TECHNICAL SCIENCES AND TECHNOLOGIES
УДК 004.043
DOI: 10.25140/2411-5363-2017-2(8)-99-109

Віталій Литвинов, Ігор Скітер, Олена Трунова, Едуард Сідін

МОДИФІКАЦІЯ МЕТОДИКИ ВЕЙВЛЕТ-АНАЛІЗУ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ У ТРАФІКУ КОМП'ЮТЕРНОЇ МЕРЕЖІ

Актуальність проблеми. *Надійна передача даних у мережі повинна базуватись на використанні адекватних методів виявлення аномалій мережевого трафіку. Кількісний аналіз мережі на основі використання статистичного підходу базується на аналізі масивів даних у вигляді динамічних рядів. Для ефективного використання методів вейвлет-аналізу необхідне удосконалення методичного забезпечення аналізу трафіку комп'ютерної мережі.*

Постановка проблеми. *Методи вейвлет-аналізу є перспективними для виявлення аномальної поведінки мережевого трафіку, так як вони базуються на декомпозиції трафіку як динамічного ряду. При цьому існують проблеми вибору відповідних масштабуючих вейвлет-функцій, способів визначення коефіцієнтів деталізації, їх трактовки та перевірки гіпотез про аномальність поведінки трафіку.*

Аналіз останніх досліджены публікацій. *Роботи, присвячені статистичним методам та технологіям аналізу та виявлення аномалій включають в себе алгоритми оцінки аномальності трафіку за такими показниками як: помилки першого роду, помилки другого роду, кількість правильно виявлених аномалій.*

Виділення не вирішених раніше частин загальної проблеми. *Використання алгоритмів вейвлет-аналізу повязане з їх складністю та ресурсоємністю, труднощами виявлення помилок другого роду. Крім того, є питання вибору максимально адекватної масштабуючої функції та необхідність трактовки апроксимуючих та деталізуючих коефіцієнтів, які утворюють окрему статистичну вибірку.*

Мета дослідження. *In this paper, classification of mobile applications was presented alongside with technologies, which can be used for development of mobile applications. Метою є використання ряду Фурьє в якості масштабуючої функції при вейвлет-аналізі трафіку; побудова ідеалізованого профілю мережі та оцінка аномальності поведінки трафіку; аналіз коефіцієнтів деталізації як окремої статистичної вибірки.*

Виклад основного матеріалу. *Проведений аналіз трафіку з використанням, як масштабуючої, функції Фурьє дав змогу отримати в явному вигляді амплітуди та початкові фази гармонічних компонент. Це дало можливість проводити порівняння «ідеального профілю» трафіку з реальним. Попередні висновки про наявність аномалії проводяться за зонами, в яких спостерігається перевищення змодельованого трафіку над реальним. Різка зміна абсолютного значення деталізуючи коефіцієнтів у аналізованих вікнах трафіку також може трактуватися як аномалія трафіку.*

Висновки. *Пропонована модифікація методики вейвлет-аналізу дає значне скорочення ресурсоємності аналізу трафіку мережі. Використання ряду Фурьє дає змогу виявляти тренди та циклічні складові в трафіку, виявляти зони аномальності. Отримані апроксимуючі та деталізуючі коефіцієнти, можуть бути використані в якості характеристик аномальності трафіку при аналізі їх зміни.*

Ключові слова: *вейвлет-аналіз; апроксимуюча функція; ряд Фурьє; деталізуючі коефіцієнти; профіль мережі; аномалії.*

Постановка проблеми. Аномалії трафіку в комп'ютерній мережі мають різні причини і можуть бути пов'язані з діяльністю хакерів, некомпетентних користувачів, несправністю апаратури і дефектами програмного забезпечення. Існують видимі аномалії, які проявляються безпосередньо в некоректній роботі інформаційно-обчислювальної системи. Аномалії можуть і не мати видимих ознак, але привести до збоїв через тривалий час.

Аналіз аномалій дозволяє виявляти суттєві відхилення трафіку мережевих пристроїв від «нормального» профілю трафіку для цього пристрою або групи пристроїв. Як правило, шаблон «нормального» трафіку мережі складається протягом певного проміжку часу на основі статистичних даних та навчальної вибірки.

Аналіз показує, що для виявлення аномалій здебільшого достатньо аналізувати основні параметри трафіку і немає необхідності вивчати вміст кожного пакета. Прикладами аномалій, виявлених на основі аналізу трафіку, є раптове збільшення інтенсивності трафіку від робочої станції або зміна структури трафіку в порівнянні зі звичайними щоденними показниками для цієї мережі або пристрою.

Застосовувані у разі виявлення та запобігання мережевих аномалій методи зводяться до аналізу сигнатур [1] та аналізу трафіку статистичними методами [2].

Аналіз сигнатур базується на понятті збігу послідовності зі зразком. Системи аналізу сигнатур [1] мають високу швидкість, на відміну від систем аналізу протоколів. Суттєвою проблемою використання методу аналізу сигнатур полягає в тому, що швидкість аналізу з часом зменшується зі збільшенням кількості перевірених сигнатур, чисельність яких може значно зростати у процесі аналізу трафіку. Крім того, список перевіре-

них сигнатур під час здійснення вторгнень чи атак нового типу різко збільшується. Тому застосування для аналізу трафіку в режимі реального часу методів аналізу сигнатур має низьку ефективність щодо своєчасного виявлення атак, неможливість самостійного виявлення нових атак і постійна необхідність оновлення бази сигнатур.

Статистичний аналіз базується на основі динамічних рядів даних про параметри трафіку мережі. Основним поняттям аналізу трафіку є визначення профілю для всіх суб'єктів аналізованої системи, як аналог еталонної поведінки мережі за відсутності аномалій, вторгнень, атак. Будь-яке відхилення використовуваного профілю від еталонного вважається несанкціонованою діяльністю. Статистичні методи універсальні, оскільки для проведення аналізу не потрібні попередні знання про можливі атаки і використовувані ними методи вразливості. Крім того, значною перевагою статистичниго аналізу трафіку є можливість визначення нових та таких, які раніше не ідентифікувалися, методів впливу на комп'ютерну мережу.

Основні переваги статистичного підходу – використання вже розробленого і зарекомендованого апарату математичної статистики і його адаптація до поведінки суб'єкта за рахунок використання відповідних методів аналізу трафіку, їх комбінацій та модифікацій. Недоліками статистичних методів є достатньо велика ймовірність помилок першого та другого роду у випадку виявлення аномалій трафіку.

