідження механізмів виявлення аномалій. Відновлений сигнал використовується для виявлення аномалій. Аналіз та виявлення проводяться з наборами даних протягом тривалого періоду часу. Однак, в режимі реального часу механізми аналізу та виявлення працюють з невеликими наборами даних. Але при використанні невеликих наборів даних зростає можливість збільшення помилкових тривог. Водночас, великі набори даних збільшують затримку виявлення тоді, коли аналіз таких наборів проводиться в реальному часі.

Для зняття цих обмежень застовується метод рухомого вікна для підвищення швидкості виявлення при одночасному зниженні помилкових тривог.

На рис. 3 зображено часову діаграму механізму виявлення аномалій, де D – тривалість нападу у ДМП-сигналі; W – розмір вікна детектування (DET) (W=q·t, де t – інтервал вибірки); T – максимальний час індикації; m – величина фактора для прийняття рішення (як правило, 1/2).

Для кожної вибірки створюється кореляція сигналу S(n). Розглядаються p вибірок, S(n-p+1), S(n-p+2), ..., S(n-1) та S(n) для розрахунку ДМП в точці вибірки n. Оголошується p, проводиться

аналізується вікно (ДМП). Для виявлення аномалій відбирається q (≤ p) вибірок, S(n-q+1), S(n-q+2), ..., S(n-1) та S(n). Оголошується q, проводиться детектування (DET) вікна. Щоб зменшити кількість помилкових тривог через миттєві шуми, використовується кілька вікон детектування при виявленні аномалій. Якщо значення вибірок для вікон детектування в q/2 або більше разів вище за поріг аномалії, вважається, що аномалія виявлена на вибірці в точці n. Більшість детекторів для успішного виявлення вимагає, щоб представлений трафік характеризувався аномальною поведінкою протягом кількох вибірок (принаймні q/2 у вікні q вибірок). При великому q рівень помилкових тривог є достатньо низьким. Проте велике значення q підвищує латентність виявлення аномалій, оскільки більшість таких функцій мають затримку виявлення атак, принаймні q/2 періоду вибірки. У результаті напади тривалістю меншою, ніж q/2 періоду вибірки, можуть бути не виявлені. Часові характеристики наведеного механізму представляються виразами:

$$D \geq m \cdot W;$$
$$T = D + (1 - 2 \cdot m) \cdot W, \text{ для } 0 < m < \min(1, D/W) \quad (4)$$
$$|W - D| \leq T < W + D.$$

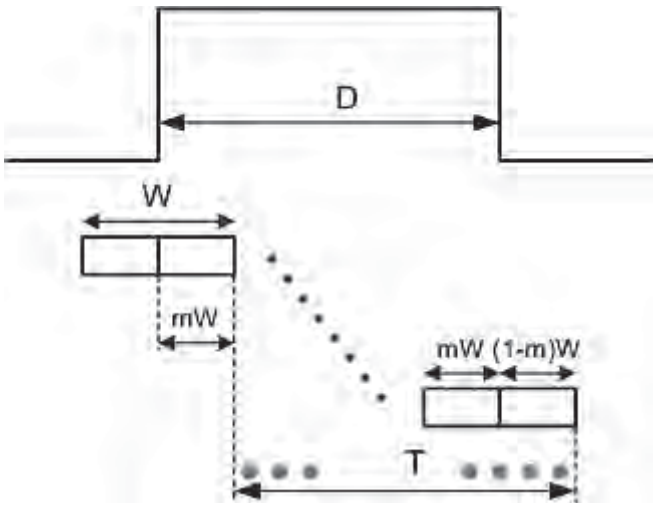

Рис. 3. Часова діаграма механізму виявлення аномалій

З рис. 3 можна побачити, що виявляються атаки тривалістю половини ширини вікна виявлення, принаймні коли m= 1/2. Індикація сигналу виявлення T розширюється від mW до D+(1-m)W. Відповідно до вибраного значення m, період T змінюється між | W-D | і W+D. Затримка виявлення становить mW. Проте, на основі результатів емпіричних досліджень встановлено, що затримка змінюється залежно від сили атаки і порогового рівня.

Особливості застосування властивостей порогу виявлення аномалій при статистичному аналізі. Розглянемо теоретичну основу для отримання порогу виявлення аномалій. При випадковій величині X(t), маючи середнє μ і дисперсію σ^2, можна записати нерівність Чебишова з погляду числа стандартних відхилень від середнього[3]:

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2 . \quad (5)$$

За допомогою нерівності Чебишова можна визначити нижню межу рівня довіри, однак в нерівності не враховано фактичний розподіл, і тому межа рівня довіри часто нестійка. Якщо припустити, що відновлений сигнал має нормальний розподіл, можна розробити відповідні методи аналізу для виявлення аномалій з високим степенем довіри при одночасному зниженні помилкових спрацювань.

