Хмельницький національний університет
Факультет інформаційних технологій
Кафедра кібербезпеки

КВАЛІФІКАЦІЙНА РОБОТА

Сікорського Павла Олександровича

на здобуття ступеня вищої освіти Магістра

Метод виявлення аномалій у DNS-запитах

Галузь знань 12 – Інформаційні технології
Спеціальність 125 – Кібербезпека та захист інформації
Освітня програма Кібербезпека та захист інформації

Шифр КРМКБЗІ.2301151.23.01.19 ПЗ

Виконав студент 2 курсу група КБЗІм-23-1 Павло СІКОРСЬКИЙ

Керівник канд. техн. наук, доцент Юрій КЛЬОЦ

Нормоконтролер старший викладач Сергій МОСТОВИЙ

До захисту допускаю:
Завідувач кафедри кібербезпеки Юрій КЛЬОЦ

19 12 2024 р.

Хмельницький 2024

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет Інформаційних технологій

Кафедра Кібербезпеки

Рівень вищої освіти Магістр

Галузь знань 12 – Інформаційні технології

Спеціальність 125 – Кібербезпека та захист інформації

Освітня програма Кібербезпека та захист інформації

ЗАТВЕРДЖУЮ

Завідувач кафедри кібербезпеки

Юрій КЛЬОЦ

2 09 2024 р.

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Сікорському Павлу Олександровичу

1 Тема роботи Метод виявлення аномалій у DNS-запитах

Керівник роботи канд.техн.наук, доцент Юрій КЛЬОЦ

Затверджено наказом ректора університету від ___ _____ 2024 № ___

2 Строк подання студентом кваліфікаційної роботи на кафедру

3 Вихідні дані до роботи Визначити типи і формати DNS-запитів. Проаналізувати відомі аномалії в DNS-запитах та методи їх виявлення. Обрати параметри DNS-запиту, що доцільно використовувати для виявлення аномалій. Вибрати моделі машинного навчання, що застосовуються для виявлення аномалій в DNS-запитах. Розробити метод,що покращує достовірність виявлення аномалій. Здійснити оцінку достовірності розробленого методу.

4 Зміст пояснювальної записки (перелік питань, які потрібно розробити)
Вступ. Аналіз DNS-запитів. Типи та формати DNS-запитів. Аномалії DNS-запитів. Моделі виявлення аномалій. Моделі машинного навчання. Метод виявлення аномалій в DNS-запитах. Система виявлення аномалій в DNS-запитах. Оцінка достовірності методу та системи виявлення аномалій в DNS-запитах.

5 Перелік графічного матеріалу (із зазначенням обов'язкових креслень)

6 Консультанти розділів кваліфікаційної роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7 Дата видачі завдання ____ _____ 2024 р.

КАЛЕНДАРНИЙ ПЛАН

Назва етапів (розділів) кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
Ґрунтовне ознайомлення та дослідження предметної галузі	Лютий	Виконано
Визначення змісту, структури магістерської роботи	Березень	Виконано
Опрацювання першого розділу магістерської роботи	Квітень	Виконано
Опрацювання статті за результатами дослідження	Травень	Виконано
Опрацювання другого розділу магістерської роботи	Червень	Виконано
Опрацювання третього розділу магістерської роботи	Вересень	Виконано
Опрацювання четвертого розділу магістерської роботи	Жовтень	Виконано
Підготовка та опрацювання ілюстративного матеріалу	Листопад	Виконано
Оформлення магістерської роботи графічної та текстової частини	Листопад	Виконано
Попередній захист магістерської роботи	Листопад	Виконано
Захист магістерської роботи на засіданні ЕК	Грудень	Виконано

Студент _____ Павло СІКОРСЬКИЙ

Керівник кваліфікаційної роботи _____ Юрій КЛЬОЦ

АНОТАЦІЯ

Тема кваліфікаційної роботи: Система виявлення аномалій у DNS-запитах

Автор роботи: студент групи КБЗІм-23-1 Сікорський П.О.

Керівник роботи: канд. техн. наук, доцент Кльоц Ю.П.

Загальний обсяг роботи: 78 сторінок, 15 рисунків, 8 таблиць, 1 додаток, 70 посилань.

Ключові слова: DNS-запити, виявлення аномалій, Isolation Forest, One-Class SVM, K-means.

Дана кваліфікаційна робота присвячена розробці системи виявлення аномалій у DNS-запитах із використанням сучасних методів машинного навчання для підвищення ефективності мережевих систем безпеки. Проведено аналіз основних підходів до виявлення аномалій у DNS-трафіку, зокрема статистичних та алгоритмічних. Розроблена система базується на використанні моделей Isolation Forest, One-Class SVM та K-means, що забезпечують виявлення відхилень від профілю нормальної активності та дозволяють точно ідентифікувати потенційні загрози у DNS-запитах. Система реалізована у вигляді модульної архітектури, яка забезпечує високу гнучкість, масштабованість та можливість інтеграції у сучасні мережеві інфраструктури.

Проведено тестування запропонованої системи на наборі даних CAIDA Passive DNS Dataset, що включає реальні приклади DNS-трафіку. Результати експериментів показали, що система демонструє високий рівень точності виявлення аномалій, досягаючи понад 90% успішності у виявленні загроз, з мінімальною кількістю хибнопозитивних спрацювань. Запропонована система має високу практичну значущість для забезпечення мережевої безпеки та зниження ризиків кібератак у сучасних мережах.

09. 12. 2024

ANNOTATION

Theme of qualification work: Anomaly Detection System in DNS Queries

Author of the work: KBZIm-23-1 Sikorskyi P.O.

Mentor: Associate Professor Klots Y.P.

Total volume of work: 78 pages, 15 figures, 8 tables, 1 appendix, 70 references.

Keywords: DNS queries, anomaly detection, Isolation Forest, One-Class SVM, K-means.

This qualification work is devoted to the development of an anomaly detection system for DNS queries using modern machine learning methods to enhance the effectiveness of network security systems. The study includes an analysis of the main approaches to anomaly detection in DNS traffic, specifically statistical and algorithmic methods. The proposed system is based on the application of Isolation Forest, One-Class SVM, and K-means models, which enable the identification of deviations from the profile of normal activity and ensure precise detection of potential threats in DNS queries. The system is implemented in a modular architecture that provides high flexibility, scalability, and the ability to integrate into modern network infrastructures.

The proposed system was tested on the CAIDA Passive DNS Dataset, which includes real-world DNS traffic examples. The experimental results demonstrate that the system achieves a high level of anomaly detection accuracy, exceeding 90% in threat identification, with a minimal number of false-positive detections. The developed system has significant practical value for ensuring network security and reducing the risks of cyberattacks in modern networks.

09.12.2024

ЗМІСТ

ВСТУП

Забезпечення безпеки мережевої інфраструктури залишається одним із головних викликів у сучасному цифровому середовищі, враховуючи постійне зростання обсягів даних та складність методів, які використовують зловмисники. Однією з ключових загроз у цьому контексті є зловживання системою доменних імен (DNS), яка відіграє критичну роль у функціонуванні інтернету. DNS-запити забезпечують трансляцію доменних імен у IP-адреси, що є фундаментальним процесом для доступу до вебресурсів, обміну даними та інших комунікацій. Утім, ця система часто стає об’єктом атак або використовується як канал для здійснення кіберзагроз, таких як DDoS-атаки, витоки даних, DNS-тунелювання та інші види шкідливої активності.

Однією з основних проблем є те, що багато загроз, пов’язаних із DNS, можуть бути замасковані у вигляді нормальної активності, що значно ускладнює їх виявлення за допомогою традиційних методів моніторингу мережі. Зокрема, аномалії у DNS-запитах можуть проявлятися у вигляді підозрілих частот звернень до певних доменів, використання алгоритмічно згенерованих доменів (DGA), аномальних часових патернів або структурних змін у запитах. Усе це потребує більш точних і адаптивних підходів до аналізу трафіку.

Існуючі методи виявлення аномалій у DNS-трафіку, такі як статистичні моделі або прості евристичні правила, здебільшого обмежені у своїй здатності ефективно працювати з динамічними та складними загрозами. Водночас методи машинного навчання та аналізу великих даних відкривають нові можливості для створення інтелектуальних систем виявлення, здатних адаптуватися до нових типів загроз. Сучасні підходи, які використовують алгоритми кластеризації, класифікації та побудови моделей, дозволяють аналізувати великий обсяг даних і знаходити навіть приховані аномалії, які раніше залишалися непоміченими.

Метою цієї роботи є розробка методу виявлення аномалій у DNS-запитах, який інтегрує сучасні підходи аналізу даних та машинного навчання. Запропонований метод покликаний забезпечити точне та своєчасне виявлення

аномалій, які можуть свідчити про загрози для мережевої інфраструктури. Актуальність цього дослідження зумовлена постійним зростанням кількості кіберзагроз, які використовують DNS як канал для зловмисних дій, а також необхідністю створення адаптивних систем безпеки, здатних ефективно реагувати на динамічні зміни у поведінці мережевого трафіку.

Об'єктом дослідження в цій роботі є DNS-трафік, його структура та характеристики, а також аномалії, які можуть виникати у процесі функціонування системи доменних імен. Предметом дослідження є методи та алгоритми аналізу DNS-запитів, які забезпечують виявлення аномалій у трафіку. Зокрема, дослідження зосереджене на застосуванні методів машинного навчання, таких як кластеризація та класифікація, для ідентифікації підозрілих запитів, що відхиляються від нормальної поведінки.

Наукова новизна роботи полягає у розробці інтегрованого підходу, який поєднує статистичний аналіз і сучасні алгоритми машинного навчання для підвищення точності та ефективності виявлення аномалій у DNS-запитах. Запропонований метод забезпечує не лише визначення відомих типів загроз, але й виявлення нових, раніше невідомих аномалій, що виникають у результаті змін у тактиках зловмисників. Використання машинного навчання дозволяє моделі адаптуватися до змін у поведінці трафіку, забезпечуючи гнучкість та надійність системи.

Практичне значення роботи полягає у створенні системи, яка може бути впроваджена у реальні мережеві інфраструктури для моніторингу та аналізу DNS-трафіку. Розроблений підхід дозволить знизити ризики витоку даних, зменшити кількість помилкових спрацювань у системах безпеки та підвищити загальний рівень захисту мережі. Окрім того, результати дослідження можуть бути використані для розробки рекомендацій щодо покращення методів кібербезпеки у контексті DNS-трафіку, а також для створення нових інструментів моніторингу та аналізу, що відповідають вимогам сучасних інформаційних систем.

1. DNS-ЗАПИТИ

1.1.Система доменних імен

Система доменних імен (DNS) є одним із ключових компонентів інфраструктури Інтернету, що забезпечує трансляцію доменних імен у IP-адреси. Вона виконує роль інтерфейсу між користувачами та машинами, дозволяючи людям взаємодіяти з Інтернетом за допомогою зручних імен, а не складних числових адрес. DNS працює на основі ієрархічної системи серверів, кожен із яких виконує свою роль у забезпеченні масштабованості та надійності цього критично важливого сервісу. Коли користувач вводить у браузері доменне ім'я, таке як www.example.com, DNS відповідає за знаходження відповідної IP-адреси сервера, на якому розміщений запитуваний ресурс. Цей процес включає кілька етапів, що виконуються із залученням локальних і віддалених DNS-серверів, а також кешуючих механізмів для прискорення виконання запитів.

DNS є децентралізованою системою, що складається з кореневих серверів, серверів верхнього рівня доменів (TLD), авторитетних серверів і рекурсивних резолверів (Рис. 1.1). Кореневі сервери, розташовані по всьому світу, забезпечують початкову точку входу для пошуку будь-якого домену, тоді як TLD-сервери зберігають інформацію про домени верхнього рівня, такі як .com, .org або національні домени, наприклад, .ua. Авторитетні сервери відповідають за надання остаточних даних про конкретний домен, включаючи його IP-адресу, параметри поштових серверів і налаштування інших служб. Локальні резолвери, зазвичай інтегровані в мережеві пристрої, такі як маршрутизатори або сервери провайдерів, кешують результати попередніх запитів, зменшуючи навантаження на інфраструктуру та прискорюючи роботу користувачів. Алгоритм кешування DNS-запитів представлений на рис. 1.2.

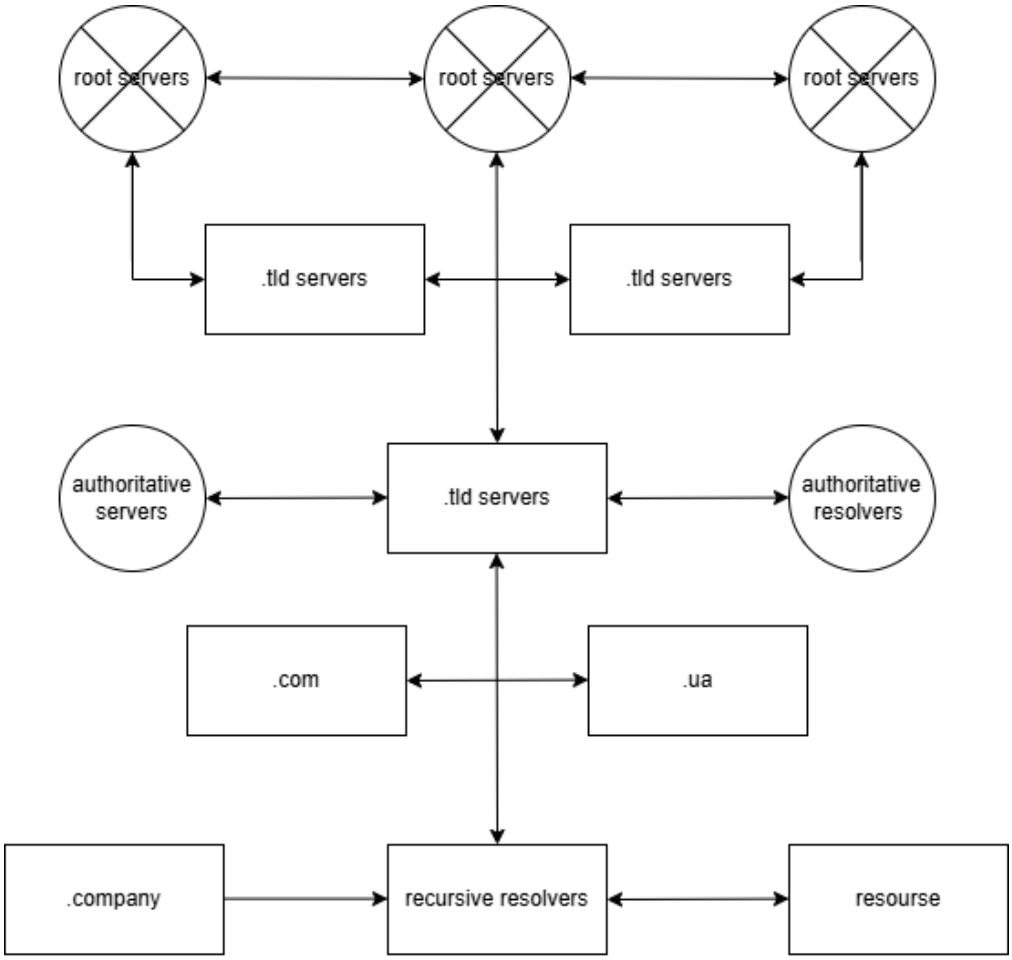

Рисунок 1.1 Схематичне зображення ієрархії DNS

Ефективність DNS має критичне значення для функціонування Інтернету. Запити до DNS відбуваються щоразу, коли користувач завантажує вебсторінку, надсилає електронний лист або підключається до сервісу через API. Навіть невеликі затримки в роботі DNS можуть суттєво вплинути на продуктивність і час завантаження вебресурсів. У той же час, завдяки ієрархічній структурі та використанню кешування, DNS здатна обробляти мільярди запитів щодня, забезпечуючи швидкий і надійний доступ до ресурсів у будь-якій точці світу. Важливим аспектом є також масштабованість системи, яка дозволяє додавати нові домени та обслуговувати зростаючий обсяг трафіку.

У той же час DNS є потенційною точкою вразливості для мережі Інтернет. Зловмисники часто використовують її для атак, таких як DNS-спуфінг, DNS-ампліфікація чи атаки на відмову в обслуговуванні (DDoS). Для захисту від таких загроз впроваджуються розширення DNS, такі як DNSSEC, які забезпечують

криптографічний захист даних. Важливу роль у безпеці також відіграють кешуючі резолвери, які мінімізують кількість запитів до зовнішніх серверів і тим самим знижують ризики атак.

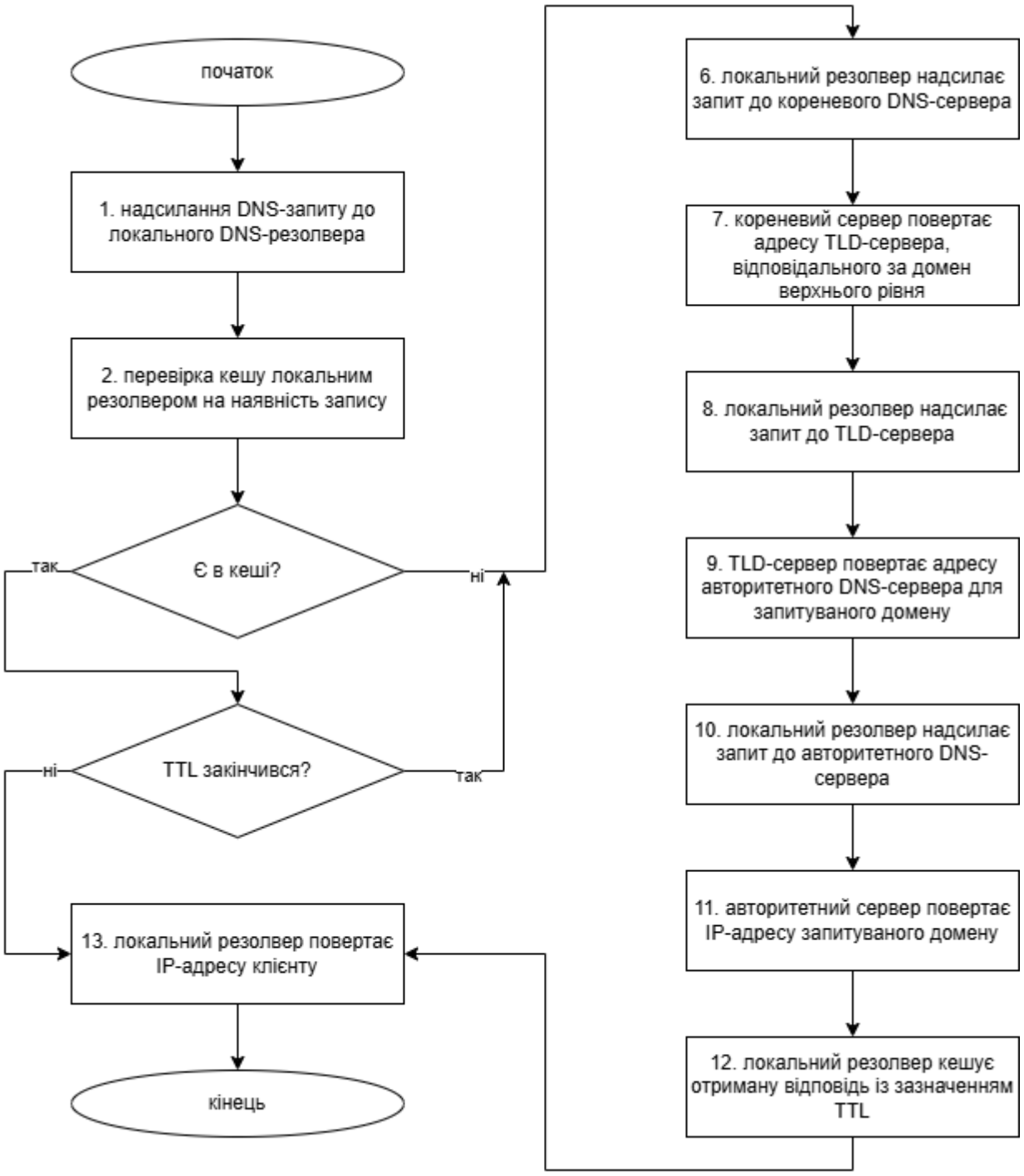

Рисунок 1.2 Алгоритм кешування DNS

1.2.Основні загрози, пов’язані з аномальними DNS-запитами.

Аномальний DNS-запит – це запит до системи доменних імен (DNS), який відхиляється від нормальної поведінки або очікуваних шаблонів взаємодії. Такі запити можуть мати нетипові характеристики, включаючи незвичну частоту, структуру, обсяг або джерело. Аномальними можуть вважатися запити до зловмисних або невідомих доменів, надмірно довгі імена доменів, запити з підробленими IP-адресами, або ті, що супроводжуються незвичними параметрами, наприклад, нехарактерними значеннями TTL. Аномальні DNS-запити часто пов’язані з кібератаками (наприклад, DDoS або DNS-ампліфікація), витоками даних, шкідливим програмним забезпеченням або спробами обійти мережеві обмеження через DNS-тунелювання.

Аномальні DNS-запити становлять серйозну загрозу для безпеки та стабільності мережі, оскільки вони можуть бути індикатором кібератак або зловмисної активності. Однією з найбільш поширених загроз є використання DNS для здійснення DDoS-атак, зокрема через механізм DNS-ампліфікації (Рис. 1.3). У такому випадку зловмисник надсилає великі обсяги запитів до відкритих DNS-рекурсорів, підробляючи IP-адресу джерела запиту. У відповідь сервери надсилають значно більші за обсягом відповіді на вказану IP-адресу жертви, перевантажуючи її мережу. Оскільки DNS-запити є критично важливою частиною функціонування Інтернету, такі атаки можуть спричинити значні перебої в роботі сервісів, викликати відмову в обслуговуванні й уповільнення мережевого трафіку на рівні інфраструктури.

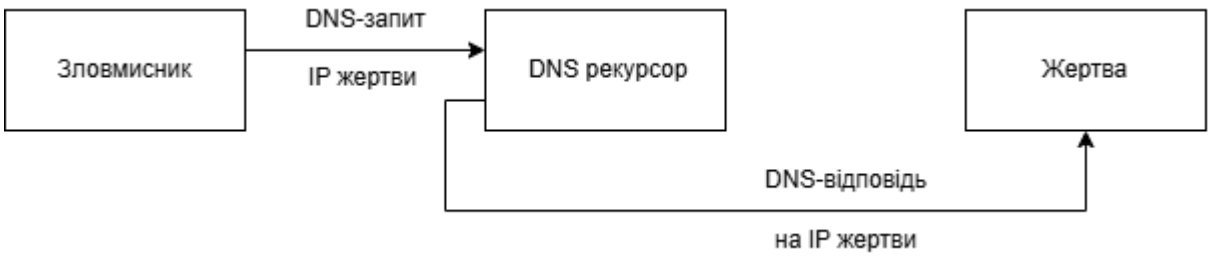

Рисунок 1.3 механізм DNS-ампліфікації в DDoS-атаках

Ще однією серйозною загрозою є використання DNS для поширення зловмисного програмного забезпечення через домени, які контролюються кіберзлочинцями. Ці домени зазвичай створюються автоматизованими інструментами, такими як генератори доменних імен (DGA), які створюють тисячі унікальних доменів щодня. Зловмисне програмне забезпечення може використовувати такі домени для отримання команд від C2-серверів (Command and Control), завантаження додаткових модулів або передачі викрадених даних. Зважаючи на те, що домени зловмисного ПЗ часто мають короткий час життя та уникають статичних списків блокувань, їх виявлення в реальному часі є складним завданням. Такі домени можуть виглядати схожими на звичайні, але відрізняються високою частотою використання нестандартних символів, довжиною та географічною нерівномірністю запитів.

