https://www.sworldjournal.com/index.php/swj/article/view/swj21-01-015
DOI: 10.30888/2663-5712.2023-21-01-015

УДК 004-049.5

INFORMATION TECHNOLOGY OF SECURE ACCESS TO DNS RESOURCES BASED ON ML-TRAINED TRAFFIC IDENTIFICATION MODELS

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ БЕЗПЕЧНОГО ДОСТУПУ ДО РЕСУРСІВ DNS НА БАЗІ ML-ТРЕНОВАНИХ МОДЕЛЕЙ ІДЕНТИФІКАЦІЇ ТРАФІКУ

Fedchuk T.B. / Федчук Т.Б.
аспірант / postgraduate
ORCID: 0009-0003-5996-6467
Korobeinikova T.I. / КоробейніковаТ.І.
c.t.s., doc. / к.т.н., доц.
ORCID: 0000-0003-2487-8742
Lviv Polytechnic National University, S. Bandera St. 12, Lviv, 79013
Національний університет «Львівська Політехніка», Львів, Бандери, 12, 79013

Анотація. *Система доменних імен (DNS) займається перетворенням IP-адреси сервера в доменне ім'я, що дає можливість кінцевому користувачу отримувати доступ до ресурсу, не запам'ятовуючи його IP-адреси. Даний протокол є основою сучасного інтернету, проте всі повідомлення між клієнтом та сервером проходять по незахищеному каналу зв'язку, що робить його вразливим до різного роду атак (Spoofing, Eavesdropping, Phishing та інших). Щоб подолати цю проблему – розробили новий протокол DNS over HTTPS (DoH), що здатний шифрувати DNS трафік між клієнтом та сервером. В даній статті будуть описані засоби дослідження виявлення, аналізу та розпізнавання шкідливого DNS-трафіку на основі аналізаторів трафіку та методу машинного навчання. Будуть запропоновані комплексні методи та подані порівняльні характеристики вже досліджених моделей виявлення DoH-трафіку. Для максимально-ефективних параметрів: Accuracy, Precison, Sensitivity, F Score, Specificity та мінімальній затримці сигналу буде використаний гібридний метод дослідження шкідливого DNS-трафіку, що базується на комбінованому застосуванні аналізаторів трафіку, машинного навчання та людського досвіду для отримання статистичних даних.*

Ключові слова: *Класифікатори трафіку, Система доменних імен, DoH, дерево Adaboost, Домен верхнього рівня, RF.*

Вступ.

Стрімкий розвиток сучасного інтернету, в тому числі, великий обсяг веб-сайтів, дозволяє усвідомити яку важливу роль відіграє ієрархічна система доменних імен DNS в інформаційному просторі. Протокол DNS по суті перетворює доменні імена в IP адреси, що дозволяє браузерам завантажувати та використовувати ресурси інтернету. Незважаючи на свою незамінність, DNS є вразливим до різного роду атак(Spoofing, Eavesdropping, Phishing), якими зловмисники постійно зловживають. Тому доставка безпечного DNS-трафіку набуває високого значення, оскільки зловмисники використовують передові підходи та швидкі методи для перехоплення, прослуховування та крадіжки інформації [1], [2].

Для подолання вразливостей DNS – використовується протокол DNS over HTTPS (DoH). Для підвищення безпеки протоколу DNS – було введено метод шифрування трафіку і його передачі через прихований канал мережі. Для ефективного дослідження, виявлення та аналізу зловмисного DoH-трафіку – є актуальними методи машинного навчання. [1], [2].

За проведеними дослідженнями Qasem Abu Al-Haija, Manar Alohaly, Ammar Odeh, які представили двоступеневу схему виявлення зловмисного трафіку DoH за допомогою гібридного підходу до навчання і запропонували двошарову систему. На першому рівні трафік досліджувався за допомогою випадкових тонких дерев (RF) та ідентифікується як трафік DoH або не DoH. На другому рівні трафік DoH додатково досліджується за допомогою дерева Adaboost(ADT) і визначені як незагрозливий DoH або шкідливий DoH. Зокрема, запропонована система працює з найменшою кількістю функцій, вибрані за допомогою аналізу головних компонентів, які мінімізують кількість зразків, отримані з використанням підходу випадкової недостатньої вибірки. Експериментальна оцінка повідомила, що високопродуктивна система з точністю прогнозування 99,4% і 100% і передбачуваними накладними витратами 0,83 с і 2,27 с для першого і другого шару відповідно. Що наглядно доводить ефективність машинного навчання у дослідженні DoH-трафіку [1].

Наукова публікація присвячена подальшому розвитку і дослідженню технології DNS із використанням протоколів шифрування та ідентифікації й аналізу шкідливого трафіку, на основі алгоритмів машинного навчання.

Метою даної роботи є підвищення ефективності виявлення та розпізнавання шкідливого DNS-трафіку, на базі алгоритмів машинного навчання для забезпечення безпеки та конфіденційності даних в межах клієнт-серверних сесій.