На рис. 4, *в* показано гістограму відновленого сигналу в завершеному режимі за відсутності атаки. Після перетворення даних отримано середнє 0 і стандартне відхилення 3.38, як показано на рис. 4, *б*. Перетворені дані мають нормальний розподіл на рівні значущості 5 %, а саме X~N (0, 3.38^2). При

порогах -10.15 і 10.15 відповідно ці цифри еквівалентні ± 3.0σ довірчого інтервалу для випадкового процесу X. Цей інтервал відповідає 99,7 % довірчого рівня:

$$P(\mu - 3.0\sigma < X \le \mu + 3.0\sigma) \approx 99.7\% \qquad (6)$$

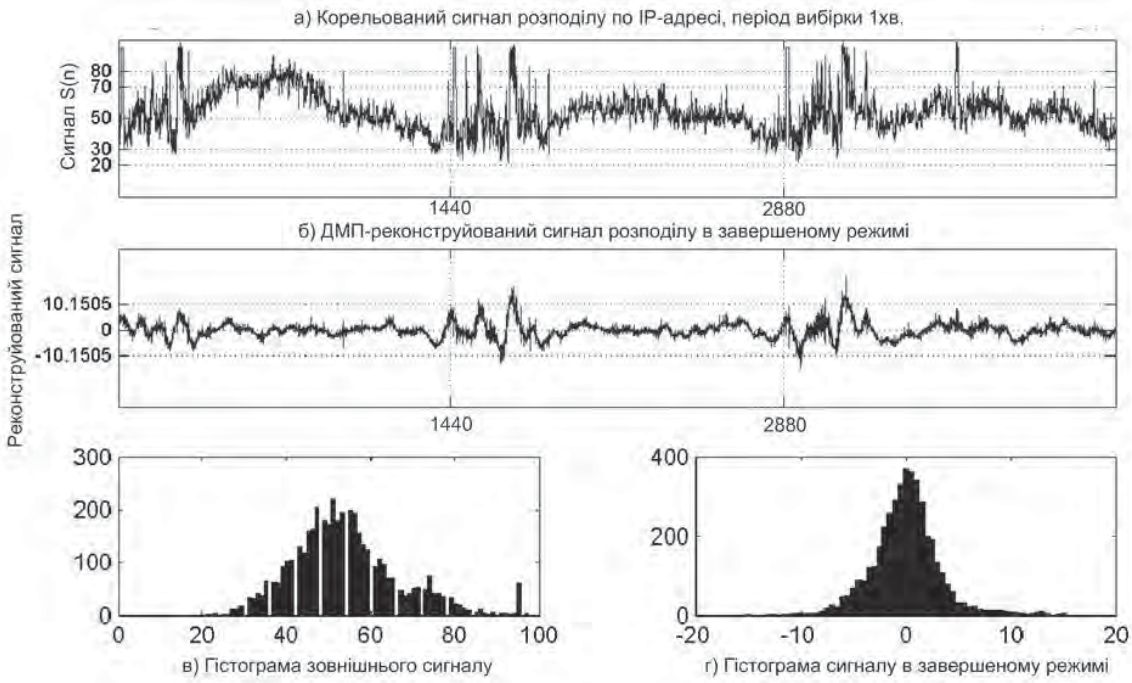

Рис. 4. Розподіл за відсутності атаки:
а, в – перед ДМП; б, г – після ДМП

За таких рівнів порогу виявлення атак проводиться з похибкою 0.3 %.

Виявлення аномалій трафіку після завершення нападу. Для проведення досліджень було взято сигнали з бази Інтернет-трафіку [6]. Результати наведено на рис. 5.

Рис. 5. Розподіл за наявності атаки

Зокрема, на рис. 5, *а* показано зважений корельований за IP-адресою сигнал за наявності атаки, на рис. 5, *б* ДМП перетворений та відновлений сигнал з результатами виявлення аномалій. Інтервал вибірки складає 1 хвилину і тривалість вибірки становить 30 секунд. Тобто, проведено вибірку протягом 30 секунд і 30 секунд – пауза. Атаки проводилися поетапно і обмежені на рис. 5 вертикальними лініями. Для аналізу та виявлення використано 3-денне широке вікно ДМП і 20-хвилинне широке вікно DET. Для аналізу було використано увесь 3-денний набір корельованих даних. Для оцінювання відновленого сигналу використано статистичний поріг ± 3.0σ. Реконструйовані сигнали перших 3 атак мають коливальну форму, тому що шаблони атаки – переривчасті, тоді як інші шість нападів мають форму пагорбів і ям, оскільки атаки проводилися безперервно в часі.