Методи вейвлет-аналізу є перспективними для виявлення аномальної поведінки мережевого трафіку, оскільки в їх основу покладений принцип декомпозиції динамічного ряду, який підлягає аналізу [3]. Поряд з достатньо розробленими алгоритмами вейвлет-аналізу існують проблеми використання відповідних утворюючих функцій для вейвлет-перетворень та способів визначення коефіцієнтів деталізації, апроксимації. Також проблемою є встановлення та перевірка гіпотез про аномальність поведінки трафіку [4] та використання відповідних типів вейвлетів залежно від обраного рівня достовірності результатів.

Аналіз останніх досліджень і публікацій. Вейвлет-аналіз (ВА) мережевого трафіку передбачає подання одновимірного цифрового масиву в різних масштабах чи рівнях дискретизації. При цьому можливе виявлення аномальних, підозрілих зон чи елементів масиву залежно від різних ступенів деталізації загального масиву – характерні деталі, які можуть залишатися непоміченими при одному ступені деталізації, можуть бути очевидними на іншому. Такі можливості зумовлені статистичним характером поведінки трафіку, який розглядається як одновимірний дискретизований цифровий масив даних у вигляді числового динамічного ряду $Y(t_i) = f(\Delta t_i)$, заданого в дискретні моменти часу Δt_i на всьому масиві спостережень ($i = \overline{1,n}$) та за весь період спостережень ($t = \overline{1,T}$). Математична модель мережевого трафіку може бути представлена у вигляді суперпозиції трьох класів складових (1):

$$Y(t_i) = tr(t_i) + C(t_i) + A(t_i) + e(t_i), \quad (1)$$

де $tr(t_i)$ – тренд, загальна тенденція;

$C(t_i)$ – періодична складова (циклічна, сезонна), визначена для обраного інтервалу Δt_i;

$A(t_i)$ – аномалії, різкі зміни трафіку.

Четверта компонента у (1) являє собою стохастичну складову, шум, який може виникати навколо основних складових, і для якої робиться припущення про рівність нулю її математичного очікування у випадкові моменти часу.

Аналіз літературних джерел [5; 6] показує, що основними алгоритмами ВА, які широко використовуються у практиці виявлення аномальної поведінки трафіку комп'ютерної мережі, є:

- алгоритм на основі дискретного вейвлет-аналізу мережевого трафіку;

- алгоритм Бродського-Дарховського;
- алгоритм на основі суми квадратів вейвлет-коефіцієнтів;
- алгоритм на основі максимуму квадратів вейвлет-коефіцієнтів.

Наведені вище алгоритми аналізують такі параметри, як: помилки першого роду, помилки другого роду, кількість правильно виявлених аномалій.

Алгоритм на основі дискретного вейвлет-аналізу мережевого трафіку [7] передбачає аналіз обсягу переданої інформації в байтах, кількість переданих пакетів, завантаження процесора тощо за певний інтервал часу Δt_i. Зміна величини інтервалу дискретизації трафіку Δt_i дає змогу виявляти аномальні зміни трафіку, які не були виявлені в попередніх аналізах. Вейвлет-модель має такий вигляд:

$$Y(t_i) = \sum_{k,\tau} b_{k,\tau}\varphi_{k,\tau}(t_i) + \sum_{k,\tau} d_{k,\tau}\psi_{k,\tau}(t_i),\ k,\tau = \overline{1,\infty}, \quad (2)$$

де $\varphi_{k,\tau}(t_i)$ – масштабуюча функція, функція апроксимації мережевого трафіку;

$\psi_{k,\tau}(t_i)$ – вейвлет-функція, деталізації мережевого трафіку і його локальних особливостей;

$b_{k,\tau}$, $d_{k,\tau}$ – апроксимуючі і деталізуючи коефіцієнти з параметрами масштабу k та зсуву τ.

Перша сума в (2) характеризує тренд і циклічні складові трафіку, а друга – значення флуктуацій на цих інтервалах, що характеризують активність (аномальність) суб'єктів мережі, з урахуванням стохастичної компоненти.

Дослідження авторів [5; 8] показали, що для моніторингу мережевого трафіку доцільно використовувати масштабуючу функцію $\varphi_{k,\tau}(t_i)$ і вейвлет Хаара $\psi_{k,\tau}(t_i)$. Виявлення поточного рівня аномальності $Y^{anomal}(t_i)$ за допомогою моделі (2) проводиться через визначення різниці рядів даних, отриманих у режимі навчання (еталонного ряду $Y^{etal}(t_i)$) та в режимі реального часу (при поточному завантаженні мережі $Y^{real}(t_i)$). За твердженням авторів такий підхід дозволяє усунути тренд і циклічну складову і на певному рівні дозволу оцінити активність суб'єктів. Таким чином, алгоритм на основі дискретного вейвлет-аналізу мережевого трафіку забезпечує покращення достовірності виявлення аномалій мережевого трафіку в умовах невизначеності.

У роботі [9] запропоновано метод виявлення розбалансування трафіку, заснований на виявленні зміни середнього значення випадкової величини у моделі виду (1). Цей метод є непараметричним. Перевагами подібних методів є їх незалежність від розподілів, відсутність необхідності наявності апріорних даних і можливість організовувати кореляційний ряд з вибіркових значень.

У загальному вигляді алгоритм Бродського-Дарховського виглядає таким чином:

$$\begin{cases} Y(n) = \left[\dfrac{n}{N}\left(1-\dfrac{n}{N}\right)\right]^{\nu}\left(\dfrac{1}{n}\sum_{i=1}^{n} Y_i - \dfrac{1}{N-n}\sum_{i=N+1}^{n} Y_i\right), \\ n_0 = \arg\max\left|Y(n)\right| \end{cases} \quad (3)$$

де n – крок, на якому проводиться оцінка;

n_0 – оцінка моменту виникнення аномалії.

$0 \le \nu \le 1$

У разі, коли виявлення розбалансування трафіку проводиться в реальному часі методом ковзаючого вікна розміру m, що зміщується зліва направо по мірі надходження даних, алгоритм доповнюється даними, зміщеними на величину вікна [10]. Висновок

про наявність аномального викиду приймається не в усій реалізації, а в кожному конкретному вікні. така схема дозволяє динамічно відстежувати зміни в мережі і швидше реагувати на них. Алгоритм чітко визначає момент початку і кінця розбалансування трафіку за відсутності шуму і рівні шуму, що дорівнює рівню сигналу. У випадку перевищення значення шуму над значення сигналу в 2 і більше разів алгоритм працює нестабільно і з'являється ймовірність появи помилок першого і другого роду. Алгоритм у режимі ковзаючого вікна дозволяє виявити викид при значному рівні шумів, у тому числі і перевищенні шумами рівня сигналу. За допомогою алгоритму Бродського-Дарховського виявляється менше помилок 1 та 2 роду, але при цьому алгоритм має великі вимоги до ресурсів.

