УДК 004.4
DOI https://doi.org/10.32782/2663-5941/2025.3.2/59

Чижмотря О.В.
Державний університет «Житомирська політехніка»

Вакалюк Т.А.
Державний університет «Житомирська політехніка»

Клименко Д.О.
Державний університет «Житомирська політехніка»

Чижмотря О.Г.
Державний університет «Житомирська політехніка»

Гордієнко І.В.
Дрогобицький державний педагогічний університет імені Івана Франка

ДОСЛІДЖЕННЯ СТРУКТУРИ ВЕБ-ТРАФІКУ ТА ВИЗНАЧЕННЯ ВИРІШАЛЬНИХ ХАРАКТЕРИСТИК ЗАПИТІВ

У статті представлено результати комплексного дослідження структури веб-трафіку з метою визначення найбільш інформативних характеристик для ефективного виявлення кіберзагроз та аномальної мережевої активності. В умовах експоненціального зростання обсягів веб-трафіку та еволюції методів кібератак традиційні сигнатурні підходи до забезпечення безпеки стають недостатніми.

Розглянуто багаторівневу архітектуру HTTP/HTTPS запитів, включаючи детальний аналіз заголовків, параметрів URL, тіла запитів, типів методів та протоколів передачі даних. Особливу увагу приділено дослідженню заголовків запитів, відповідей та загальних заголовків, які містять критично важливу інформацію про середовище клієнта та параметри авторизації.

Проведено систематизацію ключових характеристик мережевих потоків, що охоплює часові параметри (тривалість сесій, інтервали між пакетами, швидкість передачі), розмірні характеристики (статистичні показники розмірів пакетів) та контрольні біти TCP. Детально досліджено роль прапорців PSH та URG у контексті ідентифікації трафіку реального часу та виявлення специфічних типів атак.

Особливу увагу приділено аналізу ролі сесійних даних та cookie-файлів у контексті поведінкового моніторингу користувачів та виявлення аномальних змін у поведінкових патернах.

Досліджено дискримінативну здатність різних типів ознак та запропоновано їх трирівневу класифікацію за ступенем важливості для задач машинного навчання. Встановлено, що критичними ознаками є заголовки User-Agent та Authorization, характеристики тіла запитів та частотні параметри HTTP-методів.

Результати створили міцну теоретичну основу для розробки ефективних алгоритмів машинного навчання в галузі кіберзахисту. Практичне значення полягає у можливості оптимізації систем IDS/IPS та створення адаптивних систем кіберзахисту.

Ключові слова: *веб-трафік, HTTP-запити, виявлення аномалій, кіберзагрози, мережева безпека, машинне навчання, аналіз заголовків, класифікація трафіку.*

Постановка проблеми. У сучасному світі використання інтернет-ресурсів є невід›ємною частиною повноцінного існування як окремих користувачів, так і організацій. За останні десятиліття спостерігається експоненціальне зростання обсягів веб-трафіку, що зумовлено стрімким розвитком цифрових технологій, впровадженням хмарних сервісів та загальною цифровізацією суспільства.

Згідно з даними провідних телекомунікаційних компаній, щомісячний глобальний IP-трафік зростає з темпами понад 20% на рік, що створює як нові можливості, так і серйозні виклики для забезпечення кібербезпеки. Паралельно з легі-

тимним трафіком зростає і кількість кібератак: від простих спроб несанкціонованого доступу до складних багатоетапних атак, включаючи DDoS-атаки, SQL-ін'єкції, крос-сайтовий скриптинг (XSS) та атаки типу «людина посередині».

Традиційні методи виявлення загроз, що базуються на сигнатурному аналізі та правилах, вже не можуть ефективно справлятися з сучасними викликами кібербезпеки. Це пов'язано з тим, що зловмисники постійно удосконалюють свої методи, використовують поліморфні техніки та застосовують штучний інтелект для обходу систем захисту. У таких умовах особливої актуальності набувають методи поведінкового аналізу та машинного навчання, які дозволяють виявляти раніше невідомі типи атак на основі аномалій у структурі та характеристиках веб-трафіку.