Представлення сигналу у вигляді точок показують, що ці типи атак можуть бути виявлені ефективно. З іншого боку, напіввипадкові і випадкові атаки пояснюються низькою кореляцією, що означає поведінку трафіку, яка несумісна з шаблоном. Ці атаки можуть бути виявлені лише по всій ширині часу атаки. Сигнали послідовного виявлення вказують на довжину атак, а також на інтенсивність аномалій.

Виявлення аномалій в режимі реального часу. У разі аналізу трафіку в режимі реального часу немає можливості вибірково аналізувати трафік у різних часових масштабах, оскільки аномалії повинні бути виявлені, як тільки вони виникають. Тому аналіз в режимі реального часу вимагає аналізу даних на усіх часових масштабах. Зважаючи на це, необхідно забезпечити нижчі затримки при виявленні аномалій. Для того, щоб провести аналіз за короткий час, можна акцентувати увагу тільки на невеликих останніх наборах даних. Оскільки число вибірок, що піддаються перетворенню, тісно пов'язано з розміром вікна ДМП, максимально допустимі рівні обмежені значенням $\log_2(n)$, де n – число вибірок.

Для того, щоб отримати порогові рівні для виявлення аномалій, треба встановити статистичну базову лінію для зовнішнього трафіку. Крім того, необхідно класифікувати кожен рівень ДМП розкладу зовнішнього трасування і реконструювати сигнал на кожному рівні. Статистичні параметри відновленого сигналу на кожному рівні незалежно розраховуються. Ці параметри оновлюються на відповідному періоді разом з поточним значенням і старими параметрами.

Відновлений сигнал кожного рівня використовується для виявлення аномалій. Для цього проводиться статистичний аналіз кожного ДМП рівня сигналу окремо для того, щоб проаналізувати сигнал на всіх часових масштабах. Крім того, механізм виявлення працює у двох вимірах: горизонтальному і вертикальному.

Вимірюванням по горизонталі виявляють аномалії в послідовних часових вибірках на тому самому рівні малохвильового сигналу. Вимірюванням по вертикалі виявляють аномалії на різних рівнях малохвильового сигналу в одному часовому проміжку.

Для виявлення аномалії використовується комбінація горизонтального і вертикального оцінювання. Число ймовірних датчиків нападу пораховано, використовуючи 2-мірні вікна виявлення, що складаються з горизонтальних і вертикальних компонентів.

Аномалія вважається виявленою, коли число детекторів перевищує поріг у 2-мірному вікні. На рис. 6 представлено приклад проведеного аналізу в режимі реального часу. Було використано 2-годинне вікно ДМП у 2-хвилинному інтервалі вибірки та проведено аналіз до 6 рівня. При дослідженні використовувалося дискретне малохвильове перетворення з базисними функціями Добеші D6.

Проаналізовано можливості застосування дискретного малохвильового перетворення для виявлення аномалій трафіку. Зокрема, розглянуто метод, який використовує кореляцію IP-адрес призначення у трафіку на виході маршрутизатора. Результати, отримані з використанням цього методу, показують, що статистичний аналіз даних заголовка пакетів може забезпечити ефективний механізм для виявлення аномалій трафіку у мережі кампуса або на краю мережі.

Також показано ефективність цього підходу в завершеному режимі і в режимі реального часу при аналізі мережевого трафіку.