Таким чином, з’являється можливість досягти більш ефективного методу виявлення та розпізнавання шкідливого DoH-трафіку. Для цього планується розробити комплексну систему ефективності ідентифікації DNS-трафіку із використанням протоколів шифрування на основі алгоритмів машинного навчання.

1 Технічні деталі в галузі безпечного доступу до ресурсів DNS.

Сервіс DNS. Кожний хост в інтернеті має унікальну IP-адресу, яка дозволяє користувачам підключатися та спілкуватися з ним. На початку існування мережі інтернет – користувачі могли отримати доступ лише до веб-сервера використовуючи IP-адресу сервера. Наприклад, щоб відвідати веб-сайт Google, користувач повинен ввести IP-адресу сервера, 142.251.208.142, замість www.google.com.

Пізніше, у 1980-х роках, кількість інтернет-хостів зросла до сотень тисяч. У результаті це стало непрактичним запам'ятовувати та підтримувати IP-адресу кожного окремого хоста в цій мережі [3]. Paul Mockapetris вирішив цю проблему, представивши систему доменних імен (DNS). Ця система розпізнавання імен перетворює ім’я хоста на його IP-адресу [4].

Як спочатку було розроблено, DNS має ієрархічну деревоподібну структуру, що складається з трьох рівнів: кореневого рівня, верхнього рівня рівень домену (TLD) і авторитетний рівень [4]. Процес зіставлення імені з IP-адресою починається, коли інтернет-клієнти, такі як веб-браузери, ініціюють запит DNS і надсилають його до резолвера [4]. Резолвер передає запит через різні сервери для пошуку відповідної IP-адреси та надсилає її назад клієнту, як показано на рисунку 1.

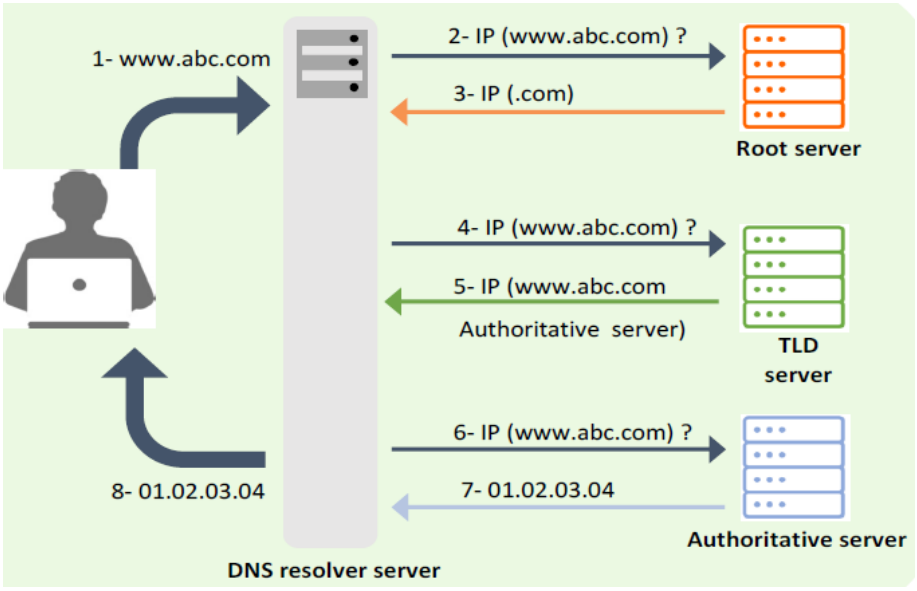

Рисунок 1 – Схема функціонування DNS

Джерело: [1]

1.1 Вразливості сервісу DNS. Трафік DNS не зашифрований. Така комунікація між клієнтом та сервером у відкритому вигляді дозволяє зловмисникам здійснювати атаки на передані пакети DNS [5]. Відповідно до рапорту IDC DNS – у звіті опитування понад 88% організацій зазнали атаки на основі DNS протягом 2022 року, в середньому сім атак на організацію. У звіті також виявлено, що DNS фішинг, hjaking/spoofing DNS, malware зі спрямуванням на DNS та DoS/DDoS-атаки були найпоширенішими атаками на основі DNS. В результаті організації втратили в середньому 942 тис. доларів за атаку [6].

1.2 Протоколи DNS, DoT та DoH. Шифрування трафіку DNS запобігає зловмисникам від перехоплення зв'єднання між кінцевим користувачем і DNS-резолвером. Щоб покращити безпеку DNS, дослідники запропонували два протоколи: DNS over TLS і DNS over HTTPS, DoT і DoH відповідно[7,8]. DNS over TLS (DoT) – це протокол безпеки, який інкапсулює DNS-запит і відповідь у стандартний пакет безпеки транспортного рівня (TLS). Використання DoT, веб-клієнта ініціює сеанс TLS із резолвером, перевіряє його сертифікати публічного ключа та обчислює секретний ключ. Після встановлення сеансу відбувається обмін зашифрованим трафіком DNS між обома сторонами через окремий порт (853) [7].