Аналіз структури веб-трафіку передбачає детальне вивчення HTTP/HTTPS запитів, включаючи заголовки, параметри, тіло запитів, часові характеристики, сесійні дані та метадані мережевих потоків. Кожен з цих компонентів містить цінну інформацію, яка може свідчити про нормальну або аномальну поведінку користувачів. Виявлення та класифікація таких аномалій є критично важливим завданням для своєчасного реагування на потенційні загрози та мінімізації можливих збитків.

Сучасні дослідження в галузі кіберзахисту показують, що ефективність систем виявлення вторгнень значно підвищується при використанні комплексного підходу до аналізу веб-трафіку, який поєднує статистичний аналіз, методи машинного навчання та глибокого навчання. Особливу увагу приділяють виділенню найбільш інформативних ознак (feature selection), що дозволяє підвищити точність класифікації та зменшити обчислювальні витрати.

Аналіз останніх досліджень і публікацій. Сучасні дослідження в галузі аналізу веб-трафіку та виявлення аномалій демонструють стрімкий розвиток методів машинного навчання та глибокого навчання для забезпечення кібербезпеки.

Так, у [1] підкреслюється критичну важливість аналізу структури даних для ідентифікації потенційних загроз. Автори систематизували основні виклики, пов›язані з обробкою великих обсягів гетерогенних даних, та запропонували комплексний підхід до виявлення аномальної поведінки на основі статистичних методів та машинного навчання.

Особливо цінним є дослідження [2], в якому розроблено програмну систему глибокої інспекції пакетів для аналізу мережевого трафіку та виявлення аномалій. Запропонована система використовує багаторівневий підхід до аналізу, що включає детальне вивчення заголовків пакетів, часових характеристик потоків та статистичних параметрів мережевої активності.

Автори [3] зосередилися на застосуванні методів глибокого навчання для прогнозування мережевого трафіку. Їхнє дослідження продемонструвало ефективність рекурентних нейронних мереж та згорткових нейронних мереж у задачах класифікації та прогнозування аномального трафіку.

Науковці [4] запропонували комплексний підхід до проектування та впровадження систем виявлення вторгнень на основі глибоких нейронних мереж. Їхнє дослідження включає детальний аналіз архітектурних рішень та оптимізацію гіперпараметрів для досягнення максимальної ефективності виявлення загроз.

У [5] досліджено вразливості систем виявлення вторгнень на основі глибоких нейронних мереж до адверсаріальних атак. Їхня робота розкриває критичні аспекти безпеки самих систем виявлення та пропонує методи підвищення стійкості до цілеспрямованих спроб обходу захисту.

У [6] представлено інноваційний підхід до розуміння внутрішньої структури глибоких нейронних мереж через концепцію «машини Джейнса». Їхнє дослідження надає теоретичну основу для оптимізації архітектури нейронних мереж у задачах класифікації мережевого трафіку.

У [7] і детально розглянуто математичні основи гауссівського розподілу та його застосування в задачах аналізу даних. Хоча дослідження носить більш теоретичний характер, воно надає важливі дані щодо статистичних методів аналізу аномалій у великих масивах даних.

Аналіз досліджень показує, що сучасні підходи до виявлення аномалій у веб-трафіку характеризуються переходом від традиційних систем до складних гібридних архітектур, що поєднують статистичні методи, машинне навчання та глибоке навчання. Однак залишається недостатньо дослідженим питання систематизації та ранжування характеристик веб-запитів.

Постановка завдання. Мета роботи – дослідити структуру веб-трафіку та визначити найбільш вирішальні характеристики HTTP/HTTPS запитів, які дозволяють ефективно ідентифікувати аномальну поведінку та потенційні кіберзагрози в мережевому трафіку.

Виклад основного матеріалу. Ефективним із способів виявлення потенційних загроз є аналіз структури запитів та вивчення активності користувачів, що дозволяє виявляти відхилення від нормальної поведінки. Це ж і допоможе у навчання моделей та класифікації трафіку.