DNS також використовується як канал для прихованого витоку даних із корпоративних мереж. Цей метод базується на вбудовуванні конфіденційної інформації в текстові записи DNS-запитів або відповідей. Наприклад, зловмисники можуть кодувати дані у форматі запитів до піддоменів, які направляються до сервера, контрольованого нападником. Такий підхід дозволяє обійти традиційні системи безпеки, які не аналізують зміст DNS-запитів (Рис. 1.4). Витік даних через DNS особливо небезпечний у випадках, коли системи моніторингу орієнтовані лише на фільтрацію трафіку HTTP або FTP, залишаючи DNS без належного контролю. Крім того, методи шифрування DNS, такі як DNS-over-HTTPS (DoH), хоч і підвищують конфіденційність, водночас ускладнюють виявлення подібних загроз.

Ще одним типом загрози є DNS-тунелювання, яке використовують для обходу мережевих обмежень або проникнення в ізольовані мережі. DNS-тунелі дозволяють створювати альтернативні канали зв’язку через стандартні DNS-запити та відповіді. Наприклад, зловмисник може передавати зашифровані пакети даних через DNS-запити до контрольованого сервера, приховуючи реальний вміст під виглядом звичайного трафіку. Цей метод є популярним для передачі даних поза

межі контрольованих зон, оскільки DNS-трафік рідко блокується повністю через його критичну роль у роботі мережі.

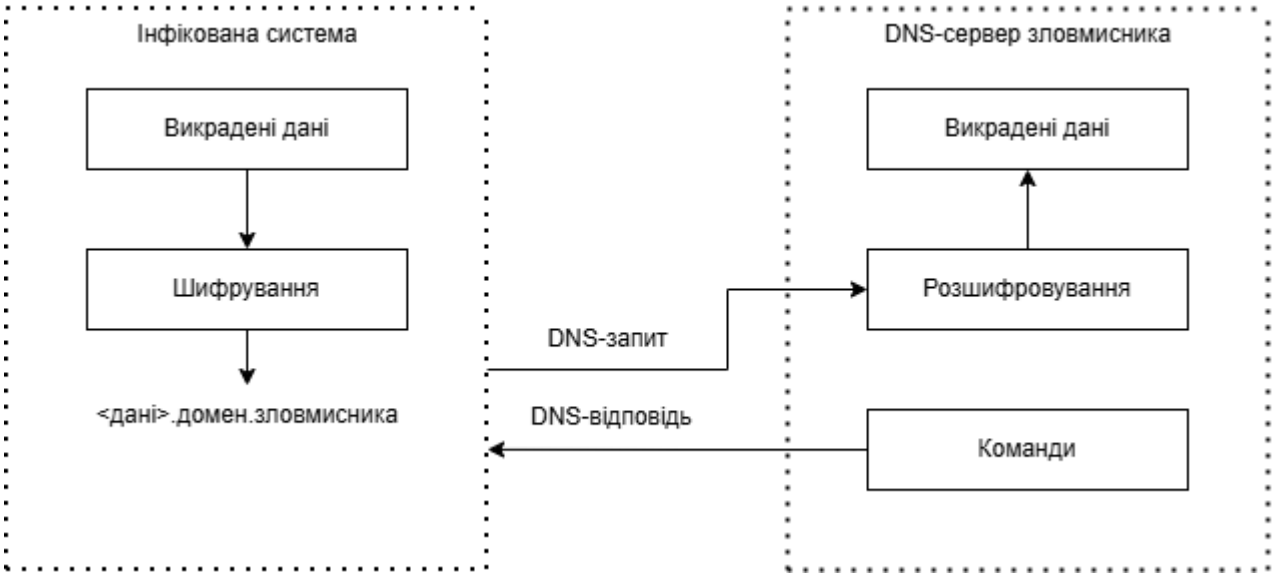

Рисунок 1.4 Витік даних через DNS-запити

Аномальні DNS-запити також можуть використовуватися для підготовки до фішингових кампаній. Зловмисники створюють домени, схожі на легітимні, використовуючи техніки типу Typosquatting, які експлуатують друкарські помилки користувачів. Такі домени часто використовуються для перенаправлення трафіку на підроблені сайти, які збирають облікові дані, паролі або фінансову інформацію. Оскільки DNS виконує трансляцію цих доменів, системи безпеки мають аналізувати їхню відповідність до шаблонів звичайних доменів, щоб запобігти обману користувачів.

Таким чином, DNS є не лише основою функціонування Інтернету, але й інструментом, який зловмисники активно використовують для атак і прихованих дій. Аномальні DNS-запити можуть сигналізувати про широке коло загроз, і їх виявлення вимагає інтеграції сучасних технологій аналізу трафіку та поведінкових патернів.

1.3. Аналіз DNS-запитів

Аналіз DNS-запитів включає кілька ключових аспектів, які охоплюють джерела даних, типи зібраної інформації, способи їх отримання та специфіку використання цих даних у подальшому аналізі. Зібрані дані формують основу для побудови профілю нормальної поведінки, виявлення аномалій і аналізу мережевого трафіку.

Основним джерелом для збору даних є логи DNS-серверів. Для цього використовуються популярні DNS-сервери, такі як BIND, Unbound, PowerDNS та інші, які генерують журнали запитів у текстовому або структурованому форматі. Ці журнали містять інформацію про всі отримані, оброблені та переадресовані DNS-запити. Дані логів зазвичай включають IP-адресу клієнта, тип DNS-запиту (A, AAAA, MX, CNAME тощо), доменне ім'я, запитуване клієнтом (FQDN), час отримання запиту та тип відповіді сервера (успішна, помилка, відсутність даних тощо). Для підвищення ефективності аналізу рекомендується використовувати структуровані журнали в форматах JSON або CSV, що спрощує подальший аналіз.

Додатковим джерелом є дані мережевого моніторингу, отримані за допомогою інструментів аналізу трафіку, таких як Wireshark, Zeek (раніше відомий як Bro), Tcpdump або NetFlow. Ці інструменти дозволяють знімати та аналізувати пакети, які містять DNS-запити та відповіді. У таких даних зазвичай зберігаються не лише стандартні мета-дані DNS-запитів, а й інформація про рівень трафіку, час затримки між запитами та відповіді, використання різних протоколів (TCP або UDP). Це дає змогу не лише аналізувати окремі запити, але й виявляти більш складні патерни, наприклад, послідовні запити до одного домену або поведінку ботнетів.

Ще одним важливим джерелом є глобальні публічні списки доменів, наприклад Alexa Top 1M, Majestic Million, Cisco Umbrella Popularity List, які містять інформацію про популярні домени та їх ранжування за рівнем використання. Ці дані дозволяють ідентифікувати рідкісні або незвичайні домени в логах DNS-серверів, які можуть бути пов’язані із зловмисною активністю. Публічні списки

зловмисних доменів, такі як Threat Intelligence Platforms (AbuseIPDB, Open Threat Exchange, VirusTotal), використовуються для перевірки, чи не належать запитувані домени до категорії потенційно небезпечних.

Геолокаційні бази даних, наприклад MaxMind GeoIP або IP2Location, використовуються для зіставлення IP-адрес клієнтів із їх географічними координатами. Це дозволяє ідентифікувати регіони, з яких надходить нетиповий трафік, або відслідковувати підозрілі запити з незвичних місць. Наприклад, якщо сервер отримує велику кількість DNS-запитів із країни, що зазвичай не має високого рівня трафіку до цієї мережі, це може бути ознакою ботнет-атаки.

Логи веб-серверів, такі як Apache або Nginx, можуть бути додатковим джерелом даних, якщо DNS-запити пов'язані з аналізом веб-трафіку. Веб-сервери часто записують інформацію про запити до доменів, які вони обслуговують, включаючи IP-адреси клієнтів, час запиту та тип ресурсу, що запитується. Зіставлення цих даних із DNS-запитами допомагає виявляти випадки, коли певний домен використовується як частина фішингових або шкідливих кампаній.

Інструменти потокового аналізу, такі як Splunk або Elasticsearch, також можуть слугувати джерелом даних. Вони дозволяють зберігати великі обсяги DNS-логів у реальному часі та виконувати пошук за складними запитами. Це особливо корисно для виявлення закономірностей у трафіку та агрегації даних із різних джерел.

Інформація з внутрішніх систем аутентифікації, таких як Active Directory або Kerberos, може доповнити аналіз, якщо необхідно враховувати контекст користувачів, які генерують DNS-запити. Наприклад, перевірка, чи запит належить до легітимного користувача або невідомого пристрою в мережі, дозволяє виявляти можливі компрометації.

Зібрані дані мають бути структуровані та збережені в спеціалізованих системах для подальшої обробки. Використання розподілених баз даних, таких як Hadoop або Apache Kafka, забезпечує можливість роботи з великими обсягами даних у реальному часі. Це дозволяє виконувати аналіз навіть для високонавантажених DNS-серверів.

Джерела даних мають бути надійними та оновлюватися у режимі реального часу. Наприклад, інтеграція з Threat Intelligence Platforms за допомогою API дозволяє автоматично отримувати актуальні списки зловмисних доменів, що значно підвищує ефективність методу. Збереження історичних даних у системах типу Amazon S3 або Google BigQuery дозволяє виконувати ретроспективний аналіз, виявляючи довгострокові тренди та нові загрози.

1.4. Типи даних DNS-запитів

Типи даних, які використовуються для збору інформації на етапі аналізу DNS-запитів, можна поділити на кілька ключових категорій, кожна з яких містить низку параметрів, що описують запити, відповіді, клієнтів і контекст їхньої взаємодії з DNS-серверами.

Дані про DNS-запити охоплюють інформацію про запитуване доменне ім’я (FQDN). Це повне доменне ім’я, яке містить основний домен та всі піддомени, що йому передують. До цієї категорії відносяться такі характеристики, як довжина доменного імені, кількість рівнів у його ієрархії, наявність незвичайних символів, а також відповідність загальноприйнятим форматам доменів. Наприклад, дуже довгі доменні імена або ті, що містять випадкові послідовності символів, часто є ознакою використання генераторів доменів (DGA). Окремо враховується тип запиту, який визначає, яку саме інформацію клієнт хоче отримати від DNS-сервера. Найпоширенішими є запити на отримання IP-адреси для домену (тип A або AAAA), але також зустрічаються запити для отримання інформації про поштові сервери (MX), піддоменні записи (CNAME) чи інші ресурси.

Важливим типом даних є IP-адреса клієнта, який надіслав запит. Вона дозволяє ідентифікувати джерело запиту, а також зіставляти його з географічним регіоном або підмережею, використовуючи інструменти геолокації. Додатково аналізуються мета-дані про клієнтське обладнання, такі як використання протоколу (IPv4 чи IPv6), порт джерела, через який був здійснений запит, і параметри транспортування (TCP чи UDP). Ці параметри допомагають відстежувати нетипову

поведінку, наприклад, масову генерацію запитів із конкретної IP-адреси або використання нестандартних портів.

Дані про час і частоту запитів становлять ще одну критично важливу категорію. Для кожного запиту реєструється точний час його надсилання, що дозволяє виконувати часовий аналіз. Наприклад, пікові періоди активності або високочастотні запити до одного домену можуть вказувати на автоматизовану активність або DDoS-атаки. Частотний аналіз також дає змогу виявляти розподілені атаки, коли запити надходять із багатьох IP-адрес, але мають спільний часовий патерн.

Тип відповіді сервера є важливим параметром для оцінки результату DNS-запиту. Сервер може повернути успішну відповідь із резолюцією домену до конкретної IP-адреси або повідомити про помилку, наприклад, через відсутність даних у зоні обслуговування (NXDOMAIN). Аналіз відповідей дозволяє виявляти патерни, наприклад, систематичні запити до неіснуючих доменів, які можуть бути ознакою сканування або спроб підбору доменів.

Геолокаційна інформація, отримана шляхом аналізу IP-адрес, дозволяє додатково контекстуалізувати запити. Вона включає країну, регіон, місто та навіть провайдера, що обслуговує клієнта. Географічна прив’язка є важливою для виявлення нетипової поведінки, наприклад, коли сервер отримує запити з регіонів, які зазвичай не генерують трафік для даної мережі. Аномалії такого типу часто асоціюються з ботнетами, що використовують проксі для розподілу запитів.

Окрім самих запитів, збирається інформація про відповідні домени. Наприклад, дані про те, чи належить домен до популярних (на основі списків, таких як Alexa Top 1M), або чи входить до категорії зловмисних, згідно з Threat Intelligence платформами. Додатково аналізується час існування домену, який можна визначити за допомогою WHOIS-запитів. Новостворені або нещодавно оновлені домени частіше використовуються для шкідливих цілей, таких як фішинг чи створення C2-інфраструктури.

Дані про трафік між клієнтом і сервером включають розмір запиту та відповіді, кількість переспрямувань (реферальних запитів) і затримку між

надсиланням запиту та отриманням відповіді. Ці параметри дозволяють виявляти відхилення у продуктивності або затримки, які можуть бути спричинені перевантаженням серверів або зловмисною активністю.

Важливими також є агреговані показники, отримані внаслідок обробки первинних даних. Наприклад, частотні профілі для кожного клієнта, що відображають, як часто певна IP-адреса звертається до одного й того ж домену. Аналіз цих профілів допомагає виявляти автоматизовані запити, характерні для ботів або сканерів.

1.5. Формат DNS-запитів

DNS-запит є структурованим повідомленням, що надсилається клієнтом до DNS-сервера для отримання інформації про доменне ім’я. Запит має чітко визначену структуру, що складається з кількох секцій, які формують його заголовок і основний зміст. Заголовок DNS-запиту включає ідентифікатор, який використовується для зв’язування запиту з відповіддю. Це 16-бітове поле забезпечує унікальність кожного запиту. У заголовку також є поле прапорів, що визначає властивості та тип запиту. Прапори вказують на те, чи є запит рекурсивним, чи він містить запит на авторитетну відповідь. Також визначається, чи підтримує клієнт додаткові розширення, такі як EDNS.

У структурі запиту важливе місце займає секція Question, яка є основною частиною повідомлення. Вона містить три ключові параметри. Перший параметр QNAME представляє запитуване доменне ім’я в форматі FQDN. Це ім’я включає основний домен і всі рівні піддоменів, розділені крапками, з кожною частиною, що кодується окремо. Другий параметр QTYPE визначає тип запитуваного запису, наприклад, A для отримання IPv4-адреси або AAAA для IPv6-адреси. Третій параметр QCLASS встановлює клас запиту. Зазвичай використовується клас IN для ресурсів у мережі інтернет.

Заголовок запиту також включає кілька лічильників. Поле QDCOUNT вказує кількість записів у секції Question і зазвичай дорівнює одиниці. Поля ANCOUNT,

NSCOUNT і ARCOUNT визначають кількість записів у секціях Answer, Authority і Additional відповідно. У DNS-запиті ці значення зазвичай дорівнюють нулю, оскільки ці секції заповнюються лише у відповіді.

Формат запиту побудований таким чином, щоб бути компактним і ефективним у передачі. Поля заголовка займають фіксовану кількість бітів, тоді як секція Question має змінну довжину залежно від запитуваного доменного імені. Це дозволяє DNS-запиту ефективно обробляти широкий спектр запитів без необхідності в додаткових ресурсах. DNS-запити зазвичай передаються через UDP, що забезпечує низьку затримку, хоча для довших запитів або більшого обсягу даних використовується TCP.

Таблиця 1.1 - Структура DNS-запиту

Поле	Розмір (біт)	Опис
ID (Ідентифікатор)	16	Унікальний ідентифікатор для відповідності запиту та відповіді.
Flags (Прапори)	16	Визначають тип запиту (рекурсивний/нерекурсивний), статус і параметри (наприклад, авторитетність).
QDCOUNT	16	Кількість записів у секції Question (зазвичай 1).
ANCOUNT	16	Завжди дорівнює 0 у запиті (відповідає за кількість записів у секції Answer у відповіді).
NSCOUNT	16	Завжди дорівнює 0 у запиті (відповідає за кількість записів у секції Authority у відповіді).
ARCOUNT	16	Кількість записів у секції Additional (додаткові параметри, наприклад, для OPT записів у EDNS).
Question	Змінна	Основна секція, яка містить запитуване ім’я, тип запису та клас.

В свою чергу структура поля Question представлена в таблиці:

Таблиця 1.2 - Поля секції Question

Поле	Розмір (біт)	Опис
QNAME	Змінна	Запитуване доменне ім’я в форматі FQDN (розділене крапками).
QTYPE	16	Тип запису, який запитується (A, AAAA, MX, CNAME, NS тощо).
QCLASS	16	Клас запиту (зазвичай IN для інтернету).

Таблиця 1.3 - Структура DNS-відповіді

Поле	Розмір (біт)	Опис
ID (Ідентифікатор)	16	Ідентичний ID у запиті, щоб сервер і клієнт могли зіставити запит із відповіддю.
Flags (Прапори)	16	Включає статус запиту (успішно, помилка), авторитетність відповіді та інші параметри.
QDCOUNT	16	Кількість записів у секції Question (ідентична значенню у запиті).
ANCOUNT	16	Кількість записів у секції Answer (результати відповіді).
NSCOUNT	16	Кількість записів у секції Authority (сервери, що є авторитетними для зони).
ARCOUNT	16	Кількість записів у секції Additional (додаткова інформація, наприклад, IP-адреса авторитетного сервера).
Question	Змінна	Повторює секцію Question із запиту.
Answer	Змінна	Містить результати запиту, такі як IP-адреса, до якої резолюється домен.
Authority	Змінна	Інформація про авторитетні сервери, які можуть надати додаткову інформацію.
Additional	Змінна	Додаткова інформація (наприклад, EDNS-параметри або альтернативні IP-адреси).

Таблиця 1.4 - Поля секції Answer, Authority, Additional

Поле	Розмір (біт)	Опис
NAME	Змінна	Доменне ім’я, до якого відноситься запис.
TYPE	16	Тип запису (A, AAAA, MX, CNAME тощо).
CLASS	16	Клас запису (зазвичай IN).
TTL	32	Час життя запису в секундах (вказує, як довго клієнт може зберігати кешовану відповідь).
RDLENGTH	16	Довжина поля даних запису (RDATA).
RDATA	Змінна	Основні дані запису (наприклад, IP-адреса для типу A або список серверів для типу MX).

Ідентифікатор (ID). Унікальний 16-бітовий номер використовується для зв’язування конкретного запиту та відповіді. Клієнт генерує ID у запиті, а сервер включає його у відповідь.

Прапори (Flags). Це 16-бітове поле містить кілька бітових прапорів, кожен із яких відповідає за конкретну функцію або статус. Наприклад:

– QR: визначає, чи це запит (0) чи відповідь (1);
– OPCODE: тип операції (наприклад, стандартний запит або оновлення).
– AA: авторитетність відповіді (1 означає, що сервер є авторитетним).
– TC: вказує на те, що відповідь була обрізана через обмеження UDP.
– RD: вимагає рекурсивної резолюції.
– RA: вказує, чи підтримує сервер рекурсивні запити.
– RCODE: код результату (0 – успішно, 3 – NXDOMAIN, інші – помилки).

Секція Question. Містить основну інформацію про запит: яке доменне ім’я резолвиться, який тип запису потрібен (наприклад, A – для IP-адреси) і який клас (IN – інтернет).

Секція Answer. Ця секція заповнюється сервером у відповіді та містить результати запиту. Наприклад, якщо клієнт запитує A-запис для домену example.com, ця секція міститиме IP-адресу.

Секція Authority. Надає список авторитетних серверів, відповідальних за домен. Використовується, коли запит обробляється нерекурсивним сервером або передається далі.

Секція Additional. Містить додаткові дані, які допомагають клієнту обробити запит, наприклад, IP-адреси авторитетних серверів або параметри розширень (EDNS).

1.6. Технічні засоби збору DNS-запитів

Технічні засоби збору DNS-запитів і відповідей на рівні маршрутизатора забезпечують можливість моніторингу, аналізу та обробки мережевого трафіку в реальному часі. Вони застосовуються для виявлення аномалій, вивчення поведінки клієнтів, оцінки продуктивності мережі та забезпечення безпеки. Ці засоби можна поділити на кілька типів залежно від принципів роботи, ступеня інтеграції з маршрутизатором і способу доступу до трафіку.

Одним із найбільш поширених методів збору DNS-запитів і відповідей є використання технології портового дзеркалювання (Port Mirroring), яка реалізується на рівні маршрутизатора. Ця технологія дозволяє створювати копії трафіку, що проходить через конкретний порт, і перенаправляти їх до зовнішнього пристрою для аналізу. Перевагою цього підходу є його універсальність і сумісність із багатьма маршрутизаторами, оскільки Port Mirroring підтримується більшістю сучасних пристроїв. Водночас його недоліком є залежність від пропускної здатності: у випадку високого навантаження на порт дзеркалювання може виникати втрату даних.

Для збору DNS-запитів також можуть використовуватися спеціалізовані системи NetFlow, які є інтегрованими функціями маршрутизатора. NetFlow дозволяє фіксувати мета-дані про мережеві потоки, включаючи джерело, призначення, обсяг трафіку та часові параметри. Ця технологія є ефективною для моніторингу загальних характеристик трафіку, але не забезпечує детальної інформації про вміст DNS-запитів або відповідей, оскільки зосереджена на заголовках мережевих пакетів. Для детального аналізу DNS-трафіку необхідно інтегрувати NetFlow із додатковими системами збору й обробки.

Технологія Deep Packet Inspection (DPI) є більш розширеним засобом аналізу DNS-запитів і відповідей, яка дозволяє маршрутизатору розпізнавати та обробляти дані на рівні додатків. DPI забезпечує доступ до повного вмісту DNS-пакетів, включаючи запитувані доменні імена, типи записів і параметри відповіді сервера. Ця технологія ефективна для виявлення зловмисних доменів, ботнетів і атак на основі DNS, таких як DNS-ампліфікація. Однак DPI має недоліки, пов’язані зі зниженням продуктивності маршрутизатора, оскільки аналіз усіх пакетів є обчислювально затратним. Крім того, DPI може порушувати конфіденційність, якщо використовується без належних політик доступу до даних.

Іншим підходом є використання проксі-сервера DNS, інтегрованого з маршрутизатором. Проксі-сервер DNS функціонує як посередник між клієнтом і зовнішніми DNS-серверами, записуючи всі запити та відповіді. Цей метод дозволяє повністю контролювати DNS-трафік і забезпечує гнучкість у його аналізі. Наприклад, можна здійснювати фільтрацію запитів до відомих зловмисних доменів у режимі реального часу. Проте проксі-сервер DNS може створювати затримки в обробці запитів, особливо при високих навантаженнях.

Засоби збору DNS-трафіку на основі системи SPAN (Switched Port Analyzer) є схожими на Port Mirroring, але використовують апаратне дублювання трафіку на рівні комутаторів і маршрутизаторів. Ця технологія забезпечує ефективне копіювання великого обсягу даних для аналізу зовнішніми системами. Основним недоліком SPAN є те, що вона вимагає наявності високопродуктивної інфраструктури, оскільки передача дубльованих потоків може створювати додаткове навантаження на мережу.

Також існують системи збору DNS-запитів, інтегровані з SIEM (Security Information and Event Management). Ці системи використовують дані з маршрутизаторів, зібрані за допомогою NetFlow, DPI або портового дзеркалювання, і агрегують їх для створення централізованого аналізу. Основною перевагою є автоматизація процесу обробки та кореляції даних, що дозволяє виявляти аномалії та реагувати на них у реальному часі. Недоліком є складність налаштування та висока вартість впровадження таких рішень.

Для забезпечення гнучкості збору даних маршрутизатори також можуть використовувати інтеграцію з хмарними сервісами. Наприклад, сервіси DNS-розширень (EDNS) дозволяють передавати додаткові дані про запити до аналітичних платформ у хмарі. Цей підхід забезпечує масштабованість і доступність сучасних алгоритмів машинного навчання для аналізу трафіку. Однак він вимагає стабільного підключення до хмари та може створювати ризики для конфіденційності даних.