У роботі [11] наголошується, що одним з перспективних методів виявлення аномалій мережевого трафіку є методи ВА, що здійснюються на основі вейвлет-декомпозиції аналізованого, в загальному випадку нестаціонарного, сигналу. Вейвлет-декомпозиція дозволяє представити аналізований мережевий трафік у вигляді набору вейвлет-коефіцієнтів, які являють собою нову статистичну вибірку, що має свої власні характеристики.

Алгоритм, заснований на сумі квадратів вейвлет-коефіцієнтів, має велику ефективність. Найбільший ефект виявляється у процесі використання коефіцієнтів апроксимації для вейвлетів Хаара на верхніх рівнях розкладання. Але збільшення розміру вікна аналізу може привести до зростання ймовірності правильного виявлення аномалії, але при цьому також зростає ймовірність помилкового виявлення.

Виділення не вирішених раніше частин загальної проблеми. Методи виявлення аномалій, які базуються на методиці ВА, реалізованого на пропонованих алгоритмах, мають максимальну ефективність порівняно з іншими алгоритмами виявлення аномалій. Поряд з цим існують зауваження до використання цих методів, пов'язаних із реалізацією пропонованих алгоритмів. У таблиці наведена оцінка алгоритмів на наявність виконання кожного з факторів, що впливають на точність виконання алгоритмів.

Таблиця

Аналіз ефективності алгоритмів вейвлет-аналізу

Назва алгоритму	Помилки першого роду	Помилки другого роду	Складність алгоритму
Алгоритм на основі дискретного вейвлет-перетворення	+	+	-
Алгоритм Бродського-Дарховського	+	+	-
Алгоритм на основі суми квадратів вейвлет-коефіцієнтів	+	-	+
Алгоритм на основі максимуму квадратів вейвлет-коефіцієнтів	+	-	+

Примітка: «+» означає здійсненність цього критерію оцінки щодо алгоритму; «-» означає нездійсненність цього критерію оцінки щодо алгоритму.

Як видно з таблиці, проблема використання алгоритмів пов'язана з частковою нездійсненністю виявлення помилок другого роду в алгоритмах, пов'язаних із аналізом вейвлет-коефіцієнтів та значною складністю та ресурсоємністю алгоритмів дискретних перетворень та алгоритму Бродського-Дарховського.

Крім того, виходячи з аналізу останніх публікацій можна зробити висновок про те, що пропоновані методи виявлення аномалій будуть максимально ефективними у процесі аналізу трафіку в режимі реального часу за умови, коли буде запропонована технологія дискретизації трафіку та одночасного використання «ковзких вікон» трафіку. При цьому постає питання вибору максимально адекватної апроксимуючої функції, яка давала б змогу проводити аналіз у межах мінімальної складності та максимальної ефективності, при низькій ресурсоємності програмної реалізації. Також існує невизначеність у трактуванні апроксимуючих та деталізуючих коефіцієнтів, які утворюють окрему статистичну вибірку з параметрами масштабу k та зсуву τ щодо виявлення на їх основі аномалій мережевого трафіку.

Мета статті. Головною метою цієї роботи є модифікація методів вейвлет-аналізу в частині використання як масштабуючої функції ряду Фур'є. За обґрунтованого та коректного використання ряду Фур'є як апроксимуючої функції такий підхід дасть змогу проводити з мінімальними затратами ресурсів процедуру дискретизації та апроксимації трафіку за вибрані проміжки часу Δt_i. Крім того, апроксимація реального трафіку на обраних часових діапазонах дасть змогу з максимальною достовірністю та мінімальною затримкою в часі оцінювати ідеалізований профіль мережі та відхилення від «ідеальності», що можна класифікувати як аномальну поведінку трафіку. Аналіз коефіцієнтів деталізації, якими будуть виступати оцінки параметрів ряду, їх змін у часі буде додатковою характеристикою оцінки аномальної поведінки комп'ютерної мережі.

Виклад основного матеріалу. Згідно з (2) вейвлет-модель включає в себе функцію апроксимації та функцію деталізації мережевого трафіку. В роботі пропонується використання як апроксимуючої функції ряду Фур'є:

$$\phi_{k,\tau}(t_i) = \frac{b_0}{2} + \sum_{i=1,k=1}^{n,\infty} \left[b_{k,\tau} \cos(\frac{2\pi}{N} k t_i) + d_{k,\tau} \sin(\frac{2\pi}{N} k t_i) \right], \quad (4)$$

де $b_0 = \frac{\sum_{i=1}^{N} T_k(t_i)}{N}$;

$$b_{k,\tau} = \frac{2}{N} \sum_{i=1}^{N} \left[T_k(t_i) \cos(k t_i) \right];$$

$$d_{k,\tau} = \frac{2}{N} \sum_{i=1}^{N} \left[T_k(t_i) \sin(k t_i) \right];$$

k – порядок вейвлету, гармоніка ряду Фур'є;

N – довжина періоду дискретизації трафіку;

$T_k(t_i)$ – кількість запитів у трафіку за відповідний деталізуючий період.

Використання ряду Фур'є як функції апроксимації дає можливість отримати найбільш точні результати та виявити закономірності в динамічному ряду, яким є трафік мережі за період спостереження чи оцінювання. Під час аналізу трафіку важливим є те, що у функції Фур'є чітко виражаються амплітуди та початкові фази гармонічних компонент. Унаслідок цього є можливість встановлювати схожість чи відмінність у різних дискретизованих частинах трафіку за різних умов, як правило, у процесі порівняння «ідеального профілю» з реальним трафіком.

Перевага використання ряду Фур'є як масштабуючої функції полягає в тому, що щодо вхідного динамічного ряду – трафіку мережі – інформація про поведінку трафіку не втрачається за будь-якого рівня деталізації чи дискретизації. Недоліком використання ряду Фур'є у вейвлет-аналізі є залежність виявлення циклічних компонент від масштабу дискретизації, ширини вікна вейвлету.

Для отримання необхідних даних для аналізу трафіка на виявлення аномалій було взято лог – файл сайта форуму (zadrots.ru). Лог – файл містить більше 1 000 000 записів і включає в себе інформацію про IP-адресу користувача, запит користувача, дату та час, коли був здійснений запит, браузер та операційну систему користувача, відповідь, реферер. Основними даними для аналізу трафіку при реалізації методу є кількість запитів $T(t_i)$ та час t_i. Для реалізації пропонованого методу аналізу обраний добовий інтервал t=24 год з погодинною дискретизацією трафіку. Ширина вейвлету становить N=6 год. Період добового спостереження розбитий погодинно на 4 періоди: $N_1 \in [06.00 - 12.00]$,

$N_2 \in]12.00-18.00]$, $N_3 \in]18.00-24.00]$, $N_4 \in]24.00-06.00]$. Дискретизований трафік за 4 доби (96 годин) погодинно представлений на рис. 1.