Подібним чином DNS over HTTPS (DoH) шифрує трафік DNS для збереження конфіденційності і цілісності DNS-з'єднання. На відміну від DoT, DoH, який є новішим, не передає DNS-дані з трафіком TLS, але передає дані за допомогою повідомлень HTTPS. Ці HTTPS-повідомлення надсилаються через порт tsp/443, як типовий трафік HTTPS [8].

За допомогою протоколу DoT виділений порт підтримує традиційну фільтрацію на основі портів. Це дозволяє мережевим адміністраторам контролювати та блокувати DNS-трафік для захисту мережі від зловмисників, зберігаючи при цьому конфіденційність зв'язку DNS. Проте він має недолік – відкриває виділений порт для зловмисників.

Таким чином, зловмисник може атакувати виділений порт (853) великим обсягом трафіку, щоб зупинити роботу DoT. З іншого боку, прийняття протоколу DoH, який інкапсулює DNS трафік у запитах HTTPS, робить DNS-з'єднання менш видимим для традиційних засобів фільтрації на основі портів. Відсутність видимості в мережі свідчить про те, що атаки можуть залишитися непоміченими. Зловмисники можуть використовувати протокол DoH для створення прихованих каналів із зовнішніми командно-контрольними серверами, для викрадання даних тощо [9].

1.3 Розвиток протоколів безпечного доступу до ресурсів доменних структур. IETF прийняв протокол DoH як документ RFC (RFC 8484 [10]) у 2018 р. Наразі існує дві суттєво відмінні реалізації. RFC 8484 використовує класичний DNS "Wireformat" [11] інкапсульований у протоколі HTTPS. Повідомлення також передаються за допомогою запитів HTTP GET або POST. Інший підхід використовує DNS-повідомлення, закодовані в описаному форматі JSON у RFC 8427 [12]. Потім дані JSON передаються через HTTPS GET. Наразі більшість DNS провайдерів (близько 90%) підтримують «Wireformat», HTTPS GET або POST версія [13]. DoH на основі JSON підтримується навколо 30% провайдерів DNS [14]. На практиці всі браузери з підтримкою DoH і більшість інших, орієнтованих на продуктивність клієнтів DoH використовують повідомлення Wireformat, сумісні з RFC 8484 разом із методом HTTPS POST.

Підхід JSON також має свої переваги. Основна причина для закодування DNS-запиту у JSON є підвищення читабельності і легке маніпулювання даними на основі текстових повідомлень. За спостереженнями дослідників Karel Hynek, Dmyto Vekshin, Jan Luxemburk, Tomas Cejka, Armin Wasicek – JSON використовується переважно для одного запиту додатками, де продуктивність і короткий час відповіді не є пріоритетом [2].

Затримка протоколу DNS безпосередньо впливає на продуктивність мережевих програм [15]. Тому багато дослідників вимірювали наслідки розгортання DoH щодо продуктивності. Ці дослідження підсумовані в таблиці 1. Налаштування – дані вимірювання та їх походження, результати – основні висновки вимірювання щодо впливу DoH на продуктивність порівняно з традиційним DNS.

Одне з перших вимірювань затримки DoH було опубліковано McManus [16] з Mozilla у 2018 році, показуючи, що середня затримка програм, спричинена DoH, становить лише 6 мс. Наступне дослідження, проведене Böttger et al. [17] зосереджено на перевиконанні DoH порівняно з традиційним DNS. Їх результати показують, що DoH додає значну затримку, коли з'єднання використовується для одного запиту. Однак, коли підключення DoH повторно використовується для кількох запитів, додаткова затримка незначна. Інше дослідження, проведене Hounsel et al. [18] показує, що затримка DoH і надійність сильно залежать від вибраного резолвера.

Це також підтверджено Jerabek et al. [19], які вивчали поведінку розпізнавача DoH і розподіл розмірів пакетів DoH залежно від використовуваних резолверів. Згідно з їх результатами, деякі резолвери DoH використовують довгі заголовки HTTP, що призводить до більших пакетів і, отже, до більших

накладних витрат.

Таблиця 1 – Порівняння досліджень, пов'язаних з продуктивністю DoH. Вимірювання.