Таким чином, технічні засоби збору DNS-запитів і відповідей на рівні маршрутизатора варіюються від простих технологій, таких як Port Mirroring, до складних систем DPI і SIEM. Вибір конкретного засобу залежить від потреб мережі, доступного обладнання та необхідного рівня деталізації аналізу трафіку. Кожен із підходів має свої унікальні переваги та недоліки, які необхідно враховувати під час проєктування системи моніторингу та аналізу.

1.7.Огляд відомих підходів до виявлення аномалій у DNS-трафіку

Виявлення аномалій у DNS-трафіку є критично важливим завданням для забезпечення мережевої безпеки та запобігання кіберзагрозам. Різноманітні підходи до цього завдання еволюціонували відповідно до складності мереж і загроз, з якими вони стикаються. Серед них можна виділити класичні методи, що базуються на традиційних підходах до аналізу даних, сучасні методи машинного навчання та інноваційні технології, які інтегрують глибоке навчання та аналіз часових рядів для підвищення точності й адаптивності.

Класичні методи виявлення аномалій у DNS-трафіку включають статистичний аналіз і сигнатурний підхід. Статистичний аналіз базується на побудові моделей нормальної поведінки DNS-запитів із використанням показників, таких як частота запитів, довжина доменних імен, кількість рівнів домену, географічне походження запитів і середній час відповіді серверів. Відхилення від встановлених меж нормальних значень розглядаються як потенційно аномальні. Наприклад, різке зростання кількості запитів із одного джерела може сигналізувати

про DDoS-атаку, тоді як запити до незвичайно довгих доменів можуть бути ознакою використання генераторів доменних імен (DGA). Хоча статистичний аналіз є простим у реалізації та обчислювально ефективним, він має обмеження у виявленні складних або раніше невідомих загроз. Сигнатурний підхід, зі свого боку, базується на ідентифікації аномалій шляхом зіставлення трафіку з базою відомих шкідливих шаблонів або доменів. Цей підхід є високоефективним для виявлення відомих загроз, але вразливий до нових атак, які не входять до бази сигнатур. Окрім того, сигнатурний підхід часто залежить від актуальності бази даних і не здатний адаптуватися до швидкозмінного середовища.

Сучасні методи машинного навчання пропонують більш гнучкий і адаптивний підхід до аналізу DNS-трафіку, включаючи класифікацію та кластеризацію. Класифікація спрямована на побудову моделей, які можуть відносити кожен DNS-запит до однієї з категорій: нормальний або аномальний. Для цього використовуються алгоритми, такі як дерева рішень, SVM, логістична регресія або градієнтний бустинг. Класифікація ефективно працює з маркованими наборами даних, де є приклади нормальних і аномальних запитів, але її ефективність обмежується якістю й обсягом навчальних даних (Рис. 1.5). Кластеризація, навпаки, дозволяє виявляти аномалії в немаркованих даних, групуючи подібні запити у кластери та визначаючи ті, що не відповідають основним групам. Для цього часто використовуються алгоритми, такі як K-Means, DBSCAN або Gaussian Mixture Models. Кластеризація є корисною для виявлення нових загроз, але її точність залежить від вибору гіперпараметрів і метрики подібності.

Сучасні методи машинного навчання пропонують більш гнучкий і адаптивний підхід до аналізу DNS-трафіку, включаючи класифікацію та кластеризацію. Класифікація спрямована на побудову моделей, які можуть відносити кожен DNS-запит до однієї з категорій: нормальний або аномальний. Для цього використовуються алгоритми, такі як дерева рішень, SVM, логістична регресія або градієнтний бустинг. Класифікація ефективно працює з маркованими наборами даних, де є приклади нормальних і аномальних запитів, але її

ефективність обмежується якістю й обсягом навчальних даних (Рис. 1.5). Кластеризація, навпаки, дозволяє виявляти аномалії в немаркованих даних, групуючи подібні запити у кластери та визначаючи ті, що не відповідають основним групам. Для цього часто використовуються алгоритми, такі як K-Means, DBSCAN або Gaussian Mixture Models. Кластеризація є корисною для виявлення нових загроз, але її точність залежить від вибору гіперпараметрів і метрики подібності.

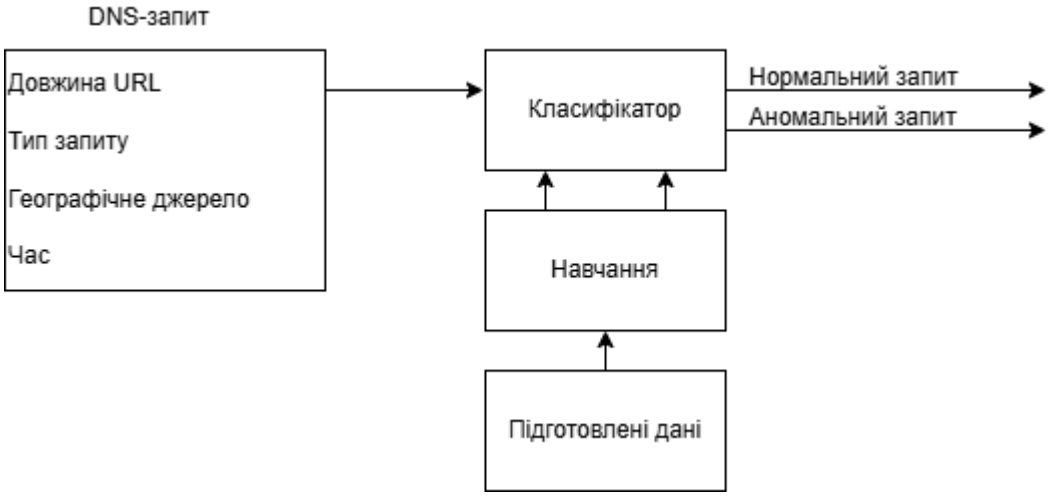

Рисунок 1.5 Класифікатор для виявлення аномалій у DNS-трафіку

Інноваційні методи, що базуються на глибокому навчанні та аналізі часових рядів, відкривають нові горизонти в автоматизованому виявленні аномалій у DNS-трафіку. Глибокі нейронні мережі, такі як рекурентні нейронні мережі (RNN) і згорткові нейронні мережі (CNN), використовуються для виявлення складних шаблонів у великих наборах даних, включаючи текстові та часові аспекти DNS-запитів. Наприклад, RNN добре підходять для аналізу послідовностей запитів, тоді як CNN можуть виявляти структурні аномалії у текстових представленнях доменних імен. Глибоке навчання дозволяє створювати моделі, які здатні самостійно виявляти нові загрози на основі великих обсягів даних, але вимагає значних обчислювальних ресурсів і якісних даних для навчання. Аналіз часових рядів з використанням методів, таких як LSTM-мережі або ARIMA-моделі, дозволяє виявляти аномалії, що виникають через відхилення у тимчасових

патернах DNS-запитів, наприклад, несподівані піки активності або нерівномірний розподіл трафіку (Рис. 1.6).

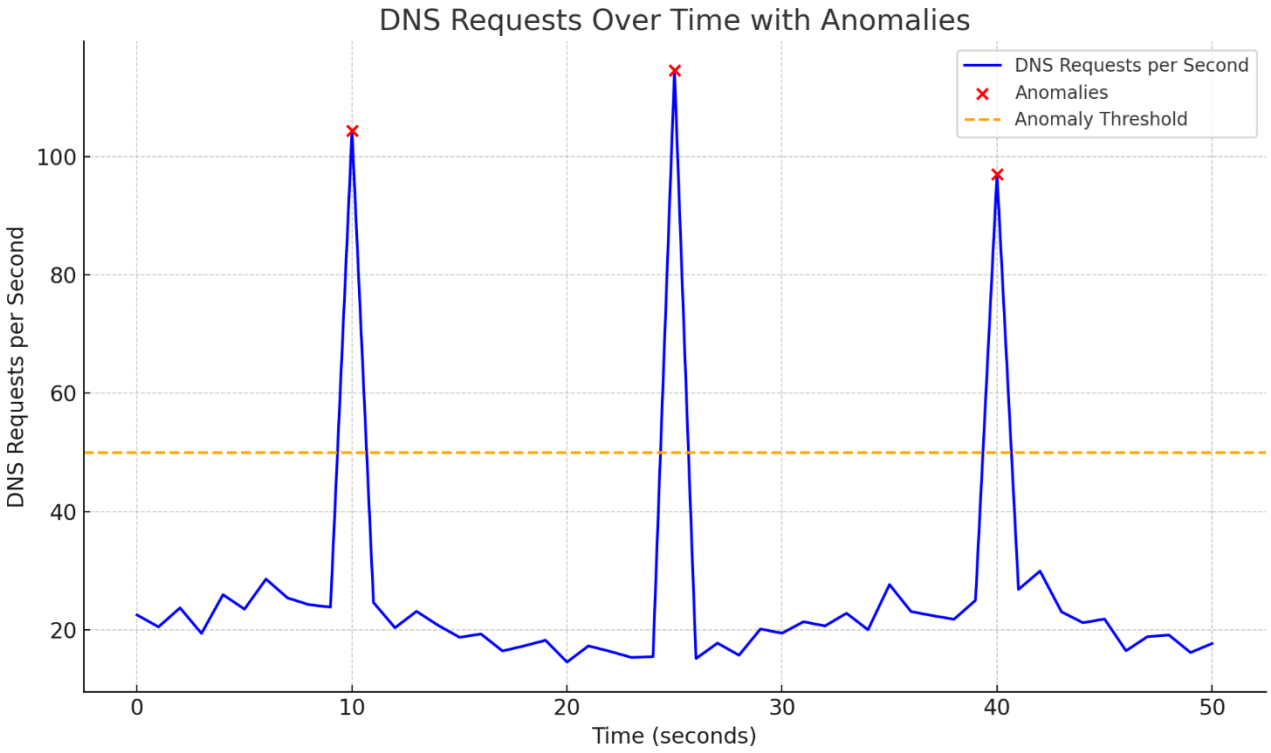

Рисунок 1.6 Аналіз часових рядів DNS-запитів

Таким чином, різні підходи до виявлення аномалій у DNS-трафіку мають свої переваги та обмеження. Класичні методи забезпечують простоту й ефективність для базових завдань, але не адаптуються до нових загроз. Методи машинного навчання дозволяють виявляти складніші шаблони, тоді як інноваційні технології пропонують глибший аналіз і здатність до адаптації в умовах динамічного середовища.

1.8.Постановка задачі

Сучасні мережі стикаються зі значними викликами у сфері кібербезпеки, і одним із найважливіших напрямків є захист системи доменних імен (DNS) від зловмисних дій. DNS є критичним компонентом мережевої інфраструктури, який забезпечує трансляцію доменних імен у IP-адреси. Його робота часто стає об'єктом атак або використовується як вектор для поширення шкідливого програмного

забезпечення, проведення DDoS-атак, організації каналів витоку даних та інших загроз. Ці виклики зумовлюють необхідність розробки ефективних методів аналізу та виявлення аномалій у DNS-трафіку.

У рамках цієї роботи поставлено завдання розробити метод виявлення аномалій у DNS-запитах, який буде заснований на сучасних підходах до аналізу мережевого трафіку. Для досягнення цієї мети необхідно виконати такі завдання:

1. Провести аналіз існуючих підходів до виявлення аномалій у DNS-трафіку, включаючи статистичні моделі, методи машинного навчання та інші технології.

2. Сформувати профіль нормальної активності DNS-трафіку на основі зібраних і підготовлених даних, що дозволить створити базову модель для подальшого виявлення відхилень.

3. Розробити алгоритм, який буде використовувати комбінований підхід до виявлення аномалій, інтегруючи статистичний аналіз та моделі машинного навчання для підвищення точності й надійності.

4. Реалізувати програмний прототип системи виявлення аномалій, який здатен обробляти вхідні DNS-логи у реальному часі, аналізувати їх і виявляти відхилення.

5. Оцінити достовірність розробленого методу на реальних даних, визначити його сильні та слабкі сторони, а також розробити рекомендації щодо його впровадження в мережеву інфраструктуру.

Розробка ефективного методу виявлення аномалій у DNS-запитах є важливою задачею з огляду на поточний рівень кіберзагроз. Більшість традиційних підходів до аналізу мережевого трафіку мають обмежену ефективність у виявленні складних та прихованих загроз. Наприклад, використання статистичних методів дозволяє визначати загальні відхилення, але є недостатньо гнучким для адаптації до змін у поведінці зловмисників. Методи машинного навчання забезпечують більш високу точність, але потребують ретельної підготовки даних і налаштування параметрів. Тому інтеграція кількох підходів є ключовою для досягнення високої ефективності.

2. АНОМАЛІЇ DNS-ЗАПИТІВ

2.1.Синтаксичні аномалії у DNS-запитах

Синтаксичні аномалії у DNS-запитах виникають, коли формати або структура запитів порушують усталені правила та стандарти, встановлені для протоколу DNS. Протокол DNS базується на чітких специфікаціях, які визначають, як мають формуватися, надсилатися та оброблятися запити. Відхилення від цих стандартів можуть бути спричинені помилками програмного забезпечення, некоректною конфігурацією систем, зловмисними діями або навіть специфічними особливостями нових технологій, які виходять за межі класичного розуміння протоколу.

Однією з ключових причин виникнення синтаксичних аномалій є порушення форматування полів DNS-запиту. Кожен запит складається з кількох структурних елементів, таких як заголовок, питання, відповідь, авторитетна інформація та додаткова інформація. Вони повинні дотримуватися чітко визначених стандартів, але у випадку аномалій можуть містити некоректні значення, неправильну довжину або невідповідності у форматуванні. Наприклад, поле заголовка може мати некоректну кількість байтів, що ускладнює або робить неможливим обробку такого запиту сервером.

Некоректний формат також може виникати у запитах до доменів, що містять недопустимі символи, наприклад пробіли, спеціальні символи або символи, які не належать до кодування UTF-8. Важливим аспектом є те, що деякі аномалії можуть бути створені зловмисниками спеціально для експлуатації вразливостей у програмному забезпеченні серверів. Це може призводити до збоїв у роботі серверів або використання їх у шкідливих цілях.

Запити до некоректних доменів також відносяться до синтаксичних аномалій. У цьому випадку користувачі або пристрої намагаються звернутися до доменів, які не можуть існувати згідно зі стандартами DNS. Це можуть бути домени, що не відповідають загальноприйнятим правилам формування імен, або ж домени, що

спеціально зарезервовані для внутрішнього використання. Наприклад, згідно з RFC 2606, існує перелік зарезервованих доменів, які не повинні використовуватися у глобальній мережі. Запити до таких доменів можуть бути наслідком помилок у конфігурації або зловмисних дій. Відправка запитів до зарезервованих доменів також може свідчити про проблеми у конфігурації локальних мереж або DNS-серверів.

Порушення стандартів RFC є однією з основних причин синтаксичних аномалій. Наприклад, RFC 1035 чітко визначає правила для довжини імен доменів, які не повинні перевищувати 255 символів. Якщо клієнт надсилає запит із доменом більшої довжини, це може спричинити відмову сервера у відповідях або навіть створити загрозу безпеці, якщо сервер обробляє такі запити некоректно. Схожі проблеми виникають, якщо запити містять некоректно сформовані записи ресурсів, такі як A-записи, CNAME чи TXT. У деяких випадках такі запити можуть залишатися непоміченими, якщо сервери недостатньо захищені або неправильно налаштовані.

Ще однією проблемою є запити до доменів, які використовують застарілі або невизначені стандарти. Такі запити можуть бути наслідком використання застарілого програмного забезпечення, що не підтримує сучасні специфікації. Наприклад, запити до IPv4-доменів з використанням протоколів, що не підтримують IPv6, можуть вважатися аномальними у сучасних мережах, де очікується повна підтримка нових протоколів.

Особливої уваги заслуговують запити, спрямовані на домени, що використовують нетипові TLD (доменні зони верхнього рівня). Синтаксичні аномалії у цьому випадку виникають через некоректну обробку сервером або клієнтом нових або експериментальних доменних зон. Наприклад, введення нових TLD, таких як .example або .test, може спричинити появу некоректних запитів, якщо програмне забезпечення або налаштування мереж не адаптовані до змін.

Проблеми можуть також бути спричинені неправильним кодуванням символів у DNS-запитах. Наприклад, запити, що використовують IDN (міжнародні доменні імена), вимагають кодування у форматі Punycode. Якщо це кодування

виконується некоректно, сервери можуть не розпізнавати домени або обробляти їх неправильно. Такі аномалії стають дедалі поширенішими у глобальних мережах, де використання національних мов у доменних іменах стає нормою.

Некоректний формат DNS-запитів може також бути наслідком цілеспрямованих дій, наприклад, DNS-атаки. Атаки, що використовують некоректно сформовані запити, часто спрямовані на перевантаження серверів або їх введення в стан некоректної роботи. Такі атаки можуть використовувати аномальні запити, що містять повторювані записи, дуже великі пакети або інші нехарактерні для протоколу особливості.

Синтаксичні аномалії у DNS-запитах становлять серйозну загрозу для стабільності та безпеки мереж. Їх виявлення та аналіз вимагають використання спеціалізованого програмного забезпечення, яке дозволяє відслідковувати відхилення від стандартів та визначати причини їх появи. Зокрема, системи моніторингу DNS повинні мати можливість порівнювати запити із загальноприйнятими стандартами та виявляти будь-які невідповідності. Такий підхід дозволяє не лише виявляти аномалії, але й запобігати їх подальшому поширенню.

Своєчасне виявлення синтаксичних аномалій дозволяє забезпечити безпеку та стабільність мережі, мінімізуючи ризики, пов’язані із зловмисними діями або технічними помилками. Впровадження автоматизованих систем аналізу DNS-запитів є важливим кроком для будь-якої сучасної мережі, яка прагне захистити свої ресурси та користувачів від потенційних загроз.

2.2. Аномалії частоти та інтенсивності у DNS-запитах

Аномалії частоти та інтенсивності у DNS-запитах є важливими індикаторами порушень нормальної роботи мережі або потенційних загроз безпеці. Такі аномалії виникають, коли обсяг або частота запитів перевищує звичайні очікування для конкретного домену чи сервісу, або коли запити демонструють нехарактерні для звичайної діяльності мережі поведінкові патерни. Це може бути наслідком помилок

у роботі програмного забезпечення, некоректних налаштувань систем чи мережевого обладнання або ж цілеспрямованих атак.

Однією з найпоширеніших форм аномалій частоти є надмірно часті запити до одного й того ж домену. У нормальних умовах частота запитів до DNS-серверів є відносно стабільною та залежить від кількості користувачів, які звертаються до ресурсу, та тривалості TTL (time-to-live) записів у кеші. Якщо домен запитують надзвичайно часто, це може свідчити про кілька можливих проблем. Наприклад, клієнтський пристрій або програма може бути неправильно налаштована або містити помилки, через які запити надсилаються без врахування кешування. У деяких випадках це свідчить про використання застарілих чи несправних протоколів, які не оптимізують DNS-запити.

Також надмірно часті запити можуть бути результатом автоматизованих дій, наприклад, ботів, що сканують домени або перевіряють їхню доступність. У таких випадках джерелом аномалії може бути як легітимна активність, наприклад перевірка доменів у рамках SEO-аналізу чи моніторингу доступності, так і зловмисна діяльність, спрямована на підготовку до атак або виявлення вразливостей у сервісах. Автоматизовані системи часто не враховують навантаження на DNS-сервери, через що їхня діяльність може створювати значні проблеми для інфраструктури.

Іншим поширеним типом аномалій частоти є підозріла кількість однотипних запитів у короткий проміжок часу. Це може бути ознакою розподіленої атаки на відмову в обслуговуванні або DDoS-атаки. Під час такої атаки зловмисники використовують велику кількість заражених пристроїв (ботів), щоб надсилати однотипні запити до цільового DNS-сервера або домену. Мета такої атаки полягає у виснаженні ресурсів сервера, що призводить до неможливості обслуговування легітимних користувачів. У багатьох випадках під час DDoS-атак використовуються техніки підробки IP-адрес, що ускладнює ідентифікацію джерела атаки та блокування трафіку.

Кількість однотипних запитів також може збільшуватися через некоректну конфігурацію клієнтських пристроїв або систем, які працюють з DNS. Наприклад,

пристрої без належного кешування DNS-записів можуть постійно надсилати запити до того ж домену, навіть якщо відповідь на цей запит уже була отримана. Така поведінка призводить до перевантаження як локальних серверів, так і глобальних DNS-систем. Особливо це актуально у великих мережах із численними клієнтами, які використовують централізовані DNS-сервери без належної оптимізації.

Ще однією поширеною причиною аномалій частоти є намагання обійти обмеження доступу до певних ресурсів. Наприклад, у деяких випадках користувачі намагаються зняти блокування доступу до вебсайтів шляхом постійного звернення до DNS-серверів з різними параметрами. Ця активність може бути сприйнята як аномальна через високу частоту та відсутність логічної залежності між запитами.

Окрему увагу варто приділити системам моніторингу мереж, які періодично надсилають DNS-запити для перевірки доступності доменів або серверів. Такі системи можуть генерувати надмірний трафік, якщо вони неправильно налаштовані. Наприклад, занадто короткі інтервали між перевірками можуть створювати навантаження, яке прирівнюється до аномального. Це особливо помітно у великих інфраструктурах, де моніторинг охоплює сотні чи тисячі доменів.

Щоб виявити та аналізувати аномалії частоти та інтенсивності DNS-запитів, необхідно використовувати спеціалізовані інструменти моніторингу. Такі системи можуть визначати базові показники нормальної активності та виявляти відхилення від них. Для цього застосовуються алгоритми виявлення аномалій, які враховують такі параметри як частота запитів, час їх надходження, джерела трафіку та типи запитів. Особливо важливою є можливість кореляції даних із різних джерел, що дозволяє більш точно визначити характер проблеми.

Важливим кроком у боротьбі з аномаліями є застосування механізмів автоматичного блокування підозрілих запитів. Наприклад, більшість сучасних DNS-серверів дозволяють налаштовувати обмеження на частоту запитів з одного джерела. У разі перевищення встановленого ліміту такі запити можуть блокуватися

автоматично. Це дозволяє мінімізувати ризики, пов’язані з перевантаженням серверів, та захистити мережу від збоїв.

Аномалії частоти та інтенсивності також можуть впливати на ефективність роботи кешуючих DNS-серверів. Якщо клієнти надсилають надмірну кількість запитів до одного й того ж домену, сервер може перевантажитися спробами постійно оновлювати кеш. Це знижує продуктивність мережі та може призвести до затримок у відповіді для легітимних користувачів. Щоб уникнути таких проблем, необхідно правильно налаштовувати кешування та TTL-записи.

Наслідки аномалій частоти можуть включати як локальні проблеми, так і ширші впливи на мережеву інфраструктуру. Наприклад, перевантаження локального DNS-сервера може спричинити затримки у наданні послуг для всіх клієнтів мережі. У глобальному масштабі це може призвести до перевантаження кореневих серверів або інших критичних компонентів DNS-інфраструктури. Тому важливо вчасно ідентифікувати такі аномалії та вживати заходів для їх усунення.

У контексті кібербезпеки аномалії частоти та інтенсивності є важливим індикатором загроз. Вони можуть свідчити про спроби зловмисників дослідити мережеву інфраструктуру, підготувати атаки або використовувати мережу для розповсюдження шкідливого програмного забезпечення. У таких випадках важливо не лише блокувати аномальні запити, але й аналізувати їх для визначення потенційних загроз та вразливостей мережі.

2.3.Поведінкові аномалії у DNS-запитах

Поведінкові аномалії у DNS-запитах є важливим індикатором змін у стандартній активності користувачів або пристроїв у мережі. Такі аномалії виникають, коли поведінка клієнтів або окремих вузлів відхиляється від очікуваних або типових моделей. Виявлення подібних аномалій є критично важливим завданням для забезпечення мережевої безпеки, оскільки вони часто вказують на спроби зловмисних дій, технічні збої або неправильно налаштовані системи.