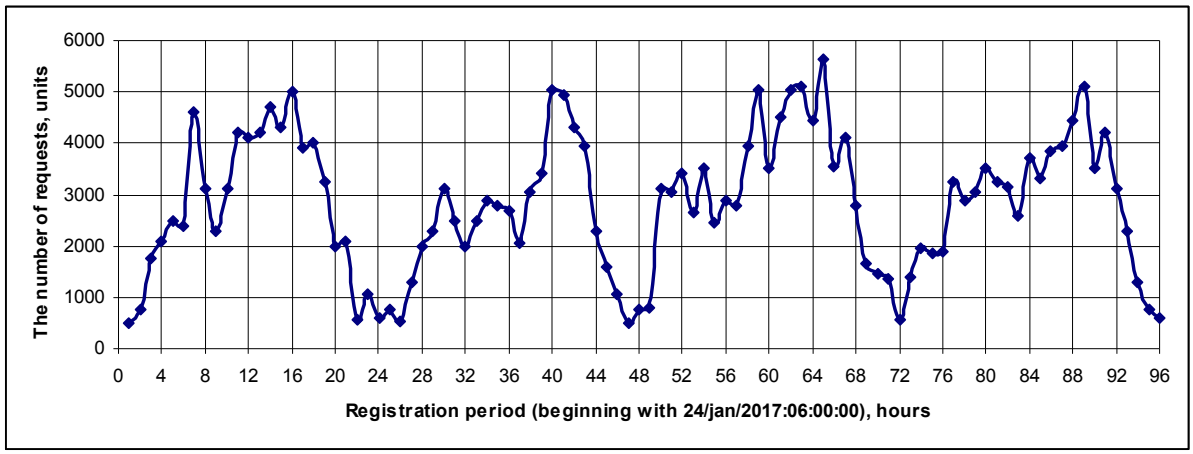
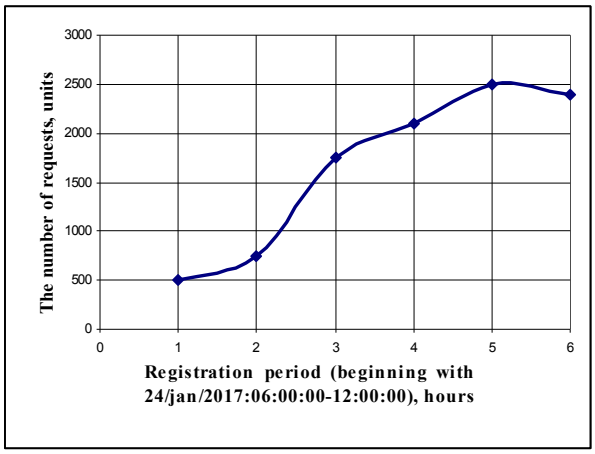
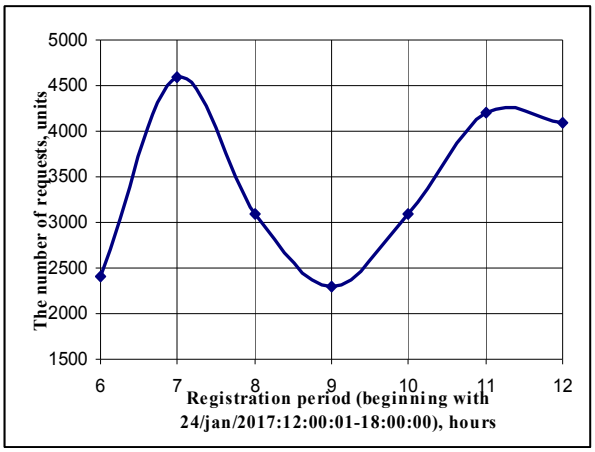


Рис. 1. Дискретизований трафік досліджуваної мережі за період 24.01.2017 р. – 27.01.2017 р.

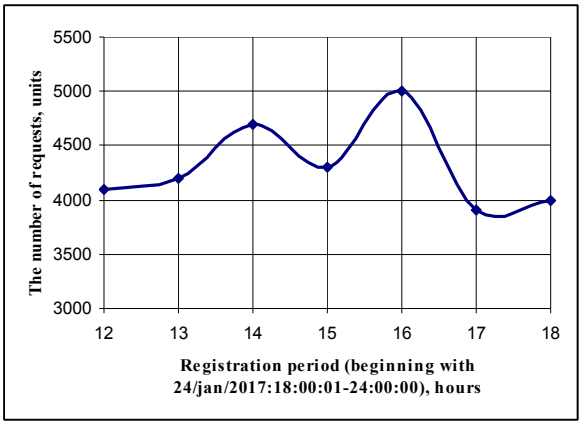
Реалізацію пропоновано підходу розглянемо на прикладі добового дискретизованого трафіку, розбитого на 4 періоди (рис. 2).



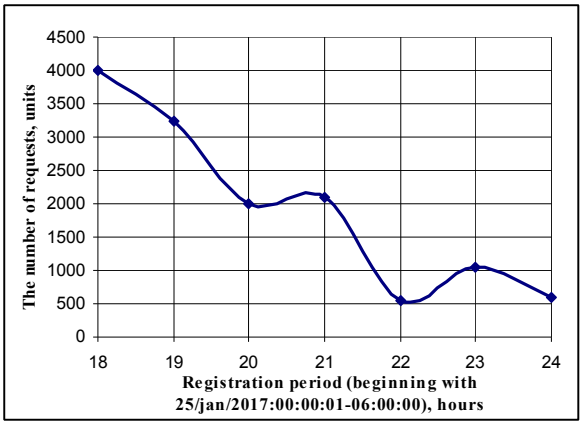
а



б



в



г

Рис. 2. Дискретизований добовий трафік досліджуваної мережі за період 24.01.2017 р. (06:00:00) - 25.01.2017 р.(06:00:00): а – $N_1 \in [06.00-12.00]$; б – $N_2 \in]12.00-18.00]$; в – $N_3 \in]18.00-24.00]$; г – $N_3 \in]18.00-24.00]$

Апроксимацію дискретизованого трафіку для обраного періоду проводимо для k-ї гармоніки ряду Фурьє ($k=\overline{1,\infty}$) за допомогою визначення апроксимуючих і деталізуючих коефіцієнтів $b_{k,\tau}$, $d_{k,\tau}$ з параметрами масштабу k та зсуву τ, синтезом масштабуючої функції виду (4) та розрахунком трендових рівнів вейвлет-моделі. Критерієм якості моделювання та вибором найкращої гармоніки апроксимуючої функції є мінімізація похибки апроксимації:

$$\sigma_k = \sqrt{\frac{\sum_{i=1}^{N}\left[T_k(t_i)-T(t_i)\right]^2}{N}} = \min, \qquad (5)$$

де $T_k(t_i)$ – трендові рівнів вейвлет-моделі для відповідного значення k.