Автор	Рік	Параметри вимірювання	Результати
MCManus [18]	2018	Користувачі Firefox	Незначний вплив, додана затримка 6 мс
Böttger et al. [19]	2019	Один клієнт	Незначний вплив на затримку під час повторного використання з'єднання
Borgolte et al. [16]	2019	Самоемульовані мережеві умови	Вибірковий вплив, в залежності від умов мережі
Hounsel et al. [23]	2020	Самоемульовані мережеві умови	Вибірковий вплив, в залежності від умов мережі
Hounsel et al. [20]	2021	Згенеровано через кінцеві точки по всій Пн. Америці	Вибірковий вплив, залежно від використовуваного резолвера DoH
Chhabra et al. [22]	2021	Глобальні вимірювання серед 224 країн	Вибірковий вплив, залежно від умов мережі
Mbewe et al. [24]	2021	Згенеровано через кінцеві точки по всій Африці	Вибірковий вплив, залежно від умов мережі
Jerabek et al. [21]	2022	Згенеровані, одна локація	Вибірковий вплив, залежно від використовуваного резолвера DoH

Джерело: [2]

Більш детальне дослідження було виконано Chhabra et al. [20]. Їх результати показують, що користувачі з країн із високим рівнем доходу та якіснішою інтернет-інфраструктурою мають меншу ймовірність сповільнити продуктивність, спричинену DoH, що призведе до непропорційного впливу на користувачів із країн із меншою економічною спроможністю. Їх висновки також підтверджуються дослідженнями, проведеними Hounsel et al. [21], Borgolte та ін. [16] і Mbewe et al. [22], які показують, що DoH має незначний вплив при малих мережевих затримках. Згідно [16], [23], [22], традиційний DNS значно перевершує DoH при роботі з перевантаженими або мобільними мережами 3G.

Станом на 2023 рік – DoH підтримується (й іноді вмикається за замовчуванням) більшістю сучасних веб-браузерів, таких як Chrome (з версії 838), Edge, Firefox, Opera та Brave. Існують також рідні резолвери з підтримкою DoH у Microsoft Windows [23] і сучасних дистрибутивах GNU/Linux (наприклад, через systemresolved). DoH підтримується основним програмним забезпеченням сервера доменних імен, таким як BIND (з версії 9.17.10), KNOT resolver (з версії 5.2.0) і Unbound (з версії 1.12.0). Існує проксі-сервер DoH від Cloudflare під назвою cloudflared. Відомо щонайменше вісім клієнтських реалізацій DoH і щонайменше шість серверних реалізацій і перераховані на dnscrypt.info9.

За цей період часу також було проведено багато досліджень по інтенсивності використання трафіку DoH і з'явилося багато наукових публікацій, проте все ще є теми для дослідження.

2 Методи та засоби аналізу та виявлення шкідливого трафіку.

2.1 Методи машинного навчання. DoH трафік – це трафік DNS, що інкапсулюється в протокол HTTPS. Таким чином DNS шифрується сертифікатом SSL й ідентифікувати чи це зловмисний трафік чи не шкідливий – складніше. Щоб виявити зловмисний трафік DoH, системі може знадобитися впровадження спеціальних інструментів і методів, таких як глибока перевірка пакетів [27,28 аналіз поведінки [29,30] або алгоритми машинного навчання [31,32].

Ефективніше використовувати метод гібридного навчання, який поєднує алгоритми машинного навчання з людським досвідом для точного виявлення та класифікації шкідливого трафіку. Гібридний метод навчання може точно виявляти та класифікувати зловмисний трафік DoH шляхом поєднання алгоритмів машинного навчання з досвідом людини. Однак важливо зазначити, що цей підхід вимагає значних даних і досвіду для ефективного впровадження.

Rawat, R. Shedbalkar, K. Moharir, M. Deepamala, N. Kumar, P.R. Tanmayananda використовували п'ять класів із масивного збору даних кібербезпеки CIRA-CIC-DoHBrw-2020 із сайту UNB як предмет дослідження: RNN, RFC, DTC, LSTM і GRU. Вони також використовували GBC, KNC і XGBoost. Точність, MAE, MSE, класифікаційні таблиці та матриці були додатковими показниками оцінки.

CIRA-CIC-DoHBrw-2020 передбачає дії з підготовки та попередньої обробки даних для налаштування даних для процесу навчання. У цьому дослідженні набір даних використовувався для оцінки запропонованої моделі ідентифікації зловмисного DNS через трафік HTTPS (DoH) за допомогою моделей навчання під наглядом. Набір даних CIRA-CIC-DoHBrw-2020 спочатку складався з двох наборів даних: (1) набір даних першого рівня, який використовується для класифікації трафіку DNS на DoH або не-DoH і складається з 269 643 зразків для трафіку DoH і 897 494 зразків для не-DoH Трафік DoH. (2) набори даних другого рівня, які використовуються для класифікації.

Безпечний трафік DoH або шкідливий і складається з 20 000 зразків для безпечного трафіку DoH і 249 836 зразків для зловмисного трафіку DoH. На рисунку 2 показано розподіл гістограми для набору даних CIRA-CIC-DoHBrw-2020.

Дослідження показують, що класифікатори XGBC і RFC працюють найкраще в цьому зборі даних. Хоча S. Alrayes, F. Maray, M. Gaddah, A. Yafoz, A. Alsini, R. Alghushairy, O. Mohsen, H. Motwakel, мали на меті класифікувати різні технології тунелювання DNS для запуску шкідливого трафіку DoH, недостатня різноманітність даних створює обмеження.