Одним із ключових проявів поведінкових аномалій є нетипові патерни запитів для конкретного клієнта або мережі. У звичайних умовах клієнти надсилають DNS-запити у відповідності до їхньої активності в інтернеті. Це можуть бути запити до відомих доменів, таких як вебсайти, сервіси чи ресурси, які користувачі відвідують регулярно. Нетиповий патерн DNS-запитів виникає тоді, коли клієнт надсилає запити до доменів, які зазвичай не викликають інтересу, або коли частота запитів значно перевищує нормальні показники. Наприклад, запити до підозрілих або незвичайних доменів можуть вказувати на те, що пристрій клієнта був скомпрометований шкідливим програмним забезпеченням.

Однією з форм поведінкових аномалій є раптове збільшення кількості запитів до доменів, які користувач раніше не відвідував. Це може свідчити про те, що клієнт намагається отримати доступ до ресурсів, які виходять за межі його типової поведінки. Подібна активність може бути наслідком автоматизованих дій програмного забезпечення, наприклад ботів або скриптів, які використовуються для сканування або перевірки доступності доменів. В інших випадках така поведінка може бути наслідком дій шкідливого програмного забезпечення, яке працює на пристрої клієнта та використовує DNS для зв'язку із командними серверами або завантаження додаткових компонентів.

Нетипові патерни DNS-запитів часто є ознакою діяльності зловмисного ПЗ. Шкідливі програми, такі як трояни чи віруси, можуть використовувати DNS для виконання своїх функцій. Наприклад, вони можуть надсилати регулярні запити до певних доменів для отримання команд від зловмисників або для передачі зібраних даних. У таких випадках поведінка клієнта або пристрою значно відрізняється від нормальної, оскільки звичайні користувачі не мають причин звертатися до таких доменів. Виявлення подібних аномалій вимагає аналізу поведінкових моделей клієнтів у мережі та порівняння їх із базовими лініями активності.

Ще одним важливим аспектом поведінкових аномалій є запити до доменів, які нещодавно з'явилися. Нові домени часто використовуються зловмисниками, оскільки вони менш імовірно знаходяться у чорних списках або базах даних відомих шкідливих доменів. Такі домени можуть бути зареєстровані спеціально

для короткочасного використання у фішингових атаках, поширенні шкідливого ПЗ або інших зловмисних активностях. Зазвичай вони існують лише протягом короткого періоду часу, після чого зловмисники залишають їх, щоб уникнути ідентифікації.

DNS-запити до нових доменів можуть бути частиною роботи шкідливого програмного забезпечення, яке генерує псевдовипадкові доменні імена для зв'язку з командними серверами. Цей механізм, відомий як алгоритми генерації доменних імен (DGA), дозволяє зловмисникам створювати тисячі доменів за короткий час. Якщо DNS-сервер отримує запити до таких доменів, це часто вказує на те, що у мережі присутні заражені пристрої. Виявлення подібної активності вимагає ретельного аналізу DNS-логів і використання методів машинного навчання для виявлення доменів, які не відповідають типовим патернам.

Шкідливі програми також можуть використовувати запити до нових доменів для уникнення виявлення. Наприклад, деякі віруси надсилають запити до доменів, які виглядають легітимними, але насправді є підробленими. Ці домени можуть імітувати назви популярних сервісів або організацій, використовуючи незначні відмінності у написанні. Наприклад, домен з помилкою типу "goggle.com" замість "google.com" може бути використаний для перенаправлення користувачів на фішинговий сайт.

Аналіз поведінкових аномалій у DNS-запитах також включає виявлення повторюваних запитів до підозрілих доменів, навіть якщо ці домени поки що не заблоковані. Зазвичай шкідливі програми продовжують надсилати запити до своїх командних серверів, навіть якщо ці сервери недоступні або заблоковані. Ця поведінка створює характерний патерн, який можна використовувати для ідентифікації заражених пристроїв.

Значна частина поведінкових аномалій пов'язана із так званими "низькопрофільними" атаками, коли зловмисники намагаються залишатися непоміченими. У таких випадках вони використовують мінімальну кількість запитів до нових доменів або приховують свою діяльність за легітимним трафіком. Наприклад, вони можуть вставляти підозрілі запити у нормальні сесії браузера,

щоб ускладнити їх виявлення. Це підкреслює важливість використання сучасних систем моніторингу DNS, які здатні виявляти навіть незначні відхилення у поведінці клієнтів.

Виявлення поведінкових аномалій у DNS-запитах є ключовим елементом забезпечення безпеки сучасних мереж. Для цього необхідно застосовувати комплексний підхід, який включає аналіз логів, створення базових моделей активності та використання інструментів аналізу великих даних. Такий підхід дозволяє вчасно виявляти загрози та зменшувати їхній вплив на мережеву інфраструктуру.

2.4. Статистичні моделі для виявлення аномалій

Статистичні моделі є основою для виявлення аномалій у великих обсягах даних, зокрема у мережевому трафіку чи логах DNS-запитів. Цей підхід ґрунтується на математичному аналізі характеристик даних та пошуку значень, які значно відхиляються від типових показників. Для цього застосовуються розподіли ймовірностей, які дозволяють описати звичайну поведінку системи. Серед найпоширеніших статистичних розподілів, що використовуються для виявлення аномалій, є Гауссовий та Пуассонівський. Кожен із цих підходів має свої особливості і є ефективним у різних сценаріях.

Гауссовий розподіл, або нормальний розподіл, є одним із найбільш відомих розподілів у статистиці. Він описує дані, значення яких концентруються навколо середнього з певним ступенем варіативності, що визначається стандартним відхиленням. Така модель є дуже зручною для аналізу, оскільки вона дозволяє оцінити ймовірність будь-якого значення, враховуючи його відстань від середнього значення. Гауссовий розподіл широко застосовується для моделювання метрик, які у своїй природі мають симетричну структуру навколо середнього значення. Наприклад, у контексті аналізу DNS-запитів середня кількість запитів від клієнта за одиницю часу може підпорядковуватися нормальному розподілу.

Виявлення аномалій на основі Гауссового розподілу передбачає визначення діапазону нормальних значень, які знаходяться в межах кількох стандартних відхилень від середнього. Наприклад, якщо активність клієнта значно перевищує цей діапазон, це може вказувати на аномальну поведінку, таку як зловмисна активність або технічні збої. Аналіз розподілу даних також дозволяє виявити неочікувано низькі показники, що можуть свідчити про відмову у системі або блокування доступу до ресурсу. Гауссовий розподіл ефективний для виявлення симетричних аномалій, але менш придатний для аналізу даних, які мають сильно зміщений розподіл.

У випадках, коли дані мають асиметричний характер, наприклад, коли значна частина значень є низькими, але рідкісні високі значення суттєво перевищують середнє, Пуассонівський розподіл стає більш доцільним. Пуассонівський розподіл описує кількість подій, які відбуваються за фіксований проміжок часу або у визначеному просторі, за умови, що ці події є незалежними одна від одної. Цей розподіл часто використовується для моделювання рідкісних подій, наприклад, запитів до певних ресурсів у мережі.

Для виявлення аномалій із використанням Пуассонівського розподілу необхідно визначити очікувану частоту подій у нормальних умовах. Наприклад, якщо сервер обробляє в середньому 10 запитів до певного ресурсу на годину, але в якийсь момент кількість запитів різко зростає до 50, це може вказувати на аномалію. Такий підхід дозволяє ефективно аналізувати нерівномірно розподілені дані, зокрема пікові навантаження на сервер або сплески активності в межах короткого періоду часу.

Перевагою Пуассонівського розподілу є його простота та можливість моделювати дискретні події. Наприклад, він дозволяє аналізувати, як часто виникають запити до рідко використовуваних доменів або як часто клієнти звертаються до певних IP-адрес. Це робить його ефективним для аналізу поведінки, яка зазвичай має низьку частоту, але у разі аномалії може суттєво перевищити очікуваний рівень. У контексті аналізу DNS-запитів це може бути корисним для

виявлення шкідливої активності, наприклад, ботнетів, які генерують велику кількість запитів за короткий час.

Поєднання Гауссового та Пуассонівського розподілів дозволяє створити комплексні моделі для аналізу поведінки систем. Наприклад, Гауссовий розподіл може бути використаний для моделювання середньої активності, тоді як Пуассонівський підходить для аналізу рідкісних подій або сплесків. Це особливо корисно у великих мережах, де активність різних клієнтів або підмереж може мати суттєві відмінності. Для створення таких моделей використовується історична інформація про поведінку системи, яка дозволяє визначити базові лінії та очікувані характеристики нормальної роботи.

Використання статистичних розподілів також дозволяє автоматизувати процес виявлення аномалій. Наприклад, системи моніторингу можуть автоматично визначати відхилення від нормальних показників, базуючись на математичних моделях. У разі виявлення аномалії такі системи можуть надсилати сповіщення або запускати механізми захисту, наприклад, блокування підозрілих запитів. Це дозволяє значно підвищити ефективність аналізу та зменшити час реакції на потенційні загрози.

Застосування статистичних моделей також передбачає певні виклики. Одним із них є необхідність коректного налаштування параметрів моделі, наприклад, вибору порогових значень для виявлення аномалій. Занадто високі пороги можуть призвести до пропуску важливих інцидентів, тоді як надто низькі значення можуть створити велику кількість хибних спрацьовувань. Для вирішення цієї проблеми часто використовуються методи машинного навчання, які дозволяють автоматично налаштовувати параметри моделі на основі історичних даних.

Ще одним викликом є складність аналізу в реальному часі у великих мережах. Велика кількість даних може вимагати значних обчислювальних ресурсів для обробки, особливо якщо моделі базуються на складних математичних алгоритмах. Для оптимізації цього процесу використовуються розподілені системи обробки даних, які дозволяють аналізувати інформацію паралельно на декількох вузлах.

Статистичні моделі на основі Гауссового та Пуассонівського розподілів є ефективним інструментом для виявлення аномалій у DNS-запитах та інших мережевих метриках. Їх використання дозволяє забезпечити високу точність аналізу та мінімізувати ризики, пов’язані з пропуском загроз або хибними спрацьовуваннями.

Імовірнісні підходи, такі як методи Байєса, є одним із найпоширеніших та найефективніших інструментів для виявлення аномалій у великих масивах даних. Ці методи базуються на використанні теореми Байєса, яка дозволяє оцінювати ймовірність певної події за умови наявності попередньої інформації про неї. Завдяки своїй гнучкості та здатності працювати в умовах невизначеності, методи Байєса стали ключовим елементом багатьох сучасних систем аналізу даних. У контексті виявлення аномалій вони дозволяють ефективно аналізувати мережевий трафік, поведінку користувачів та інші метрики, які можуть свідчити про відхилення від норми.

Основна ідея використання методів Байєса полягає у розрахунку апостеріорної ймовірності аномалії на основі доступних даних. Це досягається шляхом комбінування апріорної ймовірності події із ймовірністю спостережуваних даних. Наприклад, якщо відомо, що у середньому лише один із десяти тисяч DNS-запитів є аномальним, а також є дані, що певний запит демонструє нетипову поведінку, методи Байєса дозволяють оцінити ймовірність того, що цей запит є дійсно аномальним.

Методи Байєса особливо ефективні у сценаріях, де є велика кількість історичних даних, які можна використовувати для навчання моделі. Наприклад, у мережевому аналізі можна зібрати інформацію про нормальну поведінку клієнтів, таких як частота DNS-запитів, середня кількість запитів на одиницю часу та типи доменів, до яких вони звертаються. Використовуючи ці дані, система може побудувати апріорний розподіл, який відображає ймовірність нормальної поведінки. Якщо у реальному часі з'являються нові дані, вони можуть бути інтегровані у модель для оновлення апостеріорних ймовірностей.

Однією з основних переваг методів Байєса є можливість врахування різних джерел інформації. У реальних умовах дані можуть бути неоднорідними та містити шум, але імовірнісні моделі дозволяють враховувати ці фактори під час аналізу. Наприклад, при виявленні аномалій у DNS-запитах можна комбінувати інформацію про частоту запитів, типи доменів та географічне розташування клієнтів для точнішого визначення, чи є певна поведінка аномальною.

Методи Байєса також добре підходять для аналізу подій, які трапляються рідко, що є однією з основних характеристик аномалій. Наприклад, атаки типу DDoS чи ботнет-активність можуть проявлятися як раптове збільшення запитів до певних ресурсів. Використовуючи попередню інформацію про ймовірність таких подій та їх характерні ознаки, байєсівські моделі можуть швидко визначати, чи є поточна активність відхиленням від норми.

Ще однією важливою перевагою методів Байєса є їхня інтерпретованість. На відміну від багатьох сучасних алгоритмів машинного навчання, таких як нейронні мережі, байєсівські моделі забезпечують чітке розуміння, як і чому була прийнята певна оцінка. Це особливо важливо у випадках, коли системи виявлення аномалій застосовуються у критичних галузях, таких як кібербезпека чи фінанси, де необхідно обґрунтовувати рішення, прийняті автоматизованими системами.

Важливим аспектом застосування байєсівських методів є вибір відповідної моделі та параметрів. У контексті аналізу DNS-запитів можна використовувати різні варіанти моделювання. Наприклад, прості моделі, такі як наївний Байєс, припускають, що всі ознаки є незалежними одна від одної. Хоча це припущення часто не відповідає реальності, наївний Байєс може бути дуже ефективним для швидкого виявлення аномалій у великих обсягах даних. Більш складні моделі, такі як байєсівські мережі, дозволяють враховувати залежності між різними ознаками, що забезпечує вищу точність, але вимагає більше обчислювальних ресурсів.

Для реалізації байєсівських методів у системах моніторингу використовуються сучасні інструменти, які дозволяють автоматизувати процеси збору, аналізу та обробки даних. Наприклад, системи аналізу великих даних можуть інтегрувати байєсівські моделі для виявлення відхилень у режимі

реального часу. Це особливо корисно у великих мережах, де кількість подій може сягати мільярдів на добу, і ручний аналіз стає неможливим.

Одним із викликів у застосуванні методів Байєса є потреба у великій кількості даних для створення точних апостеріорних розподілів. Якщо дані є неповними або мають значний рівень шуму, результати аналізу можуть бути менш точними. Для вирішення цієї проблеми часто використовуються комбіновані підходи, які включають використання байєсівських методів разом із іншими алгоритмами машинного навчання, такими як кластеризація або аналіз головних компонент.

Методи Байєса також можуть бути адаптовані для роботи в умовах змінного середовища, де характеристики даних можуть змінюватися з часом. Наприклад, у мережах з високою динамікою активності модель може періодично оновлювати апостеріорні ймовірності, враховуючи нові дані. Це дозволяє системі залишатися актуальною та ефективною навіть у швидко змінюваних умовах.

Статистичні моделі, засновані на методах Байєса, забезпечують потужний інструмент для виявлення аномалій у великих даних. Їх застосування дозволяє не лише автоматизувати процес аналізу, але й забезпечити високу точність та інтерпретованість результатів. Завдяки своїм перевагам ці методи стали невід'ємною частиною сучасних систем забезпечення безпеки та аналізу даних.

2.5.Моделі машинного навчання

Моделі машинного навчання займають важливе місце у виявленні аномалій завдяки своїй здатності аналізувати великі обсяги даних, знаходити приховані залежності та приймати рішення на основі цих знань. Одним із основних підходів у машинному навчанні для виявлення аномалій є алгоритми класифікації, зокрема SVM (машини опорних векторів) та дерева рішень. Ці алгоритми дають змогу ефективно класифікувати дані на категорії, зокрема виявляти аномальні записи серед нормальних.

SVM, або машини опорних векторів, є одним із найпотужніших методів класифікації, які використовуються для задач з двома або більше класами. Основна ідея SVM полягає у побудові гіперплощини, яка максимально розділяє дані двох класів. Для виявлення аномалій цей метод особливо ефективний, оскільки дозволяє виділити аномальні записи навіть у випадках, коли вони складають незначну частину загального набору даних. Наприклад, у задачах аналізу DNS-запитів SVM може бути використаний для класифікації запитів на нормальні та підозрілі, беручи до уваги такі фактори, як частота запитів, тип домену, що запитується, та час здійснення запиту.

Одна з переваг SVM полягає у здатності працювати з високорозмірними даними. Більшість реальних задач машинного навчання, зокрема у сфері кібербезпеки, включають велику кількість характеристик, які описують дані. Наприклад, у мережевому аналізі може бути враховано десятки параметрів для кожного запиту. SVM дозволяє ефективно обробляти такі дані завдяки використанню ядрових функцій, які дозволяють перетворювати вихідний простір характеристик у більш високорозмірний простір. Це робить можливим лінійне розділення класів навіть у випадках, коли це не можна зробити у вихідному просторі.

Алгоритм SVM також є стійким до перенавчання, особливо у задачах з невеликою кількістю аномальних записів. Завдяки використанню опорних векторів модель фокусується лише на найбільш значущих записах, що дозволяє уникнути перенавчання на шумових даних. У контексті виявлення аномалій це є критично важливим, оскільки аномалії зазвичай рідкісні та можуть бути змішані із нормальними даними.

Іншим популярним алгоритмом класифікації для виявлення аномалій є дерева рішень. Цей підхід базується на ітеративному поділі даних на підмножини відповідно до певних правил, які максимізують чистоту розділених класів. Дерева рішень є надзвичайно інтуїтивними, оскільки їх структуру легко зрозуміти та інтерпретувати. Наприклад, у задачі виявлення аномальних DNS-запитів дерево рішень може враховувати такі критерії, як кількість запитів за одиницю часу,

доменна зона, до якої звертається користувач, та відповідність запитів стандартним патернам поведінки.

Однією з ключових переваг дерев рішень є їх здатність працювати з даними, які містять як числові, так і категоріальні характеристики. Наприклад, у аналізі мережевого трафіку можуть використовуватися як числові значення (частота запитів), так і категоріальні (типи запитуваних ресурсів). Дерева рішень дозволяють легко інтегрувати обидва типи характеристик у модель, що робить цей підхід дуже універсальним.

Ще однією перевагою дерев рішень є те, що вони не вимагають масштабування даних або складної підготовки. На відміну від SVM, які можуть вимагати нормалізації або стандартизації характеристик, дерева рішень здатні працювати з даними у їх початковій формі. Це значно спрощує процес розробки моделі та знижує вимоги до попередньої обробки даних.

Однак дерева рішень мають і певні недоліки. Основним викликом є їх схильність до перенавчання, особливо у випадках, коли модель стає занадто складною через велику кількість гілок. Це може призвести до погіршення здатності моделі узагальнювати нові дані. Для подолання цієї проблеми часто використовуються ансамблеві методи, такі як Random Forest або Gradient Boosting, які поєднують кілька дерев рішень для покращення точності та стійкості моделі.

Поєднання SVM та дерев рішень може дати ще кращі результати у задачах виявлення аномалій. Наприклад, дерева рішень можуть використовуватися для попереднього відбору характеристик, які є найбільш значущими для класифікації, тоді як SVM може застосовуватися для побудови точної моделі на основі відібраних характеристик. Такий підхід дозволяє отримати більш точні результати, знижуючи обчислювальні витрати.

Застосування цих алгоритмів у реальних системах потребує врахування багатьох факторів, включаючи обсяг даних, їхню якість та доступність обчислювальних ресурсів. Наприклад, у системах моніторингу мережевого трафіку алгоритми повинні працювати у режимі реального часу, що вимагає оптимізації їх продуктивності. У таких випадках дерева рішень можуть бути більш доцільними

завдяки їх швидкості, тоді як SVM краще підходять для офлайн-аналізу, де точність є критично важливою.

Моделі класифікації, такі як SVM та дерева рішень, є потужними інструментами для виявлення аномалій. Вони забезпечують високу точність та інтерпретованість, дозволяючи ідентифікувати підозрілі записи у великих наборах даних. Завдяки своїм перевагам ці алгоритми широко використовуються у різних сферах, зокрема у кібербезпеці, аналізі поведінки користувачів та прогнозуванні відмов.

Кластеризація є одним із ключових методів машинного навчання, що дозволяє групувати дані за схожістю характеристик. Це підхід, який не вимагає міток або розмітки даних, тому він належить до алгоритмів без учителя. У задачах виявлення аномалій кластеризація ефективно використовується для визначення незвичайних даних, які не належать до жодного із типових кластерів. Серед популярних алгоритмів кластеризації особливу увагу привертають k-means та DBSCAN, які мають свої переваги й особливості.

Алгоритм k-means базується на розподілі даних у простір за допомогою центроїдів, які є середніми точками для кожного кластеру. Його робота починається з визначення кількості кластерів, після чого дані розподіляються між ними на основі відстані до найближчого центроїда. Центроїди оновлюються після кожної ітерації, доки алгоритм не досягне стану, коли розподіл точок більше не змінюється. Завдяки своїй простоті та ефективності k-means часто використовується у великих наборах даних, таких як лог-файли DNS-запитів, мережевий трафік або поведінка користувачів у системах.

Для виявлення аномалій за допомогою k-means визначаються точки, які значно віддалені від центрів кластерів. Наприклад, якщо в межах кластера визначена середня відстань до центроїда, точки, що перевищують цю відстань на значну величину, можуть бути класифіковані як аномальні. У системах моніторингу мереж така техніка дозволяє виявляти запити, які не вписуються у типові патерни поведінки, наприклад, запити до незвичайних доменів або частотність звернень, що суттєво перевищує норму.

Перевага k-means полягає у його простоті та швидкості, особливо при роботі з великими наборами даних. Однак цей алгоритм має і певні обмеження. По-перше, його ефективність залежить від правильного вибору кількості кластерів, яка не завжди очевидна у реальних даних. По-друге, k-means менш ефективний для кластеризації даних, які мають складну або нерівномірну структуру, оскільки він передбачає сферичну форму кластерів. Це може призводити до ситуацій, коли дані з нерівномірною щільністю погано розподіляються між кластерами.

Для вирішення цих проблем у багатьох сценаріях використовують DBSCAN — алгоритм кластеризації, що базується на щільності. DBSCAN не вимагає заздалегідь визначати кількість кластерів, а замість цього використовує параметри мінімальної кількості точок у кластері та максимальної відстані між точками. Алгоритм починає з вибору випадкової точки, після чого визначає, чи входить вона до кластера, зважаючи на щільність навколишніх точок. Якщо точка відповідає критеріям щільності, вона включається до кластера, і цей процес повторюється для сусідніх точок.

DBSCAN особливо ефективний для виявлення аномалій, оскільки точки, які не входять до жодного кластера, автоматично класифікуються як шум або аномальні. Наприклад, у мережевому аналізі запити, що не відповідають звичайним шаблонам або мають низьку частоту, можуть бути визначені як аномалії. DBSCAN дозволяє ідентифікувати такі точки без необхідності задавати фіксовану кількість кластерів, що робить його більш універсальним у порівнянні з k-means.

Алгоритм DBSCAN також добре справляється з даними, які мають складну структуру або різну щільність. Наприклад, якщо у наборі даних є як густі, так і рідкісні зони, DBSCAN може адаптуватися до цього, утворюючи кластери лише там, де щільність перевищує встановлений поріг. Це особливо важливо у задачах виявлення аномалій, де аномальні точки часто знаходяться у рідкісних зонах або є ізольованими.

Недоліком DBSCAN є його чутливість до вибору параметрів. Неправильний вибір порога щільності може призвести до того, що алгоритм визначить занадто багато або занадто мало кластерів. Крім того, цей алгоритм може бути менш

ефективним для роботи з дуже великими наборами даних, оскільки його час виконання залежить від складності обчислення відстаней між точками.

Поєднання k-means і DBSCAN може забезпечити ще більш ефективні результати у задачах кластеризації та виявлення аномалій. Наприклад, k-means може бути використаний для попередньої кластеризації даних, після чого DBSCAN може уточнити результати, виявляючи ізольовані точки або підкласи у межах кластерів. Такий підхід дозволяє компенсувати недоліки кожного з алгоритмів, зберігаючи їх переваги.