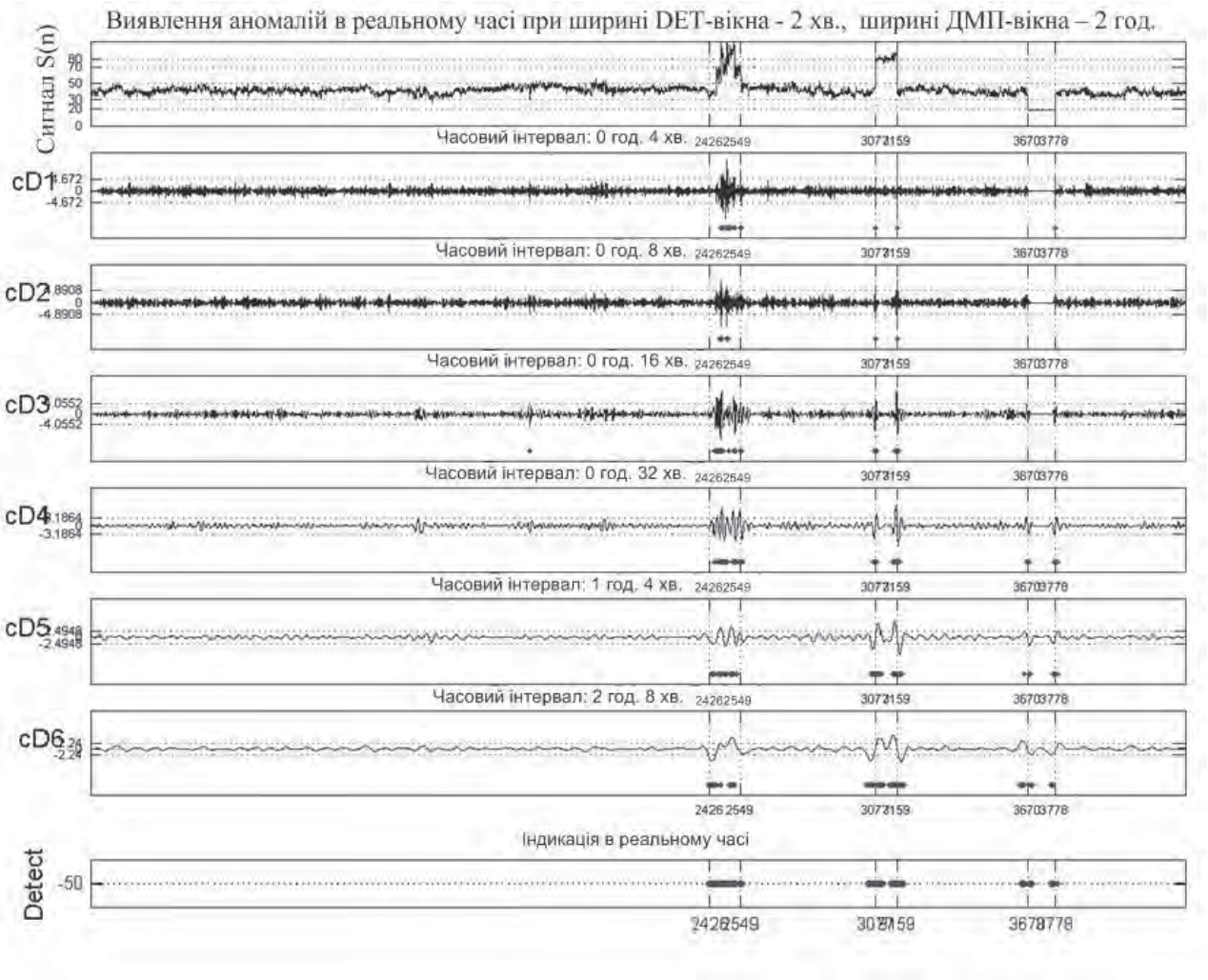

Рис. 6. Результати виявлення атак у реальному часі з використанням ДМП

Висновки. Для оцінки ефективності застосування ДМП порівняно результати виявлення аномалій трафіку схеми з використанням ДМП із схемою, яка працює безпосередньо з статистичним аналізом.

Встановлено, що при низькому довірчому рівні (нижче 90%), ДМП не дає ніяких переваг. Проте, коли довірчий рівень більший (90% ~ 99,7%) ДМП дає значно кращі результати виявлення, ніж простий статистичний аналіз. Тобто використання ДМП надає значні покращення у виявленні аномалій трафіку мереж.

1. Barford P., Kline J., Plonka D. and A. Ron. A Signal Analysis of Network Traffic Anomalies, in Proc. of ACM SIGCOMM Internet Measurement Workshop (IMW), 2002. 2. Kim S., Reddy A. L. N. and M. Vannucci. Detecting traffic anomalies using discrete wavelet transform, in Proc. of International Conference on Information Networking (ICOIN), 2004. 3. Kilpi J. and I. Norros. Testing the Gaussian approximation of aggregate traffic, in Proc. of ACM SIGCOMM Internet Measurement Workshop (IMW) 2002. 4. Малла С. Вейвлеты в обработке сигналов. – М.: Мир, 2005. – 671 с. 5. Добеши И. Десять лекций по вейвлетам. – Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. – 464 с. 6. http://ita.ee.lbl.gov/index.html, Internet Traffic Archive.