Behnke, M. Briner, N. Cullen, D. Schwerdtfeger, K. Warren, J. Basnet, R. Doleck, зменшили непотрібний шум, реалізували методи вибору функцій і запропонували зрозумілі функції, демонструючи більш точну та ефективну ідентифікацію шахрайського трафіку DoH. Ця робота спрямована на розвиток

попередніх досліджень і створення більш придатної моделі для практичного застосування шляхом усунення шуму, застосування методів вибору ознак і підкреслення можливості пояснення функцій. Результати показують, що LGBM показала найвищу точність часу навчання.

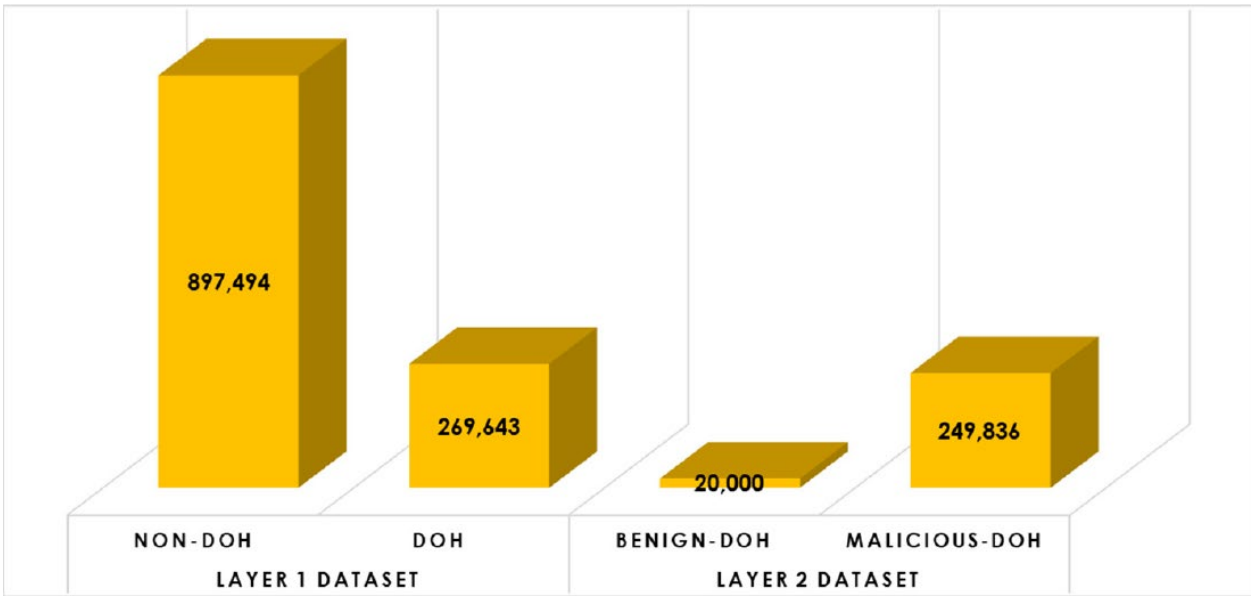

Рисунок 2 – Загальний набір даних CIRA-CIC-DoHBrw-2020

Джерело: [1]

Точна ідентифікація DoH є предметом [26] обговорення потенціалу аналізу зашифрованого трафіку. Мета полягає в тому, щоб оцінити, чи може машинне навчання витягувати будь-яку інформацію з розширених HTTPS-даних трафіку IP. Щоб визначити найкращі класифікатори DoH, можна дослідити п’ять широко використовуваних методів машинного навчання. Результати досліджень свідчать про те, що точність і розпізнавання DoH повинна становити понад 99,9%. Крім того, оскільки автори виявили (використовуючи створені набори даних) значні відмінності в поведінці Firefox, Chrome і Cloudflare, можна ідентифікувати програму, яка використовується для зв’язку DoH. З точністю 99,9% навчений класифікатор може ідентифікувати клієнтів. DoH. Deccio, C. Davis зосередилися на визначенні небезпеки використання протоколу DoH, окреслили підходи для виявлення трафіку DoH і запропонували підхід нейронної мережі для перегляду трафіку DoH. Вони використовували набір даних, отриманих з периферійних маршрутизаторів через IPFIX/NETFLOW, і досягнута точність передбачення становила 80% для ненормалізованих даних і більше 95% для очищених і нормалізованих даних. Casanova, L.F.G. Lin, P.-C. використали шість методів машинного навчання, щоб запропонувати систематику дворівневого методу ідентифікації трафіку DoH з відокремленням безпечного трафіку DoH від шкідливого. Ефективність запропонованого підходу оцінювали за точності, прецизійності, запам’ятовування, F-оцінки, матриць, криві ROC та значущощі ознак. Висновки показали, що алгоритми LGBM і XGBoost перевершують конкурентів. Практично за всіма параметрами класифікації, досягаючи

максимальної точності 100% у тестах класифікації першого та другого рівня. З 34 характеристик, взятих із набору даних CIRA-CIC-DoHBrw-2020, було виявлено, що вихідний I.P. є найважливішою характеристикою для відокремлення трафіку DoH від трафіку, що не є DoH, на першому рівні, потім IP-адреса призначення. На відміну від цього, лише IP-адреса призначення є ключовим компонентом для LGBM і алгоритмів посилення градієнта, щоб розрізняти безпечний і зловмисний трафік DoH на другому рівні.