Веб-запити, які здійснюють користувачі, містять у собі значну кількість інформації, яка може бути використана для визначення нормальної чи аномальної поведінки. Структура цих запитів, типи HTTP запитів, параметри, заголовки та інші метадані можуть мати значний вплив на класифікацію дії. Також під час запитів важливо мати можливість збору та аналізу даних, які містяться у сесійних даних, журналах веб-серверу та cookie-файлів [1].

HTTP-запити є основним елементом, який використовують клієнти для комунікації із серверами [2]. Кожен запит містить важливу інформацію, яка може бути використана для аналізу взаємодії користувачів з вебсайтом. Розберемо структуру заголовків та склад запитів (рис. 1).

Одним із елементів запиту є протокол. Він забезпечує набір правил і стандартів для здійснення комунікації між пристроями в мережі, а саме визначає яким чином передаються дані. У дослідженні роботи використовуються протоколи для інтернет передачі даних, а саме TCP та UPD [2].

Використання HTTP або HTTPS теж важливо для ідентифікації, хоча і в сучасних системах налаштовується переадресація на захищений протокол. Це більш категоріальна змінна, яка не може впливати на те чи буде запит аномальним чи ні, але при потребі може виступати елементом фільтрації, якщо моделі готуються для окремих протоколів.

Посилання містить у собі інформацію для ідентифікацію ресурсу направлення запиту. Саме посилання ідентифікують ресурс, до якого йшов запит. Таким чином при ідентифікації аномалії можна буде визначити що було ціллю зловмисних дій. На аномалію можуть вказувати підозрілі параметри запитів, або часті спроби доступу до конкретного посилання. Збереження даного елементу є важливим для ідентифікації та є частиною моніторингу.

Аналогічна важливість і збереження типу запиту. Якщо це RESTful API, то типи запиту можуть бути GET, POST, PUT, PATCH, DELETE, HEAD, OPTIONS. При виявленні аномалій можна звернути увагу на частоту запитів за певним типом, це може вказувати на DDoS або бутрфорс атаки.

Аналіз заголовків є надзвичайно важливим для вивчення при визначенні аномалій. Хоч вони і не містять дані, але визначають тип оброблення. Заголовки можна поділити на декілька категорій [3]. Заголовки запиту, які клієнт надсилає разом із запитом. Містять усобі інформацію про середовище клієнта, інтерфейс користувача, авторизацію. Заголовки відповіді, які сервер надсилає клієнту разом із відповіддю. Містять у собі

TCP Segment Header Format

Bit #	0 – 7	8 – 15	16 – 23	24 – 31
0	Source Port		Destination Port	
32	Sequence Number			
64	Acknowledgment Number			
96	Data Offset	Res	Flags	Window Size
128	Header and Data Checksum		Urgent Pointer	
160...	Options			

UDP Datagram Header Format

Bit #	0 – 7	8 – 15	16 – 23	24 – 31
0	Source Port		Destination Port	
32	Length		Header and Data Checksum	

Рис. 1. Структура заголовків запиту

інформацію про тип контенту відповіді, статус, кешування. Загальні заголовки можуть бути присутні у запитах та відповідях та містять у собі загальну мета-інформацію.

При дослідженні аномальних значень заголовки можуть вказувати на спроби обходу безпеки, наприклад фальшиві заголовки User-Agent для маскування клієнта або спроби підробити Authorization заголовки для отримання авторизованого доступу. Також про аномалію може свідчити відсутність стандартних заголовків, що може вказувати на використання ботів або заскиртованих сценаріїв. Також на аномалію може вказувати і наявність непотрібних заголовків, що може свідчити про спроби втручання в сесії.

Тіло запиту є надзвичайно важливим для визначення аномальної поведінки, адже саме в ньому зберігається інформація у складних форматах, яку сервер використовує для обробки та збереження. Для аналізу тіла, з метою забезпечення безпеки, звертають увагу на розмір, перевищення або надмірне зменшення його. Також важливим є формат, який передбачений або не передбачений системою. Саме ж наповнення тіла також може мати у собі загрози, якщо містить у собі елементи виконання коду або передаються дані, невідповідні до очікуваних.