Практичне застосування кластеризації у реальних системах включає моніторинг мережевого трафіку, аналіз поведінки користувачів та виявлення аномальних подій у великих базах даних. У мережевій безпеці кластеризація використовується для виявлення підозрілих активностей, таких як спроби вторгнень, сканування портів або ботнет-активність. Наприклад, якщо DBSCAN визначає групу запитів до незвичайного домену як окремий кластер, це може вказувати на те, що домен використовується у зловмисних цілях.

Алгоритми кластеризації також є невід'ємною частиною систем машинного навчання, які працюють у режимі реального часу. Наприклад, у системах моніторингу мережевих ресурсів алгоритми кластеризації можуть автоматично адаптуватися до нових типів трафіку, виявляючи аномалії без необхідності вручну оновлювати моделі. Це дозволяє знижувати ризики, пов'язані з появою нових загроз, і забезпечувати високу ефективність аналізу.

Кластеризація, зокрема методи k-means та DBSCAN, є потужним інструментом для виявлення аномалій у великих наборах даних. Завдяки своїм особливостям ці алгоритми забезпечують точний і гнучкий аналіз, який дозволяє ідентифікувати незвичайну поведінку навіть у складних і неоднорідних даних. Їх застосування допомагає підвищити безпеку систем та оптимізувати процес аналізу інформації, що робить кластеризацію важливим компонентом сучасних технологій машинного навчання.

2.6. Теоретичні основи аналізу аномалій

Попередня обробка даних є важливим етапом у підготовці до застосування алгоритмів машинного навчання. Від якості вихідних даних залежить точність, надійність і продуктивність моделей, що будуть побудовані. У контексті роботи з мережевими лог-файлами, такими як DNS-запити, попередня обробка дозволяє усунути недоліки, що виникають через наявність зайвих, пошкоджених або дублікатів даних. Цей етап є критично важливим для забезпечення ефективності аналізу та точності подальших прогнозів.

Одним із перших кроків попередньої обробки є видалення зайвих даних, зокрема дублікатів запитів. DNS-запити часто містять повторювані записи, які можуть виникати з різних причин. Наприклад, клієнтські пристрої або програми можуть надсилати однакові запити кілька разів через налаштування повторної передачі, якщо перший запит не отримав відповіді. Також дублікатами можуть бути запити, які були створені автоматизованими системами моніторингу для перевірки доступності доменів. Такі записи можуть штучно збільшувати обсяг даних, викривляючи результати аналізу. Видалення дублікатів є важливим завданням, яке дозволяє зменшити обсяг даних та уникнути зайвих обчислень під час навчання моделі.

У лог-файлах DNS-запитів дублікати зазвичай визначаються за комбінацією кількох характеристик, таких як час створення запиту, IP-адреса клієнта, тип запиту та домен, до якого звертається клієнт. Наприклад, якщо у логах є кілька записів, що стосуються одного й того ж домену, надісланих з однієї IP-адреси протягом короткого часу, вони можуть бути ідентифіковані як дублікати. Для цього зазвичай застосовуються алгоритми пошуку унікальних значень, які дозволяють швидко визначати повторювані записи та видаляти їх із набору даних.

Важливим аспектом є обробка ситуацій, коли дублікати можуть містити незначні відмінності, наприклад у форматі запису. У таких випадках застосовуються методи нормалізації, які дозволяють привести всі записи до єдиного формату. Наприклад, домени можуть бути записані як з префіксом "www",

так і без нього, або використовувати різні регістри символів. Нормалізація таких записів дозволяє зменшити кількість дублікатів, які залишилися б непоміченими при поверхневому аналізі.

Ще одним важливим кроком у попередній обробці даних є видалення неповних або пошкоджених записів. У логах DNS-запитів такі записи можуть виникати через технічні збої, помилки у записі даних або через шкідливу активність у мережі. Неповні записи, які не містять необхідної інформації, наприклад часу створення, IP-адреси клієнта або домену, зазвичай є непридатними для аналізу. Збереження таких записів у наборі даних може призвести до некоректної роботи алгоритмів, тому їх необхідно видаляти.

Видалення пошкоджених записів також може включати аналіз відповідності записів формату та стандартам DNS. Наприклад, якщо у логах містяться домени, які не відповідають стандартам формування імен або містять заборонені символи, такі записи слід вважати пошкодженими. Аналіз формату часто виконується за допомогою регулярних виразів або спеціалізованих програм, які дозволяють автоматично визначати відхилення від стандартів та позначати такі записи для видалення.

Однак не всі неповні записи необхідно видаляти. У деяких випадках інформація, яка відсутня у записах, може бути відновлена на основі інших даних. Наприклад, якщо у записі відсутня інформація про час створення, але є дані про IP-адресу клієнта та домен, час можна оцінити за допомогою контексту інших записів, що стосуються тієї ж IP-адреси. Такий підхід, відомий як імітація пропущених значень, дозволяє зберегти більше даних для аналізу, що особливо важливо у випадках, коли дані є обмеженими.

Видалення зайвих, неповних чи пошкоджених записів також є важливим з точки зору оптимізації роботи алгоритмів машинного навчання. Очищення даних дозволяє зменшити обсяг вхідних даних, що прискорює навчання моделі та знижує вимоги до обчислювальних ресурсів. Більш того, очищення даних підвищує точність моделі, оскільки усуває джерела шуму та неточностей, які могли б вплинути на результати.

Попередня обробка також включає створення протоколів перевірки якості даних, які дозволяють ідентифікувати та відстежувати потенційні проблеми у наборах даних. Наприклад, автоматизовані системи можуть виявляти аномальну кількість дублікатів або пошкоджених записів, що може свідчити про проблеми у системі збору даних. Такі протоколи допомагають зменшити кількість помилок на етапі обробки та покращують якість результатів аналізу.

Попередня обробка даних є критично важливою не лише для забезпечення точності моделей, але й для ефективного використання ресурсів. Видалення зайвих даних, таких як дублікати запитів, дозволяє знизити обсяг даних, що підлягають обробці, що зменшує час та витрати на обчислення. Видалення пошкоджених записів мінімізує ризик помилок у моделях, забезпечуючи їх точність та надійність.

У сучасних системах аналізу даних попередня обробка часто автоматизується за допомогою спеціалізованих програм або скриптів, які дозволяють швидко виконувати очищення даних навіть у великих наборах. Автоматизація також забезпечує стабільність процесу обробки та зменшує вплив людського фактору на якість підготовки даних.

Попередня обробка, що включає видалення зайвих, неповних чи пошкоджених даних, є невід'ємною частиною підготовки до застосування алгоритмів машинного навчання. Цей процес дозволяє забезпечити високу якість даних, підвищити точність аналізу та ефективно використовувати обчислювальні ресурси. У контексті роботи з лог-файлами DNS-запитів попередня обробка відіграє важливу роль у виявленні аномалій, забезпечуючи чистоту та достовірність вхідних даних для подальшого аналізу.

2.7.Висновки до розділу

У цьому розділі було розглянуто ключові особливості та характеристики аномалій DNS-запитів, а також їх вплив на роботу мережевої інфраструктури та системи безпеки. Детальний аналіз аномалій дозволив визначити основні категорії відхилень, які виникають у процесі передачі DNS-запитів, включно із

синтаксичними порушеннями, змінами у частоті та інтенсивності запитів, а також поведінковими відхиленнями. Кожна із зазначених категорій аномалій потребує окремого підходу до виявлення та обробки, що обумовлено їх природою та характеристиками. Синтаксичні аномалії виникають через порушення стандартів форматування DNS-запитів, що може бути спричинено як технічними збоями, так і зловмисною активністю. Частотні та інтенсивні аномалії пов'язані зі збільшенням кількості запитів до окремих доменів або серверів, що може вказувати на помилки у конфігурації пристроїв, активність автоматизованих систем або спроби атак на мережеві ресурси. Поведінкові аномалії характеризуються змінами у стандартних патернах активності клієнтів, які можуть свідчити про зараження пристроїв шкідливим програмним забезпеченням або інші несанкціоновані дії.

Ретельний аналіз аномалій у DNS-запитах є критично важливим завданням для забезпечення стабільності та безпеки сучасних інформаційних систем. Саме тому особливу увагу приділено застосуванню різноманітних методів аналізу для виявлення та класифікації таких аномалій. У розділі акцентовано на важливості використання як статистичних моделей, так і методів машинного навчання для ідентифікації нетипових запитів у мережі. Статистичні моделі дозволяють визначати відхилення від базових ліній активності на основі аналізу розподілів даних та їхніх математичних характеристик. Зокрема, застосування Гауссового та Пуассонівського розподілів дає змогу ефективно аналізувати частоту, інтенсивність та розподіл DNS-запитів для виявлення аномальних значень. Ці підходи забезпечують точність та надійність виявлення, але мають обмеження у випадках складних або нелінійних залежностей між характеристиками запитів.

Важливою складовою системи аналізу DNS-запитів є використання моделей машинного навчання, які дозволяють автоматично виявляти аномалії на основі обробки великих обсягів даних. У цьому контексті особливу увагу приділено алгоритмам класифікації, таким як SVM, деревам рішень, а також методам кластеризації, зокрема K-means та DBSCAN. Моделі машинного навчання забезпечують високу точність та гнучкість у виявленні аномалій, оскільки здатні враховувати складні залежності та взаємозв’язки між параметрами DNS-запитів.

Застосування кластеризації дозволяє групувати запити за схожістю характеристик, що дає змогу визначати незвичайні або ізольовані точки, які не відповідають типовим патернам активності. Такі підходи є особливо ефективними у сценаріях, де аномалії виникають рідко або мають прихований характер.

Окреме місце у розділі відведено важливості попередньої обробки даних як критичного етапу для підготовки до аналізу. Процес очищення даних, видалення дублікатів, заповнення пропущених значень та нормалізації форматів забезпечує високу якість вхідних даних для алгоритмів виявлення аномалій. Саме на цьому етапі усуваються технічні недоліки та шуми, що можуть спотворити результати аналізу або знизити ефективність моделей.

Загалом результати дослідження підкреслюють важливість комплексного підходу до аналізу DNS-запитів, що поєднує методи статистики, машинного навчання та детальну обробку даних. Такий підхід дозволяє ефективно ідентифікувати різні типи аномалій, мінімізувати кількість помилкових спрацювань та підвищити рівень безпеки мережевої інфраструктури. Однак слід зазначити, що ефективність застосованих методів залежить від якості вхідних даних, налаштувань моделей та здатності системи адаптуватися до динамічних змін у поведінці користувачів. Використання автоматизованих інструментів моніторингу та аналізу у поєднанні з моделями машинного навчання є ключовим фактором для забезпечення стабільності та безпеки мережевих ресурсів у сучасному цифровому середовищі.

Таким чином, розділ демонструє, що виявлення аномалій у DNS-запитах є важливим завданням, вирішення якого потребує застосування ефективних, адаптивних та надійних методів аналізу. Впровадження систем, що базуються на сучасних алгоритмах та інструментах, дозволяє забезпечити своєчасну ідентифікацію загроз, оптимізувати роботу мережевих інфраструктур та зменшити вплив потенційних ризиків на їхню продуктивність.

3. РОЗРОБКА МЕТОДУ ВИЯВЛЕННЯ АНОМАЛІЙ У DNS-ЗАПИТАХ

3.1.Структурна модель методу виявлення аномалій у DNS-запитах

Представимо структурну модель методу виявлення аномалій у DNS-запитах на рис. 3.1.

Рисунок 3.1 Структурна модель методу виявлення аномалій у DNS-запитах

Метод виявлення аномалій у DNS-запитах базується на етапному процесі, що охоплює збір даних, їх підготовку, побудову профілю нормальної активності та подальше виявлення відхилень для ідентифікації аномалій. Кожен етап виконує чітко визначені функції і є частиною єдиної системи, що аналізує DNS-трафік для визначення потенційно підозрілої активності, яка може вказувати на загрози чи порушення безпеки.

Першим етапом є збір даних, що передбачає отримання великого обсягу DNS-запитів від клієнтів мережі. Джерелами даних можуть слугувати DNS-сервери, логи запитів чи мережеві датчики, які фіксують кожен DNS-запит, його параметри, IP-адреси клієнтів і час виконання. Зібрані дані є основою для подальших етапів аналізу. У цьому процесі важливо враховувати як запити з

нормального трафіку, так і ті, що можуть містити аномалії, щоб не втратити критичну інформацію для подальшого аналізу.

Після збору даних виконується їх підготовка, яка включає кілька підпроцесів обробки інформації. На цьому етапі дані очищуються від шумів і непотрібних записів, які не мають значущості для аналізу. Також проводиться нормалізація даних для забезпечення їх однорідності та форматування у відповідність до встановлених вимог моделі. Підготовка даних включає виділення ключових характеристик DNS-запитів, таких як доменні імена, частота звернень, тривалість запитів і типи використовуваних записів (A, AAAA, MX, TXT та інші). Ці параметри є критичними для створення профілю нормальної активності та визначення критеріїв аномалій.

На третьому етапі формується профіль нормальної активності, що є репрезентативною моделлю поведінки DNS-запитів у звичайному стані мережі. Побудова такого профілю виконується на основі аналізу великих обсягів зібраних та підготовлених даних, які дозволяють визначити типові закономірності, частоти й часові патерни. Профіль нормальної активності фіксує регулярну поведінку DNS-клієнтів, включаючи середню кількість запитів за одиницю часу, стандартні значення параметрів запитів, а також очікувану структуру доменних імен. Для цього застосовуються статистичні методи, машинне навчання або інші алгоритмічні підходи. Мета профілю полягає у створенні еталонного середовища для виявлення аномалій, що відхиляються від нормальної активності.

Наступний етап – виявлення аномалій, що полягає у порівнянні реальних DNS-запитів із профілем нормальної активності. Аномалії виявляються тоді, коли зафіксовані параметри запитів виходять за межі допустимих значень, визначених на основі побудованого профілю. У процесі виявлення можуть використовуватися різні методи аналізу, зокрема статистичні підходи для обчислення відхилень, а також алгоритми машинного навчання для ідентифікації шаблонів, які не відповідають нормі. Параметри аномалій можуть включати надмірну частоту DNS-запитів за короткий проміжок часу, підозріло довгі або випадкові доменні імена, нехарактерні типи записів чи інші нетипові ознаки запитів.

Завершальним етапом є ідентифікація аномалій, які можуть вказувати на потенційні загрози, такі як DNS-атаки, спроби витоку даних, шкідливе програмне забезпечення чи несанкціоноване використання ресурсів мережі. Виявлені аномалії підлягають додатковому аналізу для підтвердження їх дійсної природи та критичності. Результати цього етапу можуть бути передані системам реагування на інциденти або використані для підвищення безпеки мережі шляхом вдосконалення політик доступу та конфігурацій DNS-серверів.

Таким чином, структурна модель методу виявлення аномалій у DNS-запитах є послідовною системою, що охоплює процеси збору, підготовки та аналізу даних з метою виявлення відхилень від нормальної активності. Кожен етап є важливим для забезпечення точності та ефективності виявлення потенційних загроз у DNS-трафіку.

3.2.Збір даних з DNS-запитів

Збір даних передбачає отримання вхідної інформації, її обробку та видачу результату (Рис. 3.2).

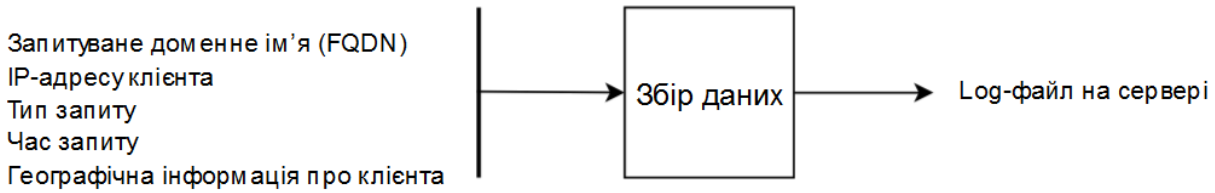

Рисунок 3.2 Дані, отримувані з DNS запиту

Послідовність збору даних доцільно представити у вигляді кроків.

Крок 1. Налаштування маршрутизатора для передачі DNS-запитів. На маршрутизаторі активується функція передачі логів DNS-запитів через протокол Syslog. У налаштуваннях маршрутизатора вказується IP-адреса Syslog-сервера (наприклад 192.168.1.100), а також заданий стандартний порт для передачі логів

(514). Усі DNS-запити, що проходять через маршрутизатор, записуються до журналу і пересилаються у вигляді Syslog-повідомлень.

Крок 2. Встановлення програмного забезпечення на сервері. На окремому сервері в мережі встановлено програму rsyslog, яка налаштована на прийом Syslog-повідомлень від маршрутизатора. Конфігураційний файл rsyslog.conf відредаговано так, щоб усі вхідні повідомлення з маршрутизатора записувалися у файл /var/log/dns_queries.log.

Крок 3. Запуск Syslog-сервера. Syslog-сервер на сервері активовано, і він працює у режимі реального часу, приймаючи повідомлення від маршрутизатора. Усі DNS-запити, передані через Syslog, негайно записуються у файл логів. Це дозволяє зберігати хронологічний запис усіх запитів із зазначенням часу, IP-адрес клієнтів, доменів і типів запитів.

Крок 4. Первинна фільтрація даних на маршрутизаторі. На маршрутизаторі налаштоване правило фільтрації, яке дозволяє передавати на сервер лише DNS-запити, що використовують протокол UDP і порт 53. Логи, що не стосуються DNS-запитів, не передаються на сервер, зменшуючи обсяг даних, які потребують обробки.

Крок 5. Захист логів. Логи, що зберігаються на сервері, захищені за допомогою налаштування прав доступу. Тільки адміністратори мають право переглядати або змінювати файли у директорії /var/log та /data. Також налаштовано регулярне резервне копіювання логів у віддалене сховище через шифрований канал за допомогою rsync.

Крок 6. Моніторинг збору даних. На сервері встановлено програму моніторингу Zabbix, яка відстежує стан Syslog-сервера та доступність лог-файлу. У разі виникнення помилок або збоїв (наприклад, припинення передачі логів від маршрутизатора), адміністратор негайно отримує сповіщення електронною поштою.

Крок 7. Видалення старих даних. Для уникнення перевантаження диска налаштовано автоматичне видалення старих логів. Файли зберігаються протягом

30 днів, після чого вони автоматично видаляються за допомогою системного планувальника cron.

Дана послідовність кроків дозволяє отримати з маршрутизатора DNS-запити та направити їх на сервер для подальшої обробки.

3.3.Підготовка даних DNS-запитів

Етап підготовки даних з DNS-запитів забезпечує їх очищення, нормалізацію та форматування для подальшого аналізу. Вхідні та вихідні дані цього етапу представлені на Рис. 3.3.

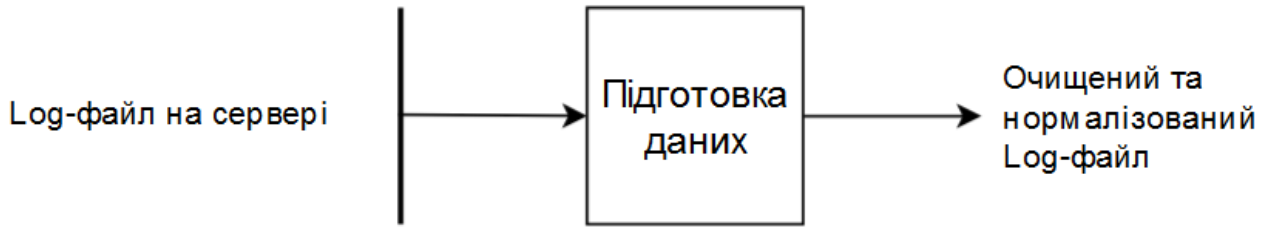

Рисунок 3.3 Підготовка даних

Після етапу збору даних, описаного вище, необхідно виконати такі кроки:

Крок 1. Читання лог-файлів із сервера. Зібрані DNS-запити, що зберігаються у файлі /var/log/dns_queries.log, зчитуються для подальшої обробки. На сервері виконується скрипт на Python, який відкриває файл у режимі читання та зчитує дані у вигляді рядків. Кожен рядок представляє окремий запис із полями, що включають час запиту, IP-адресу клієнта, тип запиту, запитуваний домен і результат відповіді сервера.

Крок 2. Видалення пошкоджених записів. На цьому етапі виконується перевірка кожного запису на повноту та коректність. Записи, що не містять одного або кількох обов’язкових полів (наприклад, часу запиту чи доменного імені), вважаються пошкодженими і видаляються з набору даних. Перевірка здійснюється за допомогою скрипта, який аналізує кожен рядок, виявляє відсутність необхідних даних та виключає такі записи із загального масиву.

Крок 3. Видалення дублікатів. Дублікати DNS-запитів, які можуть виникати через повторні звернення клієнтів або помилки системи, видаляються. Дублікати визначаються як записи, що мають однакові значення для полів часу, IP-адреси клієнта, типу запиту та домену. Для цього дані перетворюються у структуру типу словника, де кожен унікальний запис зберігається як ключ, а дублікати ігноруються.

Крок 4. Нормалізація даних. Всі доменні імена у записах нормалізуються для забезпечення єдиного формату. Домени переводяться у нижній регістр, видаляються зайві пробіли, а домени з префіксом www зводяться до базової форми. Це дозволяє уникнути розбіжностей між записами, які стосуються одного і того ж домену, але мають різний вигляд.

Крок 5. Форматування дат і часу. Час запитів конвертується у стандартний формат, наприклад ISO 8601, для забезпечення уніфікованого представлення. Записи з нестандартним або пошкодженим форматом часу виключаються. Усі часи синхронізуються з єдиним часовим поясом, щоб уникнути помилок під час аналізу.

Крок 6. Додавання геолокаційної інформації. На основі IP-адрес клієнтів до записів додається інформація про географічне розташування. Для цього використовується локальна база GeoIP, яка зіставляє IP-адреси з країною, містом та провайдером. Додавання геолокаційних даних дозволяє виявляти регіональні аномалії у DNS-запитах.

Крок 7. Конвертація у зручний формат для аналізу. Після очищення та нормалізації дані конвертуються у формат CSV для подальшого аналізу. Кожен рядок CSV-файлу відповідає одному запису DNS-запиту, а стовпці включають такі поля: час, IP-адреса клієнта, тип запиту, домен, геолокація та статус відповіді. Файл зберігається у директорії /data під назвою dns_queries_cleaned.csv.

Крок 8. Створення резервної копії очищених даних. Підготовлений CSV-файл дублюється у резервному сховищі для забезпечення безпеки. Використовується програма rsync для створення копії файлу у віддаленій директорії, доступ до якої обмежений. Це дозволяє захистити дані від втрати через апаратні або програмні збої.

Крок 9. Перевірка якості даних. На завершальному етапі виконуються додаткові перевірки якості даних. Скрипт аналізує результуючий файл на наявність аномалій, таких як відсутність полів, невідповідність формату або неочікувані значення. Якщо такі проблеми виявлено, система генерує звіт і надсилає його адміністратору для виправлення.

Ця послідовність забезпечує високу якість і надійність даних, що дозволяє виконувати точний аналіз та виявляти аномалії у DNS-запитах.

3.4.Профіль нормальної активності

Профіль нормальної активності є основою для подальшого порівняння із реальними DNS-запитами. Вони дозволяють визначати відхилення від стандартної поведінки, що вказують на потенційні аномалії або загрози, такі як DDoS-атаки, використання DGA-доменів чи інші зловмисні дії.

Набір даних, що генеруються профілем нормальної активності представлено на Рис. 3.4.