З вищедосліджених ML-моделей можна спостерігати різну ефективність від досліджень, проте забезпечення контрольованого навчання та використання досвіду фахівця нададуть моделі вищої точності і кращих показників при дослідженні, що за умови підбору правильної моделі – в кінцевому виконанні принесе показник, наближений до 100%.

2.2 Аналізатори трафіку. Для визначення виду трафіку та дослідження його закономірностей – використовують аналізатори трафіку. Вони здатні відслідковувати, фіксувати та записувати події, на основі чого в подальшому можна здійснити аналіз результатів.

Найбільш доступним і адекватним способом отримання веб-трафіку, подібного до користувача, є використання браузера. Можна використати бібліотеку Selenium [32] для автоматизації веб-браузера з підтримкою DNS over HTTPS, Mozilla Firefox. Selenium – це популярна бібліотека, доступна кількома мовами програмування, яка дозволяє контролювати та автоматизувати веб-браузери. Він діє як інтерфейс між користувачем і бінарним виконуваним файлом, який називається WebDriver [32], приховуючи внутрішню роботу певної реалізації браузера через стандартизований інтерфейс (рис. 3).

Після готовності браузера до запуску, необхідно захопити трафік. Для цього використовується утиліта tcpdump(при роботі з Winows OS чи Mac – може бути використаний аналізатор трафіку Wireshark) [32] із фільтром захоплення, оскільки необхідно зосередитись на трафіку, спрямованому на порт 443. Варто зауважити, що, оскільки трафік можна фільтрувати за допомогою простого фільтра захоплення, навіть у реальному сценарії немає необхідності вводити надлишковий трафік. Вихідний файл має формат PCAP, який можна аналізувати іншими інструментами, наприклад Wireshark.

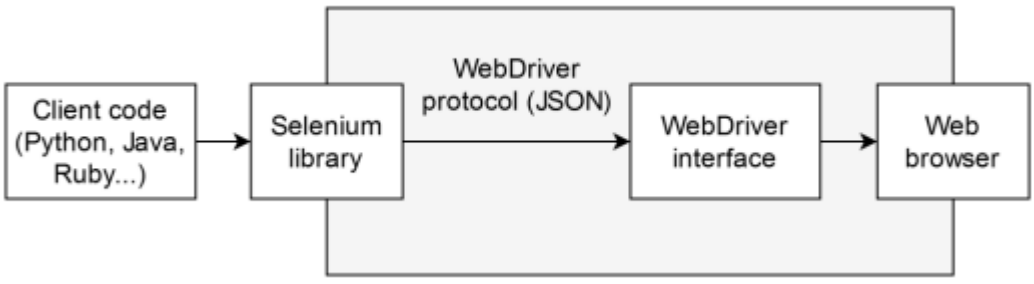

Рисунок 3 – Схема отримання веб-трафіку за допомогою бібліотеки Selenium

Джерело: [32]

Після обробки трасування мережі кожне підключення береться як зразок і позначається відповідно до IP-адреси призначення. Кожен зразок може мати

одну з двох міток: DNS over HTTPS (IP-адреса призначення резолвера Cloudflare) або HTTP (будь-яка інша адреса призначення).

Звідси можна зробити висновок, що комплексні інструменти для захоплення мережевого трафіку можуть надати більш вичерпний аналіз даних, необхідний для подальшого дослідження та його класифікації.

Висновки.

Відносно новий протокол DNS over HTTPS зарекомендував себе як такий, що забезпечує захист та конфіденційність даних в межах сесій DNS між клієнтом та сервером. Разом з цим – виникає проблема в замасковуванні шкідливого трафіку чи перенаправленні його з метою приховування істинних мотивів.

Інструменти, що аналізують та захоплюють трафік здатні виконати великий обсяг первинної роботи і здійснити розподіл трафіку на шкідливий і безпечний, а також передати необхідні результати, за якими можна створити графіки залежності.

За допомогою машинного навчання із підбором необхідної моделі та використанням класифікатора для ідентифікації типу трафіку, а також розробкою методу контрольованого навчання для блокування шкідливого трафіку – реалізований механізм, який ще не був якісно досліджений та стане в нагоді для майбутніх науковців, інженерів, фахівців з інформаційної безпеки та кінцевих користувачів. Аналіз точності механізму показує 99.4% для ідентифікації DoH та не DoH трафіку а також 100% для визначення шкідливого DoH трафіку.

Література.