Процес вибору вейвлету, який максимально достовірно відображає реальний трафік, є ітераційним і проводиться для k=1, k=2, k=3,….. k=m з визначенням на кожному кроці похибки апроксимації. Критерієм зупинки є мінімальне значення похибки апроксимації σ_k.

Алгоритм розрахунку утворюючої функції ряду Фур'є зображено на рис. 3.

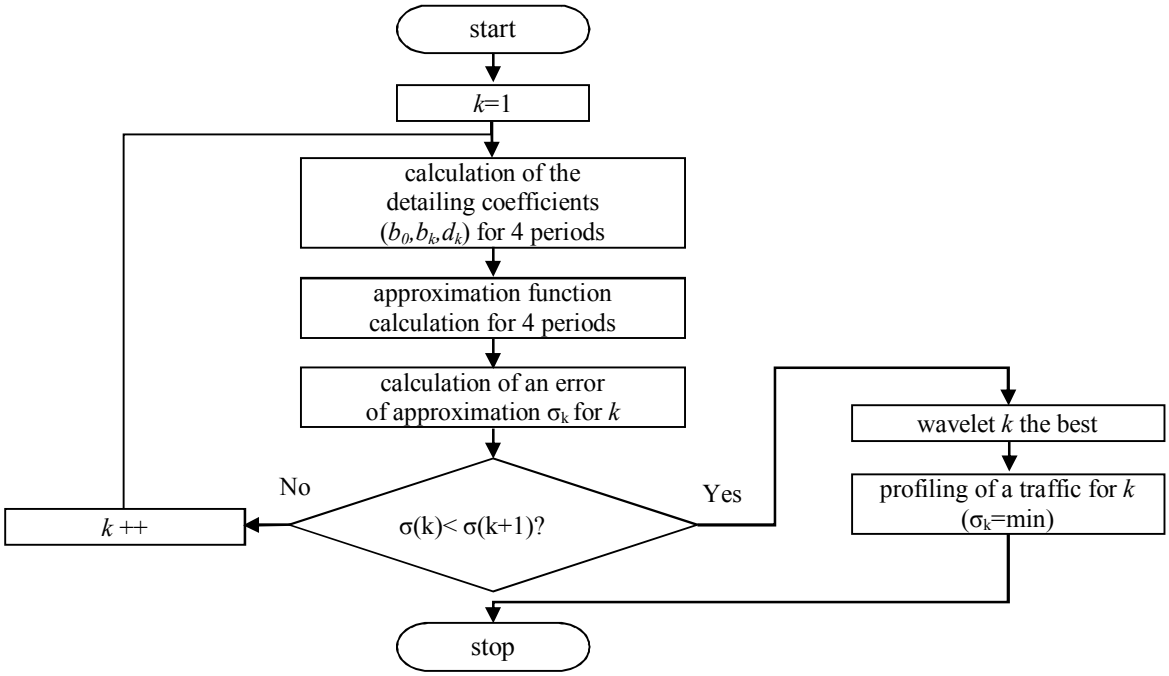


Рис. 3. Алгоритм розрахунку утворюючої функції ряду Фур'є

Для адаптації вейвлет-перетворення до аналізу трафіка використовується техніка двох вікон: *W1* (вікно порівняння) і *W2* (вікно виявлення). Вікно порівняння являє собою дискретизований реальний трафік, отриманий з лог-файла для відповідних періодів дискретизації, вікно виявлення – трендові значення трафіку для найкращого вейвлету (σ_k=min). Результати моделювання трафіку за допомогою модифікованого вейвлет-аналізу наведені на рис. 4. Попередні висновки про наявність аномалії проводяться за зонами, в яких спостерігається перевищення змодельованого трафіку над реальним [12].

Для більш математично коректного визначення аномальності трафіку у визначені часові проміжки чи моменти пропонується проводити аналіз динаміки апроксимуючих та деталізуючих коефіцієнтів.

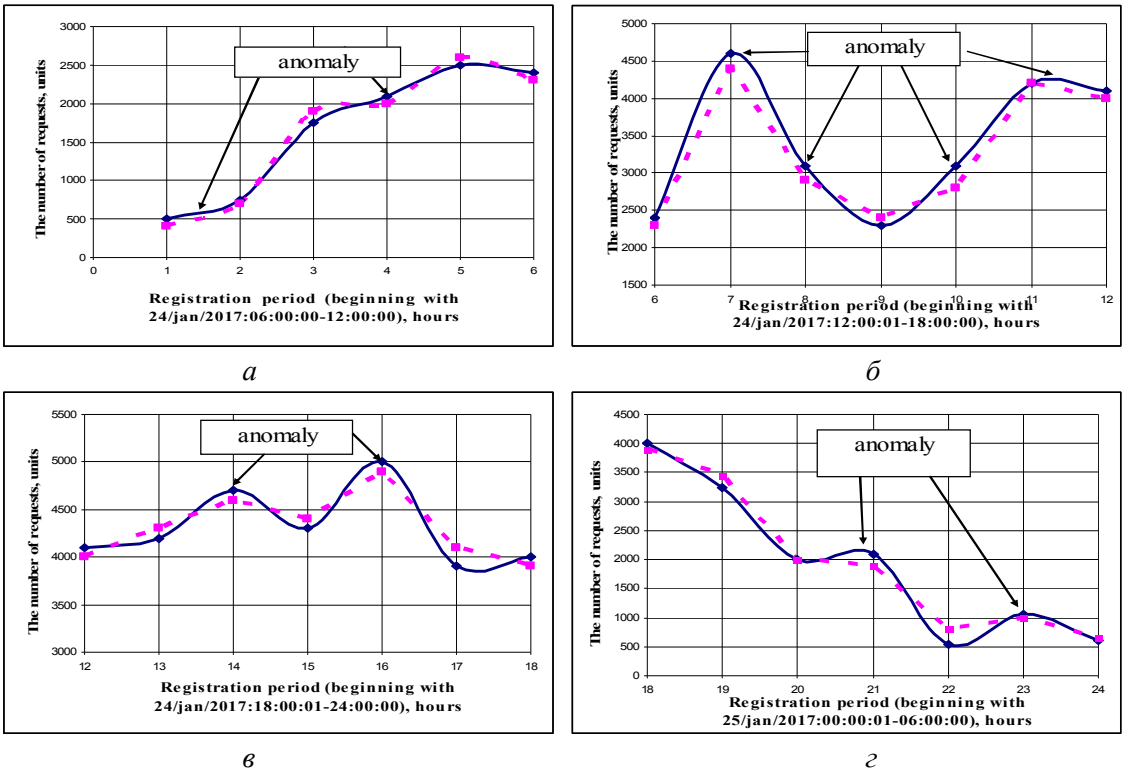


Рис. 4. Виявлення аномалій за аналізом апроксимуючої функції Фур'є: а – $N_1 \in [06.00-12.00]$*;*
б – $N_2 \in]12.00-18.00]$*; в –* $N_3 \in]18.00-24.00]$*; г –* $N_3 \in]18.00-24.00]$