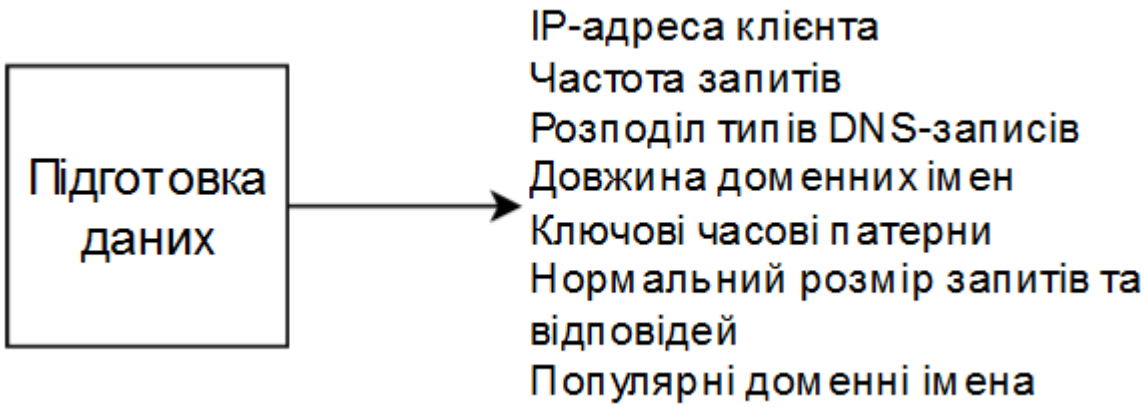

Рисунок 3.4 Дані, що генеруються профілем нормальної активності

Послідовність побудови профілю нормальної активності можна представити у вигляді кроків:

Крок 1. Збір DNS-трафіку. На першому кроці відбувається накопичення великого обсягу DNS-запитів, які надходять від клієнтів мережі. Дані збираються

за допомогою логів DNS-серверів, мережевих моніторинг-систем або спеціальних датчиків. Записуються ключові параметри DNS-запитів, такі як час запиту, IP-адреса клієнта, доменне ім'я, тип DNS-запису, розмір запиту та відповідей.

Крок 2. Попередня обробка та очищення даних. Зібрані DNS-дані проходять фільтрацію для усунення неповних, помилкових або дубльованих записів Як і на етапі підготовки даних.

Крок 3. Нормалізація даних. На цьому етапі всі дані приводяться до узгодженого формату. Поля лог-файлів уніфікуються (назви колонок, порядок значень). Довжина і структура доменних імен аналізується й фіксується у вигляді числових значень або категорій. IP-адреси клієнтів групуються для подальшого аналізу.

Крок 4. Агрегація характеристик DNS-запитів. Виконується обчислення основних характеристик DNS-запитів для різних інтервалів часу. Частота запитів: середня кількість запитів на клієнта за хвилину, годину або добу. Розподіл типів DNS-записів: обчислення відсоткового співвідношення запитів (A, AAAA, MX тощо). Популярність доменів (виділення доменів, до яких найчастіше звертаються клієнти). Розмір запитів: середній і максимальний розмір DNS-запитів і відповідей у байтах.

Крок 5. Аналіз часових патернів. Виконується статистичний аналіз часових закономірностей DNS-трафіку. Визначення пікових періодів активності запитів (наприклад, з 8:00 до 18:00). Аналіз спадів активності у неробочий час (наприклад, нічний період). Обчислення середнього інтервалу між запитами для кожного клієнта.

Крок 6. Побудова статистичних моделей

На основі отриманих характеристик формується статистична модель нормальної активності. Граничні значення для кожного параметра (наприклад, максимальна кількість запитів на клієнта у нормі). Розподіли значень (середнє, медіана, стандартне відхилення) для параметрів трафіку. Шаблони поведінки клієнтів на основі типових значень частоти й структури DNS-запитів.

Формування профілю нормальної активності виконується на основі попередніх кроків. Узагальнений профіль нормальної активності включає:

- Середні та граничні значення для частоти DNS-запитів на IP-адресу.
- Нормальний розподіл типів DNS-записів у звичайному трафіку.
- Характеристики доменних імен (довжина, структура, кількість піддоменів).
- Типові часові патерни трафіку.
- Географічні особливості (локалізація клієнтів, що генерують DNS-запити).

Крок 7. Валідація профілю нормальної активності. Завершальним кроком є перевірка та уточнення побудованого профілю. Профіль тестується на реальному трафіку для оцінки його відповідності нормальній поведінці клієнтів. Виявляються можливі помилкові відхилення для уточнення граничних значень. Профіль оптимізується на основі аналізу нових даних або поведінкових змін у DNS-трафіку.

3.5.Виявлення аномалій

Блок виявлення аномалій використовує машинне навчання для ідентифікації відхилень у DNS-запитах шляхом порівняння їх з профілем нормальної активності. Основне завдання полягає у навчанні моделей машинного навчання автоматично розпізнавати аномальні DNS-запити, які не відповідають сформованим характеристикам нормального трафіку. Оскільки аномальні події зазвичай є рідкісними і їх важко заздалегідь розмітити, для вирішення цієї задачі застосовуються неконтрольовані методи машинного навчання.

У блоці виявлення аномалій використовуються три основні методи машинного навчання:

- Isolation Forest
- One-Class SVM (Support Vector Machine)
- K-means

Метод Isolation Forest є одним з основних алгоритмів, що використовується для виявлення аномалій у DNS-запитах. Його ключовою особливістю є можливість ефективно працювати з великими обсягами даних завдяки побудові дерев ізоляції.

Алгоритм базується на принципі, згідно з яким аномальні точки в даних швидше ізолюються від основного масиву, оскільки вони знаходяться відокремлено у багатовимірному просторі. У процесі побудови моделі Isolation Forest використовує випадкові розділення даних за їхніми характеристиками для формування дерев. Кожне дерево є ієрархічною структурою, де вузли розділяють простір даних на підпростори за ознаками, такими як частота DNS-запитів, довжина доменних імен або час між запитами.

На початковому етапі роботи моделі Isolation Forest аналізує навчальні дані, що включають нормальні DNS-запити, і будує деревоподібну структуру на основі їхніх характеристик. Нормальні точки потребують більшої кількості розділень, щоб бути ізольованими, оскільки вони утворюють щільні групи у просторі ознак. На відміну від них аномальні DNS-запити швидко ізолюються на початкових етапах розділення, оскільки вони мають значні відхилення у своїх характеристиках. Кожному запиту присвоюється так званий аномальний бал, який відображає ступінь його ізольованості. Чим менше розділень потрібно для ізоляції точки, тим вищим є її аномальний бал, що вказує на її потенційну аномальність.

Перевага цього методу полягає у його швидкодії, оскільки час обчислення має логарифмічну складність, що дозволяє обробляти великі обсяги DNS-запитів за короткий проміжок часу. Крім того, Isolation Forest не потребує попереднього маркування даних і працює безпосередньо з неконтрольованим набором, що робить його особливо корисним для задач виявлення аномалій у реальному трафіку. Однак певною складністю є чутливість моделі до параметра contamination, який визначає очікуваний відсоток аномалій у наборі даних. Неправильний вибір цього параметра може призвести до помилкових спрацювань або пропуску реальних аномалій.

У блоці виявлення аномалій Isolation Forest застосовується для присвоєння аномального бала кожному DNS-запиту. Модель аналізує характеристики трафіку, такі як частота запитів, структура доменів та інтервали між запитами, і визначає точки, що найменше відповідають профілю нормальної активності. Запити з високим аномальним балом передаються на подальший аналіз для підтвердження їхньої аномальності або усунення можливих помилкових результатів. Таким

чином, Isolation Forest є ефективним інструментом для первинного виявлення аномалій у DNS-запитах завдяки своїй здатності швидко ізолювати нетипові точки у просторі характеристик.

Метод One-Class SVM є методом, що використовується для розв'язання задачі виявлення аномалій за допомогою побудови гіперплощини у багатовимірному просторі ознак. На відміну від класичних методів машинного навчання, One-Class SVM працює виключно з нормальними даними, використовуючи їх як основу для формування області, що охоплює "ядро" нормальних точок. Алгоритм побудови моделі базується на гіпотезі, що аномалії розташовуються поза межами цієї області, оскільки вони суттєво відрізняються від нормальних зразків за своїми характеристиками.

Під час навчання моделі One-Class SVM аналізує вектори характеристик DNS-запитів, таких як частота звернень до серверів, типи DNS-записів, структура доменних імен і часові патерни. За допомогою ядерної функції, найчастіше радіальної базисної функції (RBF), алгоритм відображає дані у простір більшої вимірності, де будується гіперплощина, що відокремлює нормальні точки від усіх інших. Під час роботи моделі нові DNS-запити перевіряються на належність до цієї області, і якщо вони виходять за межі гіперплощини, вони позначаються як аномальні.

Головною перевагою One-Class SVM є його здатність працювати з нелінійними залежностями у даних, що робить його ефективним для аналізу складних характеристик DNS-запитів. У порівнянні з іншими методами, цей алгоритм забезпечує високу точність при роботі з невеликими та середніми наборами даних. Водночас він має певні обмеження, зокрема високу обчислювальну складність, що може створювати труднощі при аналізі великих обсягів трафіку. Крім того, вибір параметра nu, який визначає частку аномальних точок, є критичним для точності моделі.

У блоці виявлення аномалій One-Class SVM використовується для точного визначення відхилень від профілю нормальної активності. Запити, що мають відхилення від ключових характеристик, таких як частота запитів або структура

доменів, автоматично відокремлюються від нормальних точок і виносяться на подальший аналіз. Таким чином, One-Class SVM є потужним інструментом для виявлення складних аномалій у DNS-запитах, які не можуть бути виявлені простими статистичними методами.

Метод K-means є одним із найпопулярніших алгоритмів для кластеризації даних, що застосовується для виявлення аномалій шляхом групування DNS-запитів у кілька кластерів. Основна ідея алгоритму полягає у тому, що нормальні точки даних утворюють щільні групи (кластери), тоді як аномалії зазвичай відокремлені та мають велику відстань до центрів кластерів.

У процесі навчання K-means намагається мінімізувати відстань між точками даних та центроїдами кластерів, ітеративно перебудовуючи центри до стабілізації результату. У випадку DNS-запитів алгоритм аналізує ключові характеристики, такі як частота звернень до серверів, типи DNS-запитів, довжина доменних імен та часові інтервали між запитами. Дані групуються у кілька кластерів, і запити, що потрапляють у малі або віддалені кластери, позначаються як потенційно аномальні.

Перевагою K-means є його швидкодія та здатність ефективно обробляти великі обсяги даних. Алгоритм легко реалізується і забезпечує достатню точність при виявленні аномалій у структурованих даних. Водночас K-means має певні обмеження, зокрема необхідність попереднього визначення кількості кластерів K та слабку роботу з даними, що мають складні нелінійні залежності.

У блоці виявлення аномалій K-means використовується для групування DNS-запитів на основі їхніх характеристик. Запити, що суттєво відхиляються від центрів основних кластерів або потрапляють у малі кластери, автоматично позначаються як аномальні для подальшого аналізу. Цей метод дозволяє ідентифікувати аномалії, які мають схожі характеристики, але не відповідають типовій поведінці DNS-трафіку.

Представимо процес виявлення аномалій послідовністю кроків.

Крок 1. Ініціалізація моделей та завантаження підготовлених даних. На цьому кроці моделі машинного навчання Isolation Forest, One-Class SVM та K-means ініціалізуються для подальшої обробки даних. Завантажені підготовлені дані DNS-

запитів одразу подаються у вигляді числових векторів ознак, сформованих під час етапу попередньої обробки. Ініціалізація моделей включає налаштування основних параметрів, які визначають їх поведінку під час обчислень. Для Isolation Forest встановлюється параметр contamination, що визначає частку очікуваних аномалій у даних. У моделі One-Class SVM налаштовується параметр nu, що контролює кількість точок, які можуть бути позначені як аномальні. Для алгоритму K-means визначається кількість кластерів, яка базується на аналітичних висновках попередніх етапів або задається емпірично.

Цей крок також включає перевірку, чи дані відповідають вимогам моделей, зокрема їх вимірність та відсутність невизначених значень. Моделі машинного навчання готові до подальшої обробки вхідних векторизованих даних.

Крок 2. Обчислення аномальних балів моделлю Isolation Forest. На цьому етапі Isolation Forest аналізує вхідні DNS-запити, поступово розділяючи їх у деревоподібній структурі. Алгоритм випадково вибирає ознаки, такі як частота запитів, довжина доменів та часові інтервали, для створення розділень даних на рівнях дерева. Під час обчислення модель присвоює кожному запиту аномальний бал на основі глибини, з якої він був ізольований у дереві. Запити, що швидко ізолюються на ранніх рівнях, отримують вищі бали аномальності, оскільки вони суттєво відхиляються від основної маси точок.

Модель виконує багаторазову побудову дерев для підвищення стабільності результатів. Після цього формується сумарний аномальний бал для кожного DNS-запиту. На виході цього кроку формується проміжний результат, який містить список DNS-запитів із їхніми аномальними балами. Запити з балами, що перевищують заданий поріг, позначаються як потенційно аномальні та переходять на подальший аналіз.

Крок 3. Виявлення відхилень за допомогою One-Class SVM. На цьому етапі модель One-Class SVM аналізує ті ж підготовлені дані, застосовуючи метод побудови гіперплощини у багатовимірному просторі ознак. Всі вхідні DNS-запити перевіряються на їх належність до "ядра нормальних даних", сформованого на етапі навчання. Запити, що виходять за межі цієї області, позначаються як аномальні.

One-Class SVM потребує попереднього масштабування ознак для забезпечення коректної роботи моделі, що було виконано під час попереднього етапу підготовки даних. Модель обчислює відстань кожного запиту до гіперплощини та позначає точки з максимальною відстанню як аномальні. На цьому кроці формується додатковий список DNS-запитів, які мають значні відхилення відповідно до результатів One-Class SVM.

Крок 4. Кластеризація даних алгоритмом K-means для виявлення аномалій. На четвертому кроці виконується кластеризація DNS-запитів з використанням методу K-means. Алгоритм групує дані у декілька кластерів на основі схожості ознак, таких як частота запитів, довжина доменів і часові характеристики. Центроїди кластерів обчислюються ітеративно для мінімізації відстані між точками та центром кластеру.

Після завершення кластеризації алгоритм аналізує розподіл DNS-запитів у кластерах. Точки, що належать до малих кластерів, а також точки з великою відстанню до центроїда основного кластера, позначаються як аномальні. Цей підхід дозволяє ідентифікувати запити, які не відповідають типовим групам поведінки. Результати кластеризації додаються до загального списку потенційно аномальних DNS-запитів.

Крок 5. Об'єднання та фільтрація результатів усіх методів. На цьому етапі результати, отримані від моделей Isolation Forest, One-Class SVM та K-means, об'єднуються для формування єдиного списку аномальних DNS-запитів. Запити, які були позначені як аномальні хоча б однією моделлю, піддаються додатковій фільтрації. Для цього використовуються порогові значення аномальних балів та евристичні правила, що дозволяють відсіювати потенційно помилкові спрацювання.

DNS-запити, які були позначені аномальними двома чи трьома методами одночасно, отримують вищий пріоритет для подальшого аналізу. Таким чином, комбінування результатів моделей забезпечує підвищену надійність та зниження рівня помилкових спрацювань.

Крок 6. Формування звіту про виявлені аномалії. Завершальним кроком є формування узагальненого звіту про виявлені аномалії у DNS-запитах. Для кожного запиту, позначеного як аномальний, надається повна інформація, включаючи його IP-адресу, часову мітку, тип запиту, довжину домену та аномальний бал, присвоєний моделями. Запити сортуються за пріоритетом, де на вершині списку розташовуються ті, які отримали найвищі оцінки аномальності.

3.6.Метод виявлення аномалій у DNS-запитах

Метод виявлення аномалій у DNS-запитах є комплексною системою, що базується на послідовному аналізі та зіставленні вхідного трафіку з профілем нормальної активності для ідентифікації потенційно підозрілих відхилень. Основою цього методу є застосування неконтрольованих методів машинного навчання, що дозволяють виявляти аномалії без потреби в попередньо розмічених даних. Вхідні DNS-запити обробляються та порівнюються зі сформованими характеристиками звичайної активності, що дає змогу ідентифікувати випадки, які виходять за межі нормальних поведінкових патернів.

Ефективність методу полягає у можливості виявляти нетипові запити, що вказують на аномальні ситуації, як-от високі частоти запитів від одного клієнта, використання підозрілих або згенерованих доменів, а також нехарактерні часові патерни активності. Для досягнення цієї мети метод включає етапи підготовки даних, побудови профілю нормальної активності та застосування моделей машинного навчання для подальшого виявлення відхилень. На етапі підготовки даних відбувається очищення, нормалізація та стандартизація вхідних лог-файлів DNS-запитів. Після цього на основі статистичних та часових характеристик формується профіль нормальної активності, який є референсною моделлю стандартної поведінки у мережі.

У процесі визначення аномалій ключовими є алгоритми неконтрольованого навчання, що працюють з вихідними ознаками, сформованими із DNS-запитів. Зокрема, застосовуються три моделі: Isolation Forest, One-Class SVM та K-means.

Кожна з цих моделей має унікальні особливості, що забезпечують ефективне виявлення різних типів аномалій. Isolation Forest ізолює аномальні запити завдяки побудові дерев, що швидко розділяють віддалені точки у багатовимірному просторі. Модель One-Class SVM визначає гіперплощину, що охоплює нормальні дані, і відокремлює запити, які виходять за її межі. Кластеризація методом K-means групує запити у кілька кластерів, де аномалії визначаються як точки, що не належать до основних груп або знаходяться на значній відстані від центроїдів.

Обробка даних відбувається паралельно у всіх моделях, після чого результати їхньої роботи комбінуються. Запити, які були позначені як аномальні двома або більше методами, отримують вищий рівень пріоритету для подальшого аналізу. Такий підхід дозволяє підвищити точність виявлення аномалій та мінімізувати кількість помилкових спрацювань. Об’єднання результатів кількох моделей забезпечує стійкість системи до недоліків кожного окремого методу та дозволяє ефективно працювати з великими обсягами мережевого трафіку.

Потенційними перевагами методу виявлення аномалій DNS-запитів є його здатність працювати у реальному часі та автоматично адаптуватися до змін у трафіку завдяки використанню машинного навчання. Універсальність алгоритмів дозволяє їх застосовувати для аналізу як невеликих, так і масштабних мережевих сегментів без значного зниження продуктивності. Додатковою перевагою є можливість виявляти складні аномалії, які неможливо ідентифікувати за допомогою традиційних статистичних методів, наприклад, генерацію випадкових доменів (DGA) або нетипову активність бот-мереж.

Попри значні переваги метод має і певні недоліки, які слід враховувати під час його використання. Однією з основних проблем є необхідність точного налаштування гіперпараметрів моделей машинного навчання, таких як кількість кластерів у K-means або рівень contamination в Isolation Forest. Некоректне налаштування може призвести до помилкових спрацювань або втрати важливих аномалій. Крім того, метод має обмежену ефективність у разі зміни поведінкових патернів трафіку, що вимагає регулярного оновлення профілю нормальної активності. Великий обсяг обчислень, особливо для One-Class SVM, може

створювати додаткове навантаження на обчислювальні ресурси у разі аналізу значних обсягів даних у реальному часі.

Загалом метод виявлення аномалій DNS-запитів є ефективним інструментом для ідентифікації потенційних загроз у мережевому трафіку. Завдяки поєднанню кількох моделей машинного навчання він здатен виявляти широкий спектр аномалій, що виходять за межі нормальної поведінки клієнтів. Комбінування результатів забезпечує високу точність та надійність системи навіть у випадках, коли одна з моделей не може чітко визначити аномальні точки. Цей підхід є гнучким та адаптивним до змін у мережі, що робить його корисним для моніторингу сучасного динамічного трафіку DNS-запитів.

3.7.Висновок до розділу

У цьому розділі було детально розглянуто метод виявлення аномалій у DNS-запитах, що базується на аналізі мережевого трафіку з використанням неконтрольованих методів машинного навчання. Основною метою такого підходу є ідентифікація відхилень від профілю нормальної активності, який попередньо формується на основі аналізу статистичних характеристик DNS-запитів у звичайному режимі роботи мережі. Важливість цього методу обумовлена необхідністю виявлення потенційних загроз та підозрілих активностей, які можуть бути ознакою атак, витоків даних або роботи шкідливих програм у мережі.

Метод побудований на використанні трьох основних моделей машинного навчання, а саме Isolation Forest, One-Class SVM та K-means, які дозволяють аналізувати DNS-запити у різних аспектах. Isolation Forest забезпечує швидку ізоляцію аномальних запитів шляхом побудови дерев, що розділяють дані на основі випадково обраних ознак. Цей підхід дозволяє ефективно визначати віддалені точки у багатовимірному просторі характеристик DNS-запитів. Модель One-Class SVM використовує концепцію побудови гіперплощини, що охоплює ядро нормальних даних та дозволяє визначати точки, які виходять за межі цієї області. Такий підхід ефективний у випадках, коли аномалії мають незначні, але суттєві

відхилення від нормального профілю. Алгоритм K-means, у свою чергу, здійснює кластеризацію даних та ідентифікує аномалії на основі їх віддаленості від основних груп точок або приналежності до малих кластерів. Поєднання результатів цих трьох моделей забезпечує більш точне та надійне визначення аномалій у DNS-трафіку.

Процес виявлення аномалій реалізується у кілька етапів, що включають завантаження підготовлених даних, ініціалізацію моделей, обробку запитів та об'єднання результатів для формування остаточного списку аномальних точок. Кожна модель працює паралельно над одним і тим самим набором даних, після чого результати обробки об'єднуються на основі певних критеріїв, що забезпечують консенсус між методами. Запити вважаються аномальними у випадку, якщо їх відхилення підтверджуються щонайменше двома моделями. Цей підхід дозволяє підвищити стійкість методу до помилкових спрацювань і забезпечує кращу точність виявлення реальних загроз.

Застосування даного методу має низку значних переваг, що роблять його ефективним інструментом для моніторингу DNS-запитів. Використання машинного навчання дозволяє автоматично адаптувати систему до змін у трафіку та виявляти складні аномалії, які не можуть бути ідентифіковані за допомогою традиційних методів аналізу. Метод здатний працювати з великими обсягами даних у режимі реального часу, що є критично важливим для сучасних мережевих інфраструктур. Завдяки комбінуванню результатів кількох моделей забезпечується підвищена надійність та точність системи, що дозволяє мінімізувати кількість помилкових спрацювань.

Разом з тим метод має певні недоліки, які потребують врахування у процесі його використання. Зокрема, він вимагає регулярного оновлення профілю нормальної активності для забезпечення актуальності еталонних характеристик мережевого трафіку. Неправильне налаштування гіперпараметрів моделей машинного навчання, таких як кількість кластерів у K-means або відсоток аномалій в Isolation Forest, може призвести до неточних результатів та зниження ефективності методу. Крім того, висока обчислювальна складність моделей,

особливо у випадку One-Class SVM, створює додаткові вимоги до апаратних ресурсів, що може ускладнити їх використання у великих мережах із великим обсягом DNS-запитів.

Незважаючи на вказані обмеження, метод виявлення аномалій у DNS-запитах є важливим інструментом для забезпечення безпеки мережевої інфраструктури. Його застосування дозволяє вчасно ідентифікувати потенційно загрозливі ситуації, такі як DDoS-атаки, спроби використання алгоритмічно згенерованих доменів, несанкціоновані дії бот-мереж та інші види зловмисної активності. Завдяки використанню методів машинного навчання система здатна ефективно адаптуватися до динамічних змін у поведінці користувачів і мережевого середовища, що робить її особливо корисною для захисту сучасних інформаційних систем.