1. Qasem Abu Al-Haija, Manar Alohaly,Ammar Odeh, “A Lightweight Double-Stage Scheme to Identify Malicious DNS over HTTPS Traffic Using a Hybrid Learning Approach”.

2. Hynek, K.; Vekshin, D.; Luxemburk J.; Cejka, T.; Wasicek, A.; “Summary of DNS over HTTPS Abuse”.

3. Jose, G.-L.; Mary, K.S.; Carol, A.W. Internet Protocol Handbook. In The Domain Name System (DNS) Handbook; DTIC: Fort Belvoir, VA, USA, 1989; Volume 4.

4. Paul, M. Domain Names–Implementation and Specification; Internet Engineering Task Force; ISI: Marina del Rey, CA, USA, 1987.

5. Usman Aijaz, N.; Misbahuddin, M.; Raziuddin, S. Survey on DNS-Specific Security Issues and Solution Approaches. In Data Science and Security; Jat, D.S., Shukla, S., Unal, A., Mishra, D.K., Eds.; Lecture Notes in Networks and Systems; Springer: Singapore, 2021; Volume 132, pp. 79–89, ISBN 9789811553080.

6. Romain, F. DNS Security for Business Continuity and Resilience; IDC: Needham, MA, USA, 2022.

7. Hu, Z.; Zhu, L.; Heidemann, J.; Mankin, A.;Wessels, D.; Hoffman, P.E. Specification for DNS over Transport Layer Security (TLS); Internet Engineering Task Force: Fremont, CA, USA, 2016.

8. Hoffman, P.E.; McManus, P. DNS Queries over HTTPS (DoH); Internet Engineering Task Force: Fremont, CA, USA, 2018.

9. Albulayhi, K.; Smadi, A.A.; Sheldon, F.T.; Abercrombie, R.K. IoT Intrusion Detection Taxonomy, Reference Architecture, and Analyses. Sensors 2021, 21, 6432. [CrossRef] [PubMed]

10. P. E. Hoffman and P. McManus, "DNS Queries over HTTPS (DoH)," RFC 8484, Tech. Rep. 8484, Oct. 2018.

11. P. Mockapetris, "Domain names –implementation and specification,"RFC 1035 (Internet Standard), RFC Editor, pp. 1–55. [Online]. Available: https://www:rfc-editor:org/rfc/rfc1035:txt

12. E. Brumaghin and C. Grady, "Covert channels and poor decisions: The tale of dnsmessenger," Mar 2017. [Online]. Available: https: /blog:talosintelligence:com/2017/03/dnsmessenger:html

13. C. Cimpanu, "Here's how to enable DoH in each browser, ISPs be damned," Dec 2020, https://www.zdnet.com/article/dns-over-https-willeventually-roll-out-in-all-major-browsers-despite-isp-opposition/.

14. S. García, K. Hynek, D. Vekshin, T. Cejka, and A. Wasicek, "Large scale measurement on the adoption of encrypted DNS," CoRR, vol. abs/2107.04436, 2021. [Online]. Available: https://arxiv:org/abs/2107:04436

15. K. Hynek and T. Cejka, "Privacy Illusion: Beware of Unpadded DoH," in 2020 11th IEEE Information Technology, Electronic and Mobile Communication conference (IEMCON), 2020.

16. K. Borgolte, T. Chattopadhyay, N. Feamster, M. Kshirsagar, J. Holland, A. Hounsel, and P. Schmitt, "How DNS over HTTPS is Reshaping Privacy, Performance, and Policy in the Internet Ecosystem," Performance, and Policy in the Internet Ecosystem (July 27, 2019), 2019.

17. I. N. Bozkurt, A. Aguirre, B. Chandrasekaran, P. B. Godfrey, G. Laughlin, B. Maggs, and A. Singla, "Why is the internet so slow?!" in Passive and Active Measurement, M. A. Kaafar, S. Uhlig, and J. Amann, Eds. Cham: Springer International Publishing, 2017, pp. 173–187.

18. P. McManus, Aug 2018. [Online]. Available: https://blog:nightly:mozilla:org/2018/08/28/firefox-nightly-securedns-experimental-results/

19. T. Böttger, F. Cuadrado, G. Antichi, E. L. a. Fernandes, G. Tyson, I. Castro,and S. Uhlig, "An Empirical Study of the Cost of DNS-over-HTTPS,"in Proceedings of the Internet Measurement Conference, ser. IMC '19. New York, NY, USA: Association for Computing Machinery, 2019, p.15–21. [Online]. Available: https://doi:org/10:1145/3355369:3355575

20. A. Hounsel, P. Schmitt, K. Borgolte, and N. Feamster, "Can EncryptedDNS Be Fast?" in Passive and Active Measurement, O. Hohlfeld, A. Lutu, and D. Levin, Eds. Cham: Springer International Publishing, 2021, pp. 444–459.