Згідно з теоремою Гаусса-Маркова [13] невідповідність третій умові теореми призводить до різкої зміни коефіцієнтів апроксимуючої функції при незначній зміні умов моделювання чи, наприклад, масиву спостережень. Невідповідність третій умові може свідчити про вплив сторонніх факторів на результати моделювання. Для перевірки умови стаціонарності, однорідності трафіку та відсутності впливу стохастичної складової проведений Фур'є-аналіз вейвлетів масштабу k (за умови σ_k=min) на основі «ковзких вікон» при зсуві (лагу) τ =1 година. При визначеній ширині вікна апроксимації $N = 6$ годин та погодинній дискретизації трафіку проведене моделювання трафіку протягом добового спостереження кількості запитів. У результаті отримані деталізуючі коефіцієнти $b_{k,\tau=1}$, $d_{k,\tau=1}$, які являють собою статистичну вибірку, що має свої власні характеристики (рис. 5).

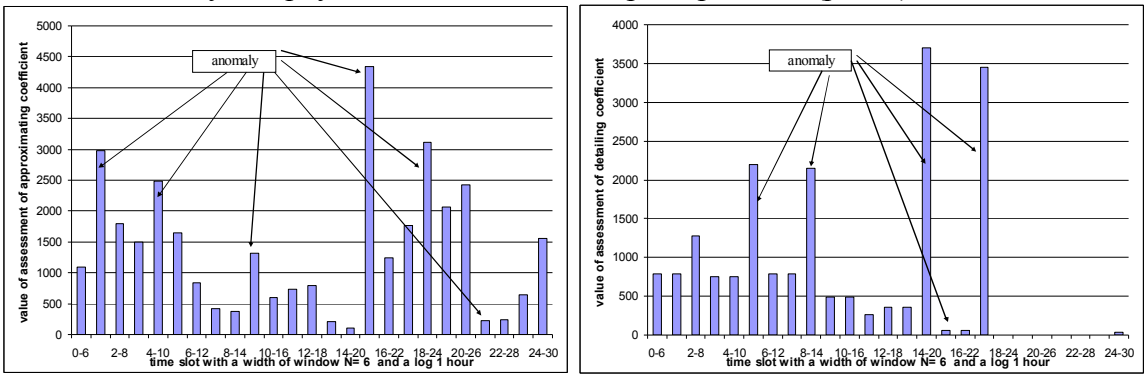


Рис. 5. Аналіз аномальної поведінки трафіку на часовому проміжку спостережень за період 24.01.2017 р.(06:00:00) – 25.01.2017 р.(06:00:00): а – апроксимуючі коефіцієнти; б – деталізуючі коефіцієнти

Як видно з рис. 5, різка зміна абсолютного значення коефіцієнтів у вказані періоди може свідчити про значний вплив сторонніх факторів, стохастичної складової, шумів, що може трактуватися як аномалія трафіку.

Висновки і пропозиції. Таким чином, за результатами пропонованої модифікації методики вейвлет-аналізу досягнуто значне скорочення ресурсоємності проведення аналізу трафіку мережі на предмет виявлення його аномальної поведінки.

Використання як апроксимуючої функції ряду Фур'є дає змогу проводити та виявляти не тільки трендів трафіку як динамічному ряду, але й виявляти його циклічні складові.

Побудова ідеалізованого профілю трафіку за критерієм мінімальної похибки апроксимації дає змогу проводити порівняння його з реальним трафіком та виявляти зони аномальності.

Отримані в результаті використання пропонованого підходу апроксимуючі та деталізуючі коефіцієнти, які утворюють власну статистичну вибірку, можуть бути використані як характеристики аномальності трафіку під час аналізу їх зміни (невиконанні третьої умови Гаусса-Маркова).

Подальшим напрямком досліджень, на нашу думку, доцільно провести аналіз автокореляційної функції, побудованої на динамічному ряду (трафіку) з метою виявлення порушень п'ятої умови Гаусса-Маркова стосовно взаємного впливу стохастичної компоненти в різних вікнах трафіку з різними значеннями лагу.

Список використаних джерел

1. *Denning Dorothy.* An Intrusion-Detection Model / Denning Dorothy // IEEE Transactions on Software Engineering. – 1987. – Vol. SE-13, No. 2. – Pp. 222–232.

2. *Amoroso Edward G.,* Intrusion Detection, 1st ed., Intrusion.Net Books, Sparta, New Jersey, USA, 1999, p. 218.

3. *Paul Barford, Jeffery Kline, David Plonka and Amos Ron.* A Signal Analysis of Network Traffic Anmalies / in Proceedings of the 2nd ACM SOGCOMM Workshop on Internet Measurement. – New York, NY, USA, ISBN:1-58113-603-X. Pp.71–82.

4. *Шелухин О. И.* Сравнительный анализ алгоритмов обнаружения аномалий трафика методами дискретного вейвлет-анализа / О. И. Шелухин, А. С. Филинова // Т-Comm-Телекоммуникации и Транспорт. – 2014. – Вып. 9, т. 8. – С. 89–97.

5. *Орлов С. И.* Вейвлет-анализ. Основы теории / С. И. Орлов. – М. : ТЕХСФЕРА, 2006. – 272 с.

6. *Sheluhyn O. I.* Measuring of Reability of Network Anomalies Detections Using Methods of Discrete Wavelet Analysis / O. I. Sheluhyn, A. V. Pankrushin // Science and Information (SAI), Conference, 2015, Longon, UK, pp. 393–397.

7. *Шелухин О. И.* Обнаружение аномальных выбросов трафика методами дискретного вейвлет – анализа / О. И. Шелухин, А. В. Гермашев // Электромагнитные волны и электронные системы. – 2012. – № 22. – С. 71–82.