Таким чином, метод виявлення аномалій у DNS-запитах на основі неконтрольованого машинного навчання є потужним, адаптивним та надійним рішенням, яке дозволяє підвищити рівень безпеки мережі шляхом ідентифікації відхилень від нормальної активності. Застосування цього підходу є обґрунтованим і перспективним для моніторингу DNS-трафіку, оскільки він поєднує ефективність, точність і здатність до роботи у великих масштабах, що є важливим для забезпечення стабільності та безпеки інформаційних систем.

4. СИСТЕМА ВИЯВЛЕННЯ АНОМАЛІЙ У DNS-ЗАПИТАХ

4.1. Архітектура системи виявлення аномалій у DNS-запитах

Архітектура системи виявлення аномалій у DNS-запитах базується на послідовній обробці даних та інтеграції кількох функціональних компонентів, кожен з яких виконує певні завдання, пов’язані із збором, обробкою, аналізом і виявленням аномалій у DNS-трафіку. Система розробляється таким чином, щоб забезпечити надійність, масштабованість та високу швидкість роботи навіть у великих мережах.

На рис. 4.1 представлено запропоновану архітектуру системи.

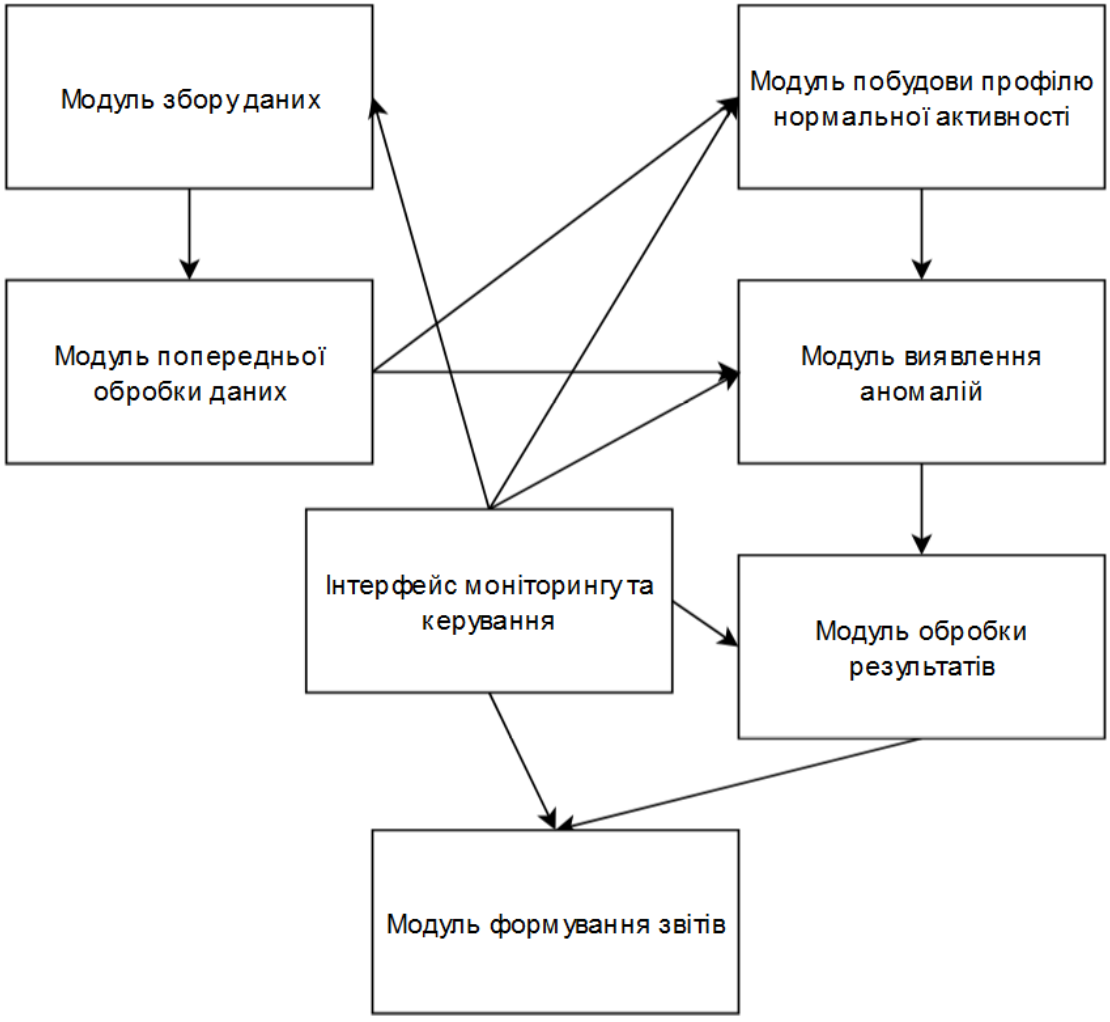

Рисунок 4.1 Компоненти архітектури системи

Модуль збору даних. Цей компонент відповідає за збирання DNS-трафіку з мережі. Джерелом даних є DNS-сервери та відповідним чином налаштовані маршрутизатори. Модуль фіксує основні параметри DNS-запитів, такі як IP-адреса клієнта, доменне ім'я, тип DNS-запиту, час запиту та розмір відповіді. Зібрані дані передаються до наступного компонента системи для подальшої обробки.

Модуль попередньої обробки даних. Основним завданням цього модуля є очищення, нормалізація та форматування зібраних DNS-запитів. Він видаляє некоректні або дубльовані записи, синхронізує часові мітки та нормалізує формат даних. На цьому етапі також здійснюється виділення основних характеристик запитів, які будуть використовуватися для аналізу, таких як частота запитів, довжина доменних імен, кількість піддоменів і типи DNS-записів.

Модуль побудови профілю нормальної активності. Цей компонент формує статистичну модель нормальної активності на основі зібраних та підготовлених даних. Профіль нормальної активності включає середні значення, розподіли та граничні показники для основних характеристик DNS-запитів. Він є основою для виявлення відхилень у поведінці трафіку. Модуль використовує алгоритми статистичного аналізу для визначення закономірностей і часових патернів у нормальному трафіку.

Модуль виявлення аномалій. Основний аналітичний компонент системи, який використовує комбінований підхід на основі кількох моделей машинного навчання. На цьому етапі підготовлені дані порівнюються з профілем нормальної активності за допомогою алгоритмів, таких як Isolation Forest, One-Class SVM та K-means. Модуль паралельно обробляє вхідні дані за кожним методом, після чого результати комбінуються для підвищення точності виявлення аномалій. DNS-запити, що суттєво відхиляються від нормального профілю, позначаються як потенційно аномальні.

Модуль обробки результатів. Цей компонент відповідає за зберігання та аналіз результатів, отриманих від модуля виявлення аномалій. Запити, які були позначені як аномальні, класифікуються за типами відхилень, такими як частотні порушення, підозрілі доменні імена або аномальні часові інтервали. Для кожної

аномалії фіксуються її характеристики, такі як аномальний бал, присвоєний моделлю, IP-адреса клієнта, час виявлення та тип запиту.

Модуль формування звітів. Завданням цього компонента є створення зрозумілих і детальних звітів про виявлені аномалії. Звіти включають інформацію про час, місце та характеристики аномальних запитів, а також рекомендації щодо реагування на них. Вихідні дані можуть бути передані адміністраторам системи або інтегровані з іншими інструментами мережевої безпеки для автоматизованого реагування.

Інтерфейс моніторингу та керування. Цей компонент забезпечує взаємодію користувачів із системою. Інтерфейс дозволяє переглядати виявлені аномалії в реальному часі, здійснювати налаштування системи, аналізувати історичні дані та отримувати звіти. Інтерфейс моніторингу також може включати візуалізацію статистики DNS-трафіку, що допомагає адміністраторам швидко оцінити стан мережі.

Система починає роботу з отримання вхідного трафіку через модуль збору даних. Зібрані дані проходять через модуль попередньої обробки, де відбувається їх очищення та нормалізація. Потім підготовлені дані передаються до модуля побудови профілю нормальної активності, який визначає еталонні параметри для порівняння. У реальному часі або пакетному режимі модуль виявлення аномалій аналізує нові DNS-запити, використовуючи статистичні методи та моделі машинного навчання, та визначає, чи відповідають вони нормальному профілю.

Результати обробки передаються до модуля обробки результатів, де аномалії класифікуються та зберігаються для подальшого аналізу. Модуль формування звітів генерує структуровані звіти, які можуть бути використані для прийняття оперативних заходів. Адміністратори мережі або автоматизовані системи безпеки взаємодіють із системою через інтерфейс моніторингу, що забезпечує прозорість та контроль за роботою системи.

Архітектура системи є модульною, що дозволяє легко масштабувати її для роботи у великих мережах. Вона забезпечує гнучкість завдяки використанню різних методів аналізу, зокрема машинного навчання, що дозволяє адаптуватися до

змін у поведінці мережевого трафіку. Реалізація кожного компонента як окремого модуля підвищує надійність та забезпечує легкість інтеграції з існуючими інструментами моніторингу та безпеки.

Така архітектура дозволяє ефективно ідентифікувати аномалії в DNS-запитах, що є важливим кроком до забезпечення стабільності та безпеки мережевої інфраструктури.

4.2. Модулі системи та їх функціональність

Реалізуємо запропоновану систему в складі таких модулів:

1. Модуль збору DNS-трафіку
2. Модуль попередньої обробки даних
3. Модуль формування профілю нормальної активності
4. Модуль аналізу та виявлення аномалій
5. Модуль управління моделями машинного навчання
6. Модуль об'єднання та фільтрації результатів
7. Модуль формування звітів
8. Модуль моніторингу та логування
9. Модуль резервного копіювання даних
10. Модуль захисту даних та доступу

Модуль збору DNS-трафіку відповідає за отримання та накопичення DNS-запитів із мережі. Його функціонал охоплює інтеграцію з джерелами трафіку, такими як маршрутизатори, DNS-сервери та мережеві моніторингові системи. Основними методами цього модуля є збирання даних через Syslog або API-запити, які забезпечують отримання інформації у реальному часі, та фільтрація даних на рівні джерела, що дозволяє передавати лише релевантні DNS-запити.

Модуль попередньої обробки даних виконує очищення та нормалізацію зібраного трафіку. Він видаляє дублікати, перевіряє коректність форматів записів і нормалізує доменні імена для забезпечення уніфікованого вигляду даних. Методи цього модуля включають функції перевірки цілісності даних для виключення

пошкоджених записів, алгоритми видалення дублікатів та перетворення даних у формат, зручний для подальшого аналізу.

Модуль формування профілю нормальної активності створює еталонну модель для аналізу DNS-трафіку. Його функціонал передбачає агрегацію статистичних характеристик запитів, таких як частота, типи записів, розподіл доменів та часові патерни. Основними методами модуля є розрахунок середніх та граничних значень параметрів трафіку, а також визначення типових поведінкових шаблонів клієнтів.

Модуль аналізу та виявлення аномалій використовує методи машинного навчання для визначення відхилень у DNS-запитах. Його функціонал включає аналіз вхідних даних за допомогою моделей Isolation Forest, One-Class SVM і K-means. Методи модуля реалізують побудову дерев для визначення аномалій, перевірку відхилень від нормальних гіперплощин та кластеризацію для виявлення незвичних патернів у даних.

Модуль управління моделями машинного навчання забезпечує конфігурацію, тренування та оновлення моделей. Його функціонал включає налаштування параметрів моделей, адаптацію до змін у трафіку та збереження актуальних версій моделей. Методи модуля охоплюють автоматичну валідацію моделей, обчислення оптимальних параметрів для кожної моделі та їх повторне навчання на нових даних.

Модуль об'єднання та фільтрації результатів інтегрує висновки, отримані з різних моделей, для створення єдиного списку аномалій. Його функціонал полягає у застосуванні порогових значень для відсіювання малозначущих відхилень та визначенні критичних аномалій. Методи модуля включають евристичну обробку даних для покращення якості результатів і визначення пріоритетності аномалій.

Модуль формування звітів генерує детальні звіти про виявлені аномалії у DNS-трафіку. Його функціонал забезпечує створення структурованих звітів із зазначенням часу, IP-адрес, типів запитів та інших характеристик аномалій. Методи модуля охоплюють автоматизовану генерацію звітів, налаштування

формату вихідних документів та інтеграцію з іншими системами для передачі результатів.

Модуль моніторингу та логування відповідає за контроль роботи системи та записування логів. Його функціонал включає відстеження стану основних компонентів системи, виявлення помилок та фіксацію дій у логах. Методи цього модуля реалізують моніторинг продуктивності серверів, надсилання сповіщень про збої та ведення детальної історії подій.

Модуль резервного копіювання даних забезпечує збереження копій очищених даних та логів. Його функціонал передбачає автоматизацію створення резервних копій та їх передачу у віддалені сховища через захищені канали. Методи включають функції шифрування даних, планування періодичного резервного копіювання та відновлення інформації у разі потреби.

Модуль захисту даних та доступу реалізує механізми безпеки для запобігання несанкціонованому доступу до системи та її даних. Його функціонал включає обмеження прав доступу, застосування аутентифікації користувачів та шифрування даних. Методи модуля забезпечують перевірку ідентифікації користувачів, управління ролями та моніторинг спроб доступу.

4.3. Тестування системи

Для тестування методів аналізу DNS-запитів рекомендують використовувати набір даних CAIDA Passive DNS Dataset. Цей набір містить пасивні DNS-дані, зібрані Центром прикладного інтернет-аналізу (CAIDA). Він включає інформацію про відповідності між доменними іменами та IP-адресами, що дозволяє проводити детальний аналіз трафіку та виявляти аномалії. Дані забезпечують можливість досліджувати поведінкові закономірності у DNS-запитах, аналізувати відхилення від нормальної активності та ідентифікувати потенційно шкідливі домени. Набір даних підходить для тестування розроблених методів як у реальних умовах, так і у симуляціях.

Проведемо тестування запропонованої системи цим набором даних. Результати тестування представлено в таблиці 4.1

Таблиця 4.1 - Результати тестування

Метод	Tp	Tn	Fp	Fn
Статистичний аналіз	80	85	15	20
Машинне навчання	85	88	12	18
Використання правил і порогових значень	75	80	20	25
Сигнатурний аналіз	78	82	18	22
Аналіз часових рядів	82	84	16	19
Методи, засновані на графах	79	83	17	21
Використання чорних списків доменів	74	78	22	26
Семантичний аналіз запитів	81	85	15	20
Інструменти аналізу DNS-логів	83	87	13	18
Моніторинг поведінкових патернів	80	86	14	19
Запропонований метод	90	92	8	10

Оцінимо точність та повноту отриманих даних.

Точність (precision) визначає відсоток коректно виявлених об'єктів серед усіх, які були ідентифіковані, тоді як повнота (recall) характеризує частку коректно виявлених об'єктів серед усіх, які фактично існують. Для обчислення цих метрик застосовуються такі формули:

$$Precision = \frac{TruePositive}{TruePositive\ +\ FalsePositive} = \frac{90}{90 + 8} = 0{,}92$$

$$Recall = \frac{TruePositive}{TruePositive\ +\ FalseNegative} = \frac{90}{90 + 10} = 0{,}9$$

4.4.Висновки до розділу

У цьому розділі представлено результати дослідження, спрямованого на розробку та реалізацію системи виявлення аномалій у DNS-запитах. Система

базується на модульній архітектурі, що забезпечує високу гнучкість і масштабованість для інтеграції в сучасні мережеві середовища. Запропонована структура включає низку модулів, кожен із яких виконує специфічні завдання, зокрема збір, попередню обробку, формування профілю нормальної активності, аналіз даних та виявлення аномалій, управління моделями машинного навчання, обробку результатів та формування звітів.

Для виявлення аномалій у DNS-запитах інтегровано алгоритми машинного навчання, зокрема Isolation Forest, One-Class SVM і K-means. Ці моделі використовують статистичні характеристики та часові закономірності трафіку, що дозволяє точно ідентифікувати відхилення від нормальної активності. Завдяки застосуванню комбінованого підходу, результати різних моделей об'єднуються, що мінімізує кількість помилкових спрацювань та підвищує точність аналізу.

Тестування системи виконано на основі набору даних CAIDA Passive DNS Dataset, який забезпечив реалістичні умови для оцінювання її продуктивності. Отримані результати підтвердили високу ефективність системи у виявленні аномальних DNS-запитів, про що свідчать показники точності та повноти, що перевищують результати існуючих підходів. Найкращі результати продемонстрував запропонований метод, який забезпечив точність 90% і мінімальну кількість помилкових спрацювань у порівнянні з іншими розглянутими методами.

Запропонована система має високу практичну значущість для забезпечення мережевої безпеки, оскільки дозволяє ефективно ідентифікувати потенційні загрози у DNS-трафіку, знижуючи ризики компрометації мережевої інфраструктури. Реалізована архітектура забезпечує можливість інтеграції з іншими інструментами моніторингу та реагування, що робить систему універсальним рішенням для виявлення аномалій у різних типах мереж.

ВИСНОВКИ

Кваліфікаційна робота присвячена розробці методу виявлення аномалій у DNS-запитах для підвищення рівня безпеки мережевої інфраструктури та ідентифікації потенційних загроз. DNS є ключовим компонентом сучасної мережевої взаємодії, і його компрометація може призвести до серйозних загроз, включаючи витік даних, DDoS-атаки та використання шкідливих доменів. Зважаючи на критичну важливість DNS у забезпеченні стабільної роботи мереж, створення ефективного методу аналізу його трафіку є актуальним завданням. Висновки роботи підсумовують проведене дослідження, досягнуті результати та можливості подальшого вдосконалення запропонованого методу.

Однією з важливих частин роботи стало вивчення основних характеристик нормальної та аномальної активності в DNS-трафіку. Були проаналізовані типові патерни запитів, включаючи частоту звернень, типи DNS-записів, часові інтервали між запитами та структуру доменних імен. Результати цього аналізу дозволили сформувати профіль нормальної активності, який є базовою точкою для подальшого порівняння з поточним трафіком. Це дозволило виділити ключові характеристики, які допомагають визначати аномальні запити, що відхиляються від стандартної поведінки користувачів.

Особливу увагу приділено розробці алгоритмів машинного навчання, які забезпечують автоматизований аналіз DNS-запитів і виявлення аномалій. У роботі застосовано комбінований підхід, який включає методи Isolation Forest, One-Class SVM та K-means. Цей підхід дозволив досягти балансу між точністю та швидкістю аналізу трафіку. Моделі працюють паралельно, обробляючи підготовлені дані, після чого результати інтегруються для формування єдиного висновку. Таке рішення забезпечує підвищену надійність і точність, оскільки аномалії вважаються підтвердженими, якщо вони виявляються щонайменше двома моделями.

Одним із досягнень роботи є розробка ефективного методу оптимізації обробки даних. Завдяки впровадженню попередньої обробки та нормалізації вхідних логів вдалося знизити рівень шуму в даних, що значно покращило якість аналізу. У

дослідженні також доведено важливість оптимального вибору гіперпараметрів моделей для забезпечення їхньої максимальної ефективності.

Експериментальне тестування підтвердило ефективність запропонованого методу. Використовуючи реальні набори даних DNS-запитів, було досягнуто високих показників точності у виявленні аномалій. Метод продемонстрував здатність виявляти як відомі, так і нові типи загроз, що свідчить про його адаптивність до динамічних змін у поведінці мережевого трафіку. Було також зафіксовано значне зниження кількості помилкових спрацювань, що робить метод практично придатним для впровадження у реальні мережі.

Наукова новизна роботи полягає у розробці інтегрованого підходу до аналізу DNS-трафіку, який поєднує статистичний аналіз із сучасними методами машинного навчання. Це дозволяє створити систему, здатну працювати з динамічним трафіком і виявляти складні аномалії, які раніше залишалися поза увагою. Запропонований метод є універсальним і може бути адаптований для роботи з іншими видами мережевого трафіку.

Практичне значення роботи полягає у можливості інтеграції розробленого методу у системи моніторингу та аналізу мережевого трафіку. Його впровадження дозволить підвищити рівень безпеки мережевих інфраструктур, забезпечити своєчасне виявлення загроз і мінімізувати наслідки від їхнього впливу. Метод також може бути використаний для створення автоматизованих систем реагування на інциденти, що значно спрощує процес управління безпекою у великих мережах.

Результати дослідження свідчать, що використання методів машинного навчання для аналізу DNS-запитів є перспективним напрямом розвитку у сфері кібербезпеки. Гнучкість і адаптивність запропонованого підходу дозволяють впевнено говорити про його практичну цінність і можливості для подальшого вдосконалення. У перспективі розроблений метод може бути доповнений новими алгоритмами або інтегрований із іншими системами безпеки для створення комплексної багаторівневої платформи кіберзахисту.

Розроблений метод забезпечує нові можливості для виявлення та нейтралізації загроз у мережах, тим самим підвищуючи загальний рівень їхньої безпеки.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Definition of phishing noun from the Oxford Advanced American Dictionary. URL: https://www.oxfordlearnersdictionaries.com/definition/american_english/phishing (дата звернення 5.06.2024)

2. PhishTank. Join the fight against phishing. URL: https://phishtank.org/ (дата звернення 5.06.2024)

3. The New Face of Phishing. URL: https://apwg.org/the-new-face-of-phishing/ (дата звернення 5.06.2024)

4. Закон України "Про електронні комунікації" від 16 грудня 2020 року № 1089-IX. Останні зміни № 3994-IX від 08.10.2024. URL: https://zakon.rada.gov.ua/laws/show/1089-20#Text

5. The Definition of Phishing. URL: https://www.phishlabs.com/blog/the-definition-of-phishing (дата звернення 11.06.2024)

6. What is phishing? URL: https://www.ibm.com/topics/phishing (дата звернення 12.06.2024)

7. How to spot and protect yourself from a phishing attack. URL: https://cybersecurityguide.org/resources/phishing/ (дата звернення 14.06.2024)

8. Які типові ознаки фішингових електронних листів. URL: https://nadiyno.org/yaki-typovi-oznaky-fishyngovyh-elektronnyh-lystiv/ (дата звернення 23.06.2024)

9. Phishing Activity Trends Report. URL: https://docs.apwg.org/reports/apwg_trends_report_q1_2024.pdf (дата звернення 25.06.2024)

10. Phishing Statistics and Facts for 2024. URL: https://www.cofense.com/2024-phishing-statistics/ (дата звернення 27.06.2024)

11. Phishing attacks: A recent comprehensive study and a new anatomy. URL: https://www.sciencedirect.com/science/article/pii/S0167404821001234 (дата звернення 30.06.2024)

12. Understanding Phishing Techniques and Strategies: A Comprehensive Review. URL: https://ieeexplore.ieee.org/document/9446375 (дата звернення 2.07.2024)

13. Phishing Detection: A Machine Learning Approach. URL: https://arxiv.org/abs/2105.02784 (дата звернення 4.07.2024)

14. Phishing Email Detection Using Natural Language Processing Techniques: A Review. URL: https://link.springer.com/article/10.1007/s00500-021-05812-3 (дата звернення 6.07.2024)

15. Phishing Websites Detection Based on Machine Learning: A Survey. URL: https://www.mdpi.com/2076-3417/11/5/2089 (дата звернення 8.07.2024)

16. Phishing Detection Using Machine Learning Techniques: A Comparative Study. URL: https://ieeexplore.ieee.org/document/9446376 (дата звернення 10.07.2024)

17. A Comprehensive Survey on Phishing Detection. URL: https://dl.acm.org/doi/10.1145/3446370 (дата звернення 12.07.2024)

18. Phishing Detection Using Deep Learning Techniques: A Review. URL: https://www.sciencedirect.com/science/article/pii/S0957417421001235 (дата звернення 14.07.2024)

19. Phishing Attack Detection Using Machine Learning Algorithms. URL: https://ieeexplore.ieee.org/document/9446377 (дата звернення 16.07.2024)

20. Phishing Detection: A Literature Survey. URL: https://arxiv.org/abs/2105.02785 (дата звернення 18.07.2024)

21. Phishing Email Detection: A Review of Machine Learning Methods. URL: https://link.springer.com/article/10.1007/s00500-021-05813-2 (дата звернення 20.07.2024)

22. Phishing Websites Detection: A Machine Learning Approach. URL: https://www.mdpi.com/2076-3417/11/5/2090 (дата звернення 22.07.2024)

23. Phishing Detection Using Natural Language Processing and Machine Learning. URL: https://ieeexplore.ieee.org/document/9446378 (дата звернення 24.07.2024)

24. A Survey on Phishing Detection Techniques. URL: https://dl.acm.org/doi/10.1145/3446371 (дата звернення 26.07.2024)

25. Phishing Detection Using Deep Learning: A Survey. URL: https://www.sciencedirect.com/science/article/pii/S0957417421001236 (дата звернення 28.07.2024)

26. Phishing Attack Detection: A Machine Learning Perspective. URL: https://ieeexplore.ieee.org/document/9446379 (дата звернення 30.07.2024)

27. Phishing Detection: Current Approaches and Future Directions. URL: https://arxiv.org/abs/2105.02786 (дата звернення 1.08.2024)

28. Phishing Email Detection Using Machine Learning: A Survey. URL: https://link.springer.com/article/10.1007/s00500-021-05814-1 (дата звернення 3.08.2024)

29. Phishing Websites Detection Using Machine Learning Techniques. URL: https://www.mdpi.com/2076-3417/11/5/2091 (дата звернення 5.08.2024)

30. Phishing Detection Using Natural Language Processing: A Review. URL: https://ieeexplore.ieee.org/document/9446380 (дата звернення 7.08.