У рамках дослідження трафіку важливо звертати увагу не лише на тіло запиту та відповіді, а і на сам потік, який може включати у себе декілька запитів [4]. Потік – це послідовність пакетів, які передаються у мережі між пристроями у межах одного безперервного з'єднання. При аналізі потоку з метою визначення важливих характеристик для ідентифікації аномалій важливо звернути увагу на такі особливості [2].

Основні характеристики потоку, що включає тривалість потоку, загальну кількість пакетів відправлення та отримання, а також загальний розмір пакетів відправлення та отримання.

Часові показники потоку, що включають швидкість передачі байтів та пакетів при відправленні, отриманні та в загальному; часові інтервали між потоками (середній, максимальний, мінімальний, відхилення від середнього), а також часові інтервали між пакетами в обох напрямках окремо та в загальному.

Також при дослідженні трафіку важливо звернути увагу на прапорці, такі як PSH або URG [3]. Ці контрольовані біти впливають на чергу оброблення всіх або окремих пакетів у TCP запитах відповідно. Зазвичай вони використовуються при обміні даних у реальному часі, що може включати відео або голосові дзвінки, а також передачі сигналів на термінове виконання певних команд. Таким чином ці дані можуть мати одне із вирішальних значень при класифікації аномальної поведінки.

Сесійні дані також є важливою складовою для відстеження взаємодії користувача з веб-сайтом. За допомогою них можна зберігати стан користувача впродовж певного часу в різних ситуаціях, таких як авторизація, кошик, налаштування браузера на сайті. Завдяки моніторингу компонентів сесії можна визначити послідовність змін, які можуть впливати на класифікацію дій користувача. Для визначення аномальної поведінки можна слідкувати за такими даними сесії: ідентифікатор сесії, тривалість, активність користувача на сторінках [3; 6].

Аналогічно до сесій також можна слідкувати за зміною файлів cookies. Ці невеликі файли зберігають у собі інформацію про стан сесії, персональні налаштування браузера, часто використовуються для ідентифікації користувачів. Таким чином різка зміна файлів cookies між запитами може слідчити про аномальну активність.

Використовуючи сесії та файли cookies можна зберігати інформацію для моніторингу для подальшого аналізу. Для прикладу можна додати додаткові змінні, які будуть визначати поведінкові характеристики користувача. Це може бути час перебування на сайті, частота входів, певні дії з інтерфейсом вебсайту, які можуть бути вирішальними.

Аналіз структури веб-трафіку, зокрема зазначених вище елементів, можуть дати повну інформацію для ідентифікації класу активності у веб-системі.

Висновки. У результаті проведеного дослідження структури веб-трафіку та визначення вирішальних характеристик для ідентифікації аномальної поведінки було встановлено, що сучасні веб-запити містять багаторівневу ієрархію інформативних елементів, кожен з яких має різну цінність для виявлення потенційних кіберзагроз.

Детальний аналіз структури HTTP/HTTPS запитів показав, що найбільш значущими компонентами для виявлення аномалій є заголовки запитів, які містять критично важливу інформацію про клієнтське середовище, авторизацію та параметри сесії. URL-параметри та тіло запитів дозволяють ефективно виявляти спроби ін›єкцій та несанкціонованої передачі даних, тоді як ано-

мальна частота певних типів HTTP-методів може свідчити про DDoS або брутфорс атаки.

Дослідження ключових характеристик мережевих потоків виявило, що для ефективної класифікації трафіку необхідно враховувати комплекс часових та розмірних параметрів. Часові характеристики, такі як тривалість потоків, інтервали між пакетами та швидкість передачі даних, виявилися особливо інформативними для розпізнавання аномальної активності. Розмірні характеристики пакетів, включаючи їх максимальні, мінімальні та середні значення разом зі статистичними відхиленнями, також надають цінну інформацію для класифікації. Контрольні біти TCP, зокрема PSH та URG, виявилися критично важливими для ідентифікації трафіку реального часу та можуть слугувати індикаторами специфічних типів атак.