8. *Соловьев С. А.* Метод идентификации угроз безопасности информационных ресурсов АСУ на основе мультиразрешающего анализа / С. А. Соловьев, Л. А. Юркевская // Вестник Самарского государственного технического университета. – 2007. – № 2 (20). – С. 70–76.

9. *Дарховский Б. С.* Апостериорное обнаружение момента «разладки» случайной последовательности / Б. С. Дарховский, Б. Е. Бродский // Теория вероятностей и ее применение. – 1980. – Т. 25, № 3 – С. 635–639.

10. *Скітер І. С.* Ідентифікація аномальної поведінки трафіку комп'ютерної мережі на основі EWMA-статистики / І. С. Скітер, І. В. Бальченко // Перша Міжнародна конференція «Проблеми виведення з експлуатації об'єктів ядерної енергетики та відновлення оточуючого середовища» INUDECO'16 25-27 квітня 2016 : зб. матеріалів. – Славутич : СФ НТУУ «КПІ», 2016. – С. 171–178.

11. *Микова С. Ю.* Обзор алгоритмов выявления сетевых атак / С. Ю. Микова, В. С. Оладько // Актуальные проблемы гуманитарных и естественных наук. – 2015. – № 9-1. – С. 59–62.

12. *Моделювання* та аналіз безпеки розподілених інформаційних систем : навч. посіб. / Литвинов В. В., Казимир В. В., Стеценко І. В., Трунова О. В. та ін. – Чернігів : ЧНТУ, 2016. – 254 с.

13. *Вапник В. Н.* Восстановление зависимостей по эмпирическим данным / В. Н. Вапник. – М. : Наука, 1979. – 358 с.

References

1. Denning Dorothy (1987). An Intrusion-Detection Model. *IEEE Transactions on Software Engineering,* vol. SE-13, no. 2, pp. 222–232.

2. *Amoroso, E.G.* (1999). *Intrusion Detection [Intrusion Detection]*. 1st ed., Intrusion.Net Books, Sparta, New Jersey, USA.

3. Barford, P. & Kline, J. & Plonka, D. & Ron, A. (2002). A Signal Analysis of Network Traffic Anmalies. Proceedings of *the 2nd ACM SOGCOMM Workshop on Internet Measurement* (USA, NY, New York, November 06-08), pp. 71–82.

4. Sheluhin, O.I. & Filinova, A.S. (2014). Sravnitelnii analiz algoritmov obnaruzheniya anomaliy trafika metodami diskretnogo veivlet-analiza [Comparative Analysis of Algoritms of Detection of Anomalies of the Traffic by Methods of the Discrete of Veyvlt-analysis]. *T-Comm-Telekommunikatsii i Transport – T-Comm-Telecommunications and Transport,* issue 9, vol. 8, pp. 89–97 (in Russian).

5. Orlov, S.I. (2006). *Veivlet-analys. Osnovy teorii [Veyvlt-analysis. Basic theory]*. Moscow: TEHSFERA (in Russian).

6. Sheluhin, O.I. & Pankrushin, A.V. (2015). Measuring of Reability of Network Anomalies Detections Using Methods of Discrete Wavelet Analysis. Proceedings from *Science and Information (SAI), Conference*. (UK, Longon, July 28-30), pp. 393–397.

7. Sheluhin, O.I. & Germashev, A.V. (2012). Obnaruzhenie anomalnykh vybrosov trafika metodami diskretnogo veivlet-analiza [Detection of the abnormal bursts of a traffic by methods of the discrete veyvlt-analysis]. *Elektromagnitnye volny i elektronnye sistemy – Electromagnetic waves and electronic systems*, no. 22, pp. 71–82 (in Russian).

8. Soloviov, S.A. & Iurkevskaia, L.A. (2007). Metod identifikatsii ugroz bezopasnosti informatsionnykh resursov ASU na osnove multirazreshaiushchego analiza [Method of identification of security risks of information resources of ACS on the basis of the multiallowing analysis]. *Vestnik Samarskogo gosudarstvennogo tehnicheskogo universiteta – Bulletin of the Samara state technical university,* no. 2 (20), pp. 70–76 (in Russian).

9. Darhovsky, B.S. & Brodsky, B.E. (1980). Aposteriornoe obnaruzhenie momenta «razladki» sluchainoi posledovatelnosti [A posteriori discovery of moment of "discord" of casual sequence]. *Teoriia veroiatnostei i ee primenenie – Theory of chances and her application,* vol. 25, no. 3, pp. 635–639 (in Russian).

10. Skiter, I.S. & Balchenko, I.V. (2016). Identyfykatsiia anomalnoi povedinky trafiku kompiuternoi merezhi na osnovi EWMA-statystyky [Identification of the aberrant behavior of a computer network on the basis of EWMA statistics]. Proceedings from *Persha Mizhnarodna konferentsiya "Problemy vivetdennya z ekspluatatsii obektiv yadernoi enerhetyky ta vidnovlennia otochuiuchoho seredovyshcha" INUDECO'16 – Problems of leadingout are from exploitation of objects of nuclear energy and proceeding in an environment. Conference proceedings of the 1st International conference INUDECO'16.* (Slavutich, April 25-27, 2016). Slavutich: SD NTUU «KPI», pp. 171–178 (in Ukrainian).

11. Mikova, S.Y. & Oladko, V.S. (2015). Obzor algoritmov vyiavleniia setevykh atak [Review of algorithms of exposure of network attacks]. *Aktualnye problemy gumanitarnykh i estestvennykh nauk – Actual problems of humanitarian and natural sciences,* no. 9-1, pp. 59–62 (in Russian).

12. Lytvynov, V.V., Kazymyr, V.V., Stetsenko, I.V., Trunova, O.V. et al. (2016). *Modelivannia ta analiz bezpeki rozpodilenykh informatsisnykh system [Modeling and analysis of security of distributed information systems]*. Chernihiv: ChNTU (in Ukrainian).

13. Vapnik, V.N. (1979). *Vosstanovlenie zavisimostei po empiricheskim dannym [Addiction recovery from empirical data]*. Moscow: Nauka (in Russian).