21. K. Jerabek, O. Rysavy, and I. Burgetova, "Measurement and characterization of DNS over HTTPS traffic," 2022. [Online]. Available: https://arxiv:org/abs/2204:03975

22. R. Chhabra, P. Murley, D. Kumar, M. Bailey, and G. Wang, "Measuring DNS-over-HTTPS Performance around the World," in Proceedings of the 21st ACM Internet Measurement Conference, ser. IMC '21. New York, NY, USA: Association

for Computing Machinery, 2021, p. 351–365. [Online]. Available: https://doi:org/10:1145/3487552:3487849

23. A. Hounsel, K. Borgolte, P. Schmitt, J. Holland, and N. Feamster, Comparing the Effects of DNS, DoT, and DoH on Web Performance. New York, NY, USA: Association for Computing Machinery, 2020, p. 562–572. [Online]. Available: https://doi:org/10:1145/3366423:3380139

24. E. S. Mbewe and J. Chavula, "On QoE Impact of DoH and DoT in Africa: Why a User's DNS Choice Matters," in Towards new e-Infrastructure and e-Services for Developing Countries, R. Zitouni, A. Phokeer, J. Chavula, A. Elmokashfi, A. Gueye, and N. Benamar, Eds. Cham: Springer International Publishing, 2021, pp. 289–304.

25. T. Jensen, "Windows Insiders can now test DNS over HTTPS," May 2020. [Online]. Available: https://techcommunity:microsoft:com/t5/networkingblog/windows-insiders-can-now-test-dns-over-https/ba-p/1381282

26. Vekshin, D.; Hynek, K.; Cejka, T. Doh insight: Detecting dns over https by machine learning. In Proceedings of the 15th International Conference on Availability, Reliability and Security, Virtual Event, 25–28 August 2020; pp. 1–8.

27. Rawat, R.; Shedbalkar, K.; Moharir, M.; Deepamala, N.; Kumar, P.R.; Tanmayananda, M. Analysis and detection of malicious activity on doh traffic. In Proceedings of the 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 1–3 October 2021; pp. 1–5.

28. Parra, G.D.L.T.; Rad, P.; Choo, K.-K.R. Implementation of deep packet inspection in smart grids and industrial Internet of Things: Challenges and opportunities. J. Netw. Comput. Appl. 2019, 135, 32–46. [CrossRef]

29. Naz, N.; Khan, M.A.; Alsuhibany, S.A.; Diyan, M.; Tan, Z.; Khan, M.A.; Ahmad, J. Ensemble learning-based IDS for sensors telemetry data in IoT networks. Math. Biosci. Eng. 2022, 19, 10550–10580. [CrossRef]

30. Fisher,W.W.; Piazza, C.C.; Roane, H.S. Handbook of Applied Behavior Analysis; Guilford Publications: New York, NY, USA, 2021.

31. Behnke, M.; Briner, N.; Cullen, D.; Schwerdtfeger, K.; Warren, J.; Basnet, R.; Doleck, T. Feature Engineering and Machine Learning Model Comparison for Malicious Activity Detection in the DNS-Over-HTTPS Protocol. IEEE Access 2021, 9, 129902–129916. [CrossRef]

32. DNS Over HTTPS Traffic Analysis and Detection. Carlos López Romera, Carlos Hernández Gañán,Víctor García Font 2nd June, 2020.

Abstract: *Domain Name System has a great role in accessing Internet resources, providing granularity, hierarchy and consistency.*

This protocol is an integral part of the information space, which cannot be replaced, but due to its peculiarity, which was developed back in the 1980s, it contains a number of vulnerabilities that are easily used by criminals.

The introduction of the new DNS over HTTPS protocol eliminates the threats inherent in traditional DNS, ensuring data encryption and privacy within client-server connections, but creates an additional traffic load that can be felt with limited 3G/LTE network connectivity.

The development of significantly new models for the study and analysis of DoH traffic with the provision of high performance indicators can provide network engineers and specialists in the field of computer network protection with qualitative knowledge necessary for management, monitoring

and protection of the information space.

Tools that analyze and capture traffic are able to perform a large amount of initial work and perform the classification of traffic into malicious and safe, as well as transmit the necessary results that can be used to create dependency graphs. These include Tcpdump, Wireshark using the necessary filters (by port or by protocol) to capture DoH traffic, the Selenium library for automating the collecting of logs in a web browser controlled by one of the Python, Java or Ruby programming languages.

Using machine learning with model matching, the first step uses a random spanning tree (RF) method to identify traffic as DoH or non-DoH, the second step analyzes malicious and benign DoH traffic, giving us an accuracy score of 99.4% and 100% respectively . The use of the classifier itself to identify the type of traffic, as well as the development of a supervised learning method, is the basis for configuring the blocking of malicious traffic. The implementation of this mechanism has not yet been qualitatively researched and it will be useful for future scientists, engineers, information security specialists and end users.

Key words: *Traffic classifier, Domain Name System, DNS over HTTPS, ADT, Top Level Domain, RF.*

Науковий керівник: к.т.н., доц. Коробейнікова Т.І.

Стаття надіслана: 14.09.2023 р.