Аналіз ролі сесійних даних та cookie-файлів продемонстрував їх високу цінність для поведінкового аналізу користувачів. Моніторинг ідентифікаторів сесій дозволяє ефективно виявляти спроби перехоплення сесій, а аналіз тривалості та активності користувачів на сторінках надає можливість створення детальних поведінкових профілів. Різкі зміни в cookie-файлах між запитами часто індикують аномальну активність та можуть свідчити про компрометацію облікового запису або спроби несанкціонованого доступу.

Список літератури:

1. Min Yang and Jiajie Zhang. Data Anomaly Detection in the Internet of Things: A Review of Current Trends and Research Challenges. International Journal of Advanced Computer Science and Applications (IJACSA), 14 (9), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140901

2. Song, W., Beshley, M., Przystupa, K., Beshley, H., Kochan, O., Pryslupskyi, A., Pieniak, D., & Su, J. A Software Deep Packet Inspection System for Network Traffic Analysis and Anomaly Detection. *Sensors*, 20 (6), 2020, 1637. https://doi.org/10.3390/s20061637

3. R. Vinayakumar, K. P. Soman and P. Poornachandran. Applying deep learning approaches for network traffic prediction. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017, P. 2353–2358, DOI: 10.1109/ICACCI.2017.8126198.

4. Edosa Osa, Patience E. Orukpe, Usiholo Iruansi. Design and implementation of a deep neural network approach for intrusion detection systems, e-Prime – Advances in Electrical Engineering, Electronics and Energy, Volume 7, 2024, 100434, https://doi.org/10.1016/j.prime.2024.100434

5. Xingwei Zhang, Xiaolong Zheng, Desheng Dash Wu. Attacking DNN-based Intrusion Detection Models. IFAC-PapersOnLine, Volume 53, Issue 5, 2020, P. 415–419, https://doi.org/10.1016/j.ifacol.2021.04.118.

6. Venkat Venkatasubramanian, N. Sanjeevrajan, Manasi Khandekar, Abhishek Sivaram, Collin Szczepanski. Jaynes machine: The universal microstructure of deep neural networks. Computers & Chemical Engineering, Volume 192, 2025, 108908, https://doi.org/10.1016/j.compchemeng.2024.108908.

7. Chen,S. The Gaussian distribution: Derivation method and its applications in selected examples. Theoretical and Natural Science, 38, 2024, P. 196–202. 10.54254/2753-8818/38/20240595

Chizhmotria O.V., Vakaliuk T.A., Klymenko D.O., Chizhmotria O.H., Hordienko I.V. RESEARCH OF WEB TRAFFIC STRUCTURE AND DETERMINATION OF DECISIVE CHARACTERISTICS OF REQUESTS

The article presents the results of a comprehensive study of web traffic structure to identify the most informative characteristics for effectively detecting cyber threats and abnormal network activity. With the exponential growth of web traffic and the evolution of cyberattack methods, traditional signature-based approaches to security are becoming insufficient.

The multi-level architecture of HTTP/HTTPS requests is considered, including a detailed analysis of headers, URL parameters, request bodies, method types, and data transfer protocols. Particular attention is paid to the study of request headers, responses, and general headers that contain critical information about the client environment and authorisation parameters.

The key characteristics of network flows are systematised, covering time parameters (session duration, packet intervals, transmission speed), dimensional characteristics (statistical indicators of packet sizes) and TCP control bits. The role of PSH and URG flags in real-time traffic identification and detection of specific types of attacks has been studied in detail.

Particular attention has been paid to analysing the role of session data and cookie files in monitoring user behaviour and detecting abnormal changes in behaviour patterns.

The discriminative ability of different types of features has been investigated, and a three-level classification of their importance for machine learning tasks has been proposed. It has been established that the critical features are User-Agent and Authorization headers, request body characteristics and HTTP method frequency parameters.

The results have created a solid theoretical basis for the development of effective machine learning algorithms in cyber security. The practical significance lies in the possibility of optimising IDS/IPS systems and creating adaptive cyber security systems.

Key words: *web traffic, HTTP requests, anomaly detection, cyber threats, network security, machine learning, header analysis, traffic classification.*