Vitalii Lytvynov, Igor Skiter, Helen Trunova, Eduard Sidin

MODIFICATION OF METHODOLOGY OF WAVELET-ANALYSIS IS FOR EXPOSURE OF ANOMALIES IN TRAFFIC OF COMPUTER NETWORK

Urgency of the research. *Reliable data transmission in the network should be based on the usage of appropriate methods of anomaly detection network traffic. Quantitative analysis of the network on the basis of a statistical approach is based on the data analysis in the form of dynamic series. Methodological support improvement of computer network traffic analysis is necessary for effective usage of wavelet analysis methods.*

Target setting. *Methods of wavelet analysis is prospective for detection of network traffic abnormal behavior because they are based on the decomposition of traffic as dynamic series of. In this case there are set of problems such as appropriate scalable wavelet functions selecting, ways of detail coefficients determining, its interpretation and hypotheses testing for the traffic behavior anomaly.*

Actual scientific researches and issues analysis. *Works which were devoted to statistical methods and techniques for analysis and anomaly detection algorithms include evaluation of anomalous traffic indicators such as: the first kind of errors, errors of the second kind, the number of correctly identified anomalies..*

Uninvestigated parts of general matters defining. *Usage wavelet analysis algorithms is associated with its specific resource consumption and complexity, the difficulties of the second kind errors detecting. In addition, there is the question of choice of the*

most appropriate scaling function and necessity of interpretation of approximating and detailing factors, which are formed a separate statistical sample.

***The research objective**. The aim is the usage Fourier series as the scaling function in the traffic wavelet analysis; construction of idealized network profile and evaluation of anomalous traffic behavior; coefficient detailing analysis as a separate statistical sample.*

***The statement of basic materials**. Traffic analysis using Fourier functions as scaling was conducted. This provided the possibility to get an amplitude in an explicit form and initial phases of harmonic component. This made it possible to compare the "ideal profile" with real traffic. Preliminary conclusions about the presence of anomalies on areas where there is an exceeding of simulated traffic over the real are conducted. Sharp change of the absolute value of detailing coefficients in the analyzed traffic windows can also be interpreted as anomaly traffic.*

***Conclusions**. The proposed Wavelet analysis method modification provides a significant reduction of analysis capacity resources of network traffic. Usage of Fourier series allows to identify trends and cyclical components in traffic, identify anomalous zones. Obtained approximating and detailing coefficients can be used as anomalous traffic characteristics in the analysis of its changes.*

***Key words**: wavelet-analysis; approximating function; Fourier series; detailing coefficients; profile of network; anomaly.*

Виталий Литвинов, Игорь Скитер, Елена Трунова, Эдуард Сидин

МОДИФИКАЦИЯ МЕТОДИКИ ВЕЙВЛЕТ-АНАЛИЗА ДЛЯ ВЫЯВЛЕНИЯ АНОМАЛИЙ В ТРАФИКЕ КОМПЬЮТЕРНОЙ СЕТИ

Для обеспечения надежной передачи данных в сети необходимо использование адекватных методов выявления аномалий, которые дадут возможность обнаруживать аномальный сетевой трафик, оценивать величину и параметры аномалии. Статистические методы анализа наиболее распространены для реализации технологий выявления аномалий. Количественный анализ сети на основе использования статистического подхода базируется на анализе массивов данных в виде динамических рядов – статистической информации по прохождению трафика.

Работа посвящена решению вопросов усовершенствования методического обеспечения анализа трафика компьютерной сети с использованием модифицированного метода вейвлет-анализа, в котором в качестве аппроксимирующей функции выступает ряд Фурьє, идентификация аномалий проводится на основе отклонений идеализирующего профиля от реального и за резкими изменениями, детализирующих коэффициентов.

***Ключевые слова:** вейвлет-анализ; аппроксимирующая функция; ряд Фурьє; детализирующие коэффициенты; профиль сети; аномалии.*

Литвинов Віталій Васильович – доктор технічних наук, професор, завідувач кафедри інформаційних технологій та програмної інженерії, Чернігівський національний технологічний університет (вул. Шевченка, 95, м. Чернігів, 14027, Україна).

Литвинов Виталий Васильевич – доктор технических наук, профессор, заведующий кафедрой информационных технологий и программной инженерии, Черниговский национальный технологический университет (ул. Шевченко, 95, г. Чернигов, 14027, Украина).

Lytvynov Vitalii – Doctor of Technical Sciences, Professor, Head of Department of Information Technology and Software Engineering, Chernihiv National University of Technology (95 Shevchenka Str., 14027 Chernihiv, Ukraine).

E-mail: vlitvin@ukrsoft.ua

ORCID: http://orcid.org/0000-0003-2334-2275

Скітер Ігор Семенович – кандидат фізико-математичних наук, доцент, доцент кафедри інформаційних технологій та програмної інженерії, Чернігівський національний технологічний університет (вул. Шевченка, 95, м. Чернігів, 14027, Україна).

Скитер Игорь Семенович – кандидат физико-математических наук, доцент, доцент кафедры информационных технологий и программной инженерии, Черниговский национальный технологический университет (ул. Шевченко, 95, г. Чернигов, 14027, Украина).

Skiter Igor – PhD in Physical and Mathematical Sciences, Assistant Professor, Assistant Professor of Department of Information Technology and Software Engineering, Chernihiv National University of Technology (95 Shevchenka Str., 14027 Chernihv, Ukraine).

E-mail: skiterigors@gmail.com

ORCID: http://orcid.org/0000-0003-2334-2276

ResearcherID: F-5950-2014

Трунова Олена Василівна – кандидат педагогічних наук, доцент, доцент кафедри інформаційних технологій та програмної інженерії, Чернігівський національний технологічний університет (вул. Шевченка, 95, м. Чернігів, 14027, Україна).

Трунова Елена Васильевна – кандидат педагогических наук, доцент, доцент кафедры информационных технологий и программной инженерии, Черниговский национальный технологический университет (ул. Шевченко, 95, г. Чернигов, 14027, Украина).

Trunova Helen – PhD in Pedagogical Sciences, Assistant Professor, Assistant Professor of Department of Information Technology and Software Engineering, Chernihiv National University of Technology (95 Shevchenka Str., 14027, Chernihiv, Ukraine).

E-mail: e.trunova@gmail.com

ORCID: http://orcid.org/0000-0003-0689-8846

Сідін Едуард Пилипович – кандидат технічних наук, доцент, науковий співробітник, Державний науково-випробувальний центр Збройних сил України (вул. Стрілецька, 1, м. Чернігів, 14033, Україна).

Сидин Эдуард Филиппович – кандидат технических наук, доцент, научный сотрудник, Государственный научно-испытательный центр Вооруженных сил Украины (ул. Стрелецкая, 1, г. Чернигов, 14033, Украина).

Sidin Eduard – PhD in Technical Sciences, Assistant Professor, Researcher of State Research and Test Center of the armed forces of Ukraine (1 Striletska Str., 14033 Chernihiv, Ukraine).

E-mail: sidin.e.f@mail.ru

Модифікація методики вейвлет-аналізу для виявлення аномалій у трафіку комп'ютерної мережі / В. Литвинов, І. Скітер, О. Трунова, Е. Сідін // Технічні науки та технології. – 2017. – № 2 (8). – С. 99-109.