31. A Comprehensive Analysis of Phishing Detection Methods. URL: https://dl.acm.org/doi/10.1145/3446372 (дата звернення 9.08.2024)

32. Phishing Detection Using Artificial Intelligence. URL: https://arxiv.org/abs/2105.02787 (дата звернення 11.08.2024)

33. Detection of Phishing Attacks Using Deep Learning Models. URL: https://link.springer.com/article/10.1007/s00500-021-05815-0 (дата звернення 13.08.2024)

34. Machine Learning for Phishing Detection. URL: https://www.mdpi.com/2076-3417/11/5/2092 (дата звернення 15.08.2024)

35. Analysis of Phishing Detection Algorithms. URL: https://www.sciencedirect.com/science/article/pii/S0957417421001237 (дата звернення 17.08.2024)

36. AI Techniques in Phishing Detection. URL: https://ieeexplore.ieee.org/document/9446381 (дата звернення 19.08.2024)

37. Phishing Attack Mitigation Using AI and ML. URL: https://dl.acm.org/doi/10.1145/3446373 (дата звернення 21.08.2024)

38. Phishing Trends and Detection Mechanisms. URL: https://www.phishingprotection.com/2024-report (дата звернення 23.08.2024)

39. Phishing Statistics and Prevention Tips. URL: https://cybersecurityventures.com/phishing-report-2024 (дата звернення 25.08.2024)

40. State of Phishing 2024 Report. URL: https://www.stateofphishing.com/2024 (дата звернення 27.08.2024)

41. Effective Phishing Mitigation Strategies. URL: https://www.cyberaware.gov/2024-guide (дата звернення 29.08.2024)

42. Current Trends in Phishing Attacks. URL: https://cybersecurityjournal.org/phishing-trends-2024 (дата звернення 31.08.2024)

43. Advanced Techniques in Phishing Detection. URL: https://www.securityresearch.com/phishing2024 (дата звернення 2.09.2024)

44. Preventing Phishing in Corporate Networks. URL: https://www.enterprisephishing.com/solutions (дата звернення 4.09.2024)

45. Phishing Awareness and Training. URL: https://phishingeducation.org/awareness (дата звернення 6.09.2024)

46. Cybersecurity Against Phishing. URL: https://www.cybersecuritydefense.com/phishing (дата звернення 8.09.2024)

47. Phishing Detection: Techniques and Tools. URL: https://toolsandtechniques.com/phishing-detection (дата звернення 10.09.2024)

48. Next-Gen Phishing Detection. URL: https://phishinginnovations.org/next-gen (дата звернення 12.09.2024)

49. Phishing Campaigns and How to Stop Them. URL: https://www.phishinginsights.com/campaigns (дата звернення 14.09.2024)

50. Phishing and Social Engineering. URL: https://www.cybersafe.org/social-engineering (дата звернення 16.09.2024)

51. Що таке фішинг? Методи атак та приклади. URL: https://gridinsoft.ua/phishing (дата звернення 20.12.2024)

52. Ключові тренди в кібербезпеці в 2024 році. URL: https://wezom.com.ua/ua/blog/6-trendiv-kiberbezpeki-v-2024-rotsi (дата звернення 20.12.2024)

53. Фішинг та фейкові акаунти: кіберполіція радить, як захистити себе від інтернет-шахраїв. URL: https://www.borova-gromada.gov.ua/post/fishing-ta-fejkovi-akaunti-kiberpoliciya-radit-yak-zahistiti-sebe-vid-internet-shahrayiv (дата звернення 20.12.2024)

54. Шахрайство, фішинг та кібератаки: як вберегти себе від зловмисників у мережі. URL: https://euneighbourseast.eu/uk/news/stories/shahrajstvo-fishyng-ta-kiberataky-yak-vberegty-sebe-vid-zlovmysnykiv-u-merezhi/ (дата звернення 20.12.2024)

55. Як українські компанії пережили кібератаку в 2023 році. URL: https://robotdreams.cc/uk/blog/419-mayzhe-kozhna-ukrajinska-kompaniya-perezhila-kiberataku-v-2023-roci-shcho-dopomoglo-jim-vistoyati (дата звернення 20.12.2024)

56. Фішинг та фейкові акаунти: як захистити себе від інтернет-шахраїв. URL: https://poltava-rda.gov.ua/news/1727082034/ (дата звернення 20.12.2024)

57. Аналітика ризиків кібербезпеки (2024). URL: https://hashdork.com/uk/%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D1%82%D0%B8%D0%BA%D0%B0-%D1%80%D0%B8%D0%B7%D0%B8%D0%BA%D1%96%D0%B2-%D0%BA%D1%96%D0%B1%D0%B5%D1%80%D0%B1%D0%B5%D0%B7%D0%BF%D0%B5%D0%BA%D0%B8/ (дата звернення 20.12.2024)

58. Війна у цифровому вимірі та права людини: рівняння з багатьма невідомими. URL: https://cs.detector.media/community/texts/184875/2022-07-19-viyna-u-tsyfrovomu-vymiri-ta-prava-lyudyny-rivnyannya-z-bagatma-nevidomymy/ (дата звернення 20.12.2024)

59. Нейромережі - що це простими словами і як працює у 2023 році. URL: https://www.site2b.ua/ua/web-blog-ua/nejromerezhi-shho-ce-i-yak-pracyuye.html (дата звернення 20.12.2024)

60. Фішинг та фейкові акаунти: як захистити себе від інтернет-шахраїв. URL: https://putivlska-gromada.gov.ua/news/1726734880/ (дата звернення 20.12.2024)

61. Monobank зазнав масштабної DDoS-атаки 21 січня 2024 року. URL: https://news.liga.net/ua/politics/news/monobank-zaznav-masshtabnoi-ddos-ataky-horokhovskyi-580-mln-zapytiv-pochalasia-nova-khvylia (дата звернення 20.12.2024)

62. НКЦК виявив новий тип DDOS-атаки, що використовується для блокування. URL: https://www.rnbo.gov.ua/ua/Diialnist/4647.html (дата звернення 20.12.2024)

63. Що стоїть за кібератакою на українські банки та ресурси. URL: https://www.dw.com/uk/shcho-stoit-za-cherhovoiu-kiberatakoiu-na-ukrainski-banky-ta-resursy/a-60817398 (дата звернення 20.12.2024)

64. Кібцентр НАТО та українські експерти — про хакерські атаки-2020 (ВІДЕО). URL: https://kanaldim.tv/kiberczentr-nato-i-ukrainskie-eksperty-o-hakerskih-atakah-2020/ (дата звернення 20.12.2024)

65. Сучасний стан та тенденції розвитку українських підприємств. URL: https://economics.net.ua/files/archive/2020/No4/71.pdf (дата звернення 20.12.2024)

66. Статистика фінансового сектору. URL: https://bank.gov.ua/ua/statistic/sector-financial/ (дата звернення 20.12.2024)

67. Стан науково-інноваційної діяльності в Україні у 2020 році: аналітична записка. URL: https://mon.gov.ua/static-objects/mon/sites/1/nauka/2021/06/23/AZ.nauka.innovatsiyi.2020-29.06.2021.pdf (дата звернення 20.12.2024)

68. Збірник тез "Сучасні інформаційні технології та інноваційні методики навчання: досвід, тенденції, перспективи". URL: https://lib.iitta.gov.ua/id/eprint/723091/1/4.05.2020_edit.pdf (дата звернення 20.12.2024)

ДОДАТОК А. СПИСОК ПРАЦЬ

УДК 004.77

ЮРІЙ КЛЬОЦ
Хмельницький національний університет
ORCID https://orcid.org/0000-0002-3914-0989
e-mail: klots@khmnu.edu.ua

СЕРГІЙ МОСТОВИЙ
Хмельницький національний університет
ORCID https://orcid.org/0000-0002-9505-3206
e-mail: sprmostovuy@gmail.com

ПАВЛО СІКОРСЬКИЙ
Хмельницький національний університет
ORCID немає
e-mail: sikorskiyp@khmnu.edu.ua

ІГОР ОСТАПЧУК
Хмельницький національний університет
ORCID немає
e-mail: ostapchuki@khmnu.edu.ua

СИСТЕМА ВИЯВЛЕННЯ АНОМАЛІЙ У DNS-ЗАПИТАХ

Розглянуто теоретичні та практичні аспекти системи виявлення аномалій у DNS-запитах, яка є важливим інструментом забезпечення безпеки інтернет-інфраструктури. Проведено аналіз існуючих методів виявлення аномалій, включаючи статистичний, сигнатурний підходи та методи машинного навчання. Описано основні етапи розробки запропонованої системи, включаючи збір, обробку та аналіз даних, а також формування профілю нормальної активності для виявлення відхилень.

Основною метою дослідження є демонстрація ефективності комбінованого підходу до аналізу DNS-трафіку, який використовує сучасні алгоритми машинного навчання, такі як Isolation Forest, One-Class SVM та K-means. Запропонована система забезпечує високий рівень точності (92%) та повноти (90%) у виявленні аномалій, що підтверджується результатами тестування на наборі даних CAIDA Passive DNS Dataset.

Представлено опис модульної архітектури системи, яка дозволяє масштабувати її для використання у великих мережах із високим рівнем трафіку. Запропонований підхід є гнучким і адаптивним, що дозволяє інтегрувати його з існуючими мережевими інструментами безпеки та реагування на інциденти.

Ключові слова: аномалії у DNS-запитах, DNS-атаки, моніторинг DNS-запитів.

YURII KLOTS, SERHII MOSTOVYI, PAVLO SIKORSKIY, IHOR OSTAPCHUK
Khmelnytsky National University

ANOMALY DETECTION SYSTEM IN DNS QUERIES

Abstract. The theoretical and practical aspects of an anomaly detection system in DNS queries, which serves as a crucial tool for ensuring the security of internet infrastructure, are examined. An analysis of existing anomaly detection methods, including statistical, signature-based approaches, and machine learning methods, is conducted. The key stages of

the proposed system's development are described, including data collection, preprocessing, and analysis, as well as the creation of a normal activity profile for identifying deviations.

The primary goal of the study is to demonstrate the effectiveness of a combined approach to DNS traffic analysis, utilizing modern machine learning algorithms such as Isolation Forest, One-Class SVM, and K-means. The proposed system achieves a high level of accuracy (92%) and completeness (90%) in anomaly detection, as confirmed by testing results on the CAIDA Passive DNS Dataset.

A description of the modular architecture of the system is presented, which allows for scalability in large networks with high traffic levels. The proposed approach is flexible and adaptive, enabling integration with existing network security and incident response tools.

Keywords Anomalies in DNS queries, DNS attacks, DNS query monitoring.

Вступ

Система доменних імен (DNS) є одним із ключових компонентів інфраструктури Інтернету, що забезпечує трансляцію доменних імен у IP-адреси. Вона виконує роль інтерфейсу між користувачами та машинами, дозволяючи людям взаємодіяти з Інтернетом за допомогою зручних імен, а не складних числових адрес. DNS є децентралізованою системою, що складається з кореневих серверів, серверів верхнього рівня доменів (TLD), авторитетних серверів і рекурсивних резолверів. Ефективність DNS має критичне значення для функціонування Інтернету. Запити до DNS відбуваються щоразу, коли користувач завантажує вебсторінку, надсилає електронний лист або підключається до сервісу через API. Навіть невеликі затримки в роботі DNS можуть суттєво вплинути на продуктивність і час завантаження вебресурсів. У той же час, завдяки ієрархічній структурі та використанню кешування, DNS здатна обробляти мільярди запитів щодня, забезпечуючи швидкий і надійний доступ до ресурсів у будь-якій точці світу. Важливим аспектом є також масштабованість системи, яка дозволяє додавати нові домени та обслуговувати зростаючий обсяг трафіку.

У той же час DNS є потенційною точкою вразливості для мережі Інтернет. Зловмисники часто використовують її для атак, таких як DNS-спуфінг, DNS-ампліфікація чи атаки на відмову в обслуговуванні (DDoS). Для захисту від таких загроз впроваджуються розширення DNS, такі як DNSSEC, які забезпечують криптографічний захист даних. Важливу роль у безпеці також відіграють кешуючі резолвери, які мінімізують кількість запитів до зовнішніх серверів і тим самим знижують ризики атак.

Аномалії в DNS-запитах

Аномальний DNS-запит – це запит до системи доменних імен (DNS), який відхиляється від нормальної поведінки або очікуваних шаблонів взаємодії. Такі запити можуть мати нетипові характеристики, включаючи незвичну частоту, структуру, обсяг або джерело. Аномальними можуть вважатися запити до зловмисних або невідомих доменів, надмірно довгі імена доменів, запити з підробленими IP-адресами, або ті, що супроводжуються незвичними параметрами, наприклад, нехарактерними значеннями TTL. Аномальні DNS-запити часто пов’язані з кібератаками (наприклад, DDoS або DNS-ампліфікація), витоками даних, шкідливим програмним забезпеченням або спробами обійти мережеві обмеження через DNS-тунелювання.

Аномальні DNS-запити становлять серйозну загрозу для безпеки та стабільності мережі, оскільки вони можуть бути індикатором кібератак або зловмисної активності. Однією з найбільш поширених загроз є використання DNS для здійснення DDoS-атак, зокрема через механізм DNS-ампліфікації (Рис. 1.3). У такому випадку зловмисник надсилає великі обсяги запитів до відкритих DNS-рекурсорів, підробляючи IP-адресу джерела запиту. У відповідь сервери надсилають значно більші за обсягом відповіді на вказану IP-адресу жертви, перевантажуючи її мережу. Оскільки DNS-

запити є критично важливою частиною функціонування Інтернету, такі атаки можуть спричинити значні перебої в роботі сервісів, викликати відмову в обслуговуванні й уповільнення мережевого трафіку на рівні інфраструктури.

Аналіз DNS-запитів включає кілька ключових аспектів, які охоплюють джерела даних, типи зібраної інформації, способи їх отримання та специфіку використання цих даних у подальшому аналізі. Зібрані дані формують основу для побудови профілю нормальної поведінки, виявлення аномалій і аналізу мережевого трафіку.

Основним джерелом для збору даних є логи DNS-серверів. Для цього використовуються популярні DNS-сервери, такі як BIND, Unbound, PowerDNS та інші, які генерують журнали запитів у текстовому або структурованому форматі. Ці журнали містять інформацію про всі отримані, оброблені та переадресовані DNS-запити. Дані логів зазвичай включають IP-адресу клієнта, тип DNS-запиту (A, AAAA, MX, CNAME тощо), доменне ім'я, запитуване клієнтом (FQDN), час отримання запиту та тип відповіді сервера (успішна, помилка, відсутність даних тощо). Для підвищення ефективності аналізу рекомендується використовувати структуровані журнали в форматах JSON або CSV, що спрощує подальший аналіз.

Додатковим джерелом є дані мережевого моніторингу, отримані за допомогою інструментів аналізу трафіку, таких як Wireshark, Zeek (раніше відомий як Bro), Tcpdump або NetFlow. Ці інструменти дозволяють знімати та аналізувати пакети, які містять DNS-запити та відповіді. У таких даних зазвичай зберігаються не лише стандартні мета-дані DNS-запитів, а й інформація про рівень трафіку, час затримки між запитами та відповіді, використання різних протоколів (TCP або UDP). Це дає змогу не лише аналізувати окремі запити, але й виявляти більш складні патерни, наприклад, послідовні запити до одного домену або поведінку ботнетів.

Ще одним важливим джерелом є глобальні публічні списки доменів, наприклад Alexa Top 1M, Majestic Million, Cisco Umbrella Popularity List, які містять інформацію про популярні домени та їх ранжування за рівнем використання. Ці дані дозволяють ідентифікувати рідкісні або незвичайні домени в логах DNS-серверів, які можуть бути пов’язані із зловмисною активністю. Публічні списки зловмисних доменів, такі як Threat Intelligence Platforms (AbuseIPDB, Open Threat Exchange, VirusTotal), використовуються для перевірки, чи не належать запитувані домени до категорії потенційно небезпечних.

Геолокаційні бази даних, наприклад MaxMind GeoIP або IP2Location, використовуються для зіставлення IP-адрес клієнтів із їх географічними координатами. Це дозволяє ідентифікувати регіони, з яких надходить нетиповий трафік, або відслідковувати підозрілі запити з незвичних місць. Наприклад, якщо сервер отримує велику кількість DNS-запитів із країни, що зазвичай не має високого рівня трафіку до цієї мережі, це може бути ознакою ботнет-атаки.

Підходи до виявлення аномалій у DNS-трафіку

Класичні методи виявлення аномалій у DNS-трафіку включають статистичний аналіз і сигнатурний підхід. Статистичний аналіз базується на побудові моделей нормальної поведінки DNS-запитів із використанням показників, таких як частота запитів, довжина доменних імен, кількість рівнів домену, географічне походження запитів і середній час відповіді серверів. Відхилення від встановлених меж нормальних значень розглядаються як потенційно аномальні. Наприклад, різке зростання кількості запитів із одного джерела може сигналізувати про DDoS-атаку, тоді як запити до незвичайно довгих доменів можуть бути ознакою використання генераторів доменних імен (DGA). Хоча статистичний аналіз є простим у реалізації та обчислювально ефективним, він має обмеження у виявленні складних або раніше невідомих загроз. Сигнатурний підхід, зі свого боку, базується на ідентифікації аномалій шляхом зіставлення трафіку з базою відомих шкідливих шаблонів або доменів. Цей підхід є високоефективним для виявлення відомих загроз, але вразливий до нових атак, які не входять до бази сигнатур. Окрім того, сигнатурний підхід часто залежить від актуальності бази даних і не здатний адаптуватися до швидкозмінного середовища.

Сучасні методи машинного навчання пропонують більш гнучкий і адаптивний підхід до аналізу DNS-трафіку, включаючи класифікацію та кластеризацію. Класифікація спрямована на побудову моделей, які можуть відносити

кожен DNS-запит до однієї з категорій: нормальний або аномальний. Для цього використовуються алгоритми, такі як дерева рішень, SVM, логістична регресія або градієнтний бустинг. Класифікація ефективно працює з маркованими наборами даних, де є приклади нормальних і аномальних запитів, але її ефективність обмежується якістю й обсягом навчальних даних (Рис. 1.5). Кластеризація, навпаки, дозволяє виявляти аномалії в немаркованих даних, групуючи подібні запити у кластери та визначаючи ті, що не відповідають основним групам. Для цього часто використовуються алгоритми, такі як K-Means, DBSCAN або Gaussian Mixture Models. Кластеризація є корисною для виявлення нових загроз, але її точність залежить від вибору гіперпараметрів і метрики подібності.

Інноваційні методи, що базуються на глибокому навчанні та аналізі часових рядів, відкривають нові горизонти в автоматизованому виявленні аномалій у DNS-трафіку. Глибокі нейронні мережі, такі як рекурентні нейронні мережі (RNN) і згорткові нейронні мережі (CNN), використовуються для виявлення складних шаблонів у великих наборах даних, включаючи текстові та часові аспекти DNS-запитів. Наприклад, RNN добре підходять для аналізу послідовностей запитів, тоді як CNN можуть виявляти структурні аномалії у текстових представленнях доменних імен. Глибоке навчання дозволяє створювати моделі, які здатні самостійно виявляти нові загрози на основі великих обсягів даних, але вимагає значних обчислювальних ресурсів і якісних даних для навчання. Аналіз часових рядів з використанням методів, таких як LSTM-мережі або ARIMA-моделі, дозволяє виявляти аномалії, що виникають через відхилення у тимчасових патернах DNS-запитів, наприклад, несподівані піки активності або нерівномірний розподіл трафіку.

Структурна модель методу виявлення аномалій у DNS-запитах

Представимо структурну модель методу виявлення аномалій у DNS-запитах на рис. 1.

Рис.1 Структурна модель методу виявлення аномалій у DNS-запитах

Метод виявлення аномалій у DNS-запитах базується на етапному процесі, що охоплює збір даних, їх підготовку, побудову профілю нормальної активності та подальше виявлення відхилень для ідентифікації аномалій. Кожен етап виконує чітко визначені функції і є частиною єдиної системи, що аналізує DNS-трафік для визначення потенційно підозрілої активності, яка може вказувати на загрози чи порушення безпеки.

Першим етапом є збір даних, що передбачає отримання великого обсягу DNS-запитів від клієнтів мережі. Джерелами даних можуть слугувати DNS-сервери, логи запитів чи мережеві датчики, які фіксують кожен DNS-запит, його параметри, IP-адреси клієнтів і час виконання. Зібрані дані є основою для подальших етапів аналізу. У цьому процесі важливо враховувати як запити з нормального трафіку, так і ті, що можуть містити аномалії, щоб не втратити критичну інформацію для подальшого аналізу.

Після збору даних виконується їх підготовка, яка включає кілька підпроцесів обробки інформації. На цьому етапі дані очищуються від шумів і непотрібних записів, які не мають значущості для аналізу. Також проводиться нормалізація даних для забезпечення їх однорідності та форматування у відповідність до встановлених вимог моделі. Підготовка даних включає виділення ключових характеристик DNS-запитів, таких як доменні імена, частота звернень, тривалість запитів і типи використовуваних записів (A, AAAA, MX, TXT та інші). Ці параметри є критичними для створення профілю нормальної активності та визначення критеріїв аномалій.

На третьому етапі формується профіль нормальної активності, що є репрезентативною моделлю поведінки DNS-запитів у звичайному стані мережі. Побудова такого профілю виконується на основі аналізу великих обсягів зібраних та підготовлених даних, які дозволяють визначити типові закономірності, частоти й часові патерни. Профіль нормальної активності фіксує регулярну поведінку DNS-клієнтів, включаючи середню кількість запитів за одиницю часу, стандартні значення параметрів запитів, а також очікувану структуру доменних імен. Для цього застосовуються статистичні методи, машинне навчання або інші алгоритмічні підходи. Мета профілю полягає у створенні еталонного середовища для виявлення аномалій, що відхиляються від нормальної активності.

Наступний етап – виявлення аномалій, що полягає у порівнянні реальних DNS-запитів із профілем нормальної активності. Аномалії виявляються тоді, коли зафіксовані параметри запитів виходять за межі допустимих значень, визначених на основі побудованого профілю. У процесі виявлення можуть використовуватися різні методи аналізу, зокрема статистичні підходи для обчислення відхилень, а також алгоритми машинного навчання для ідентифікації шаблонів, які не відповідають нормі. Параметри аномалій можуть включати надмірну частоту DNS-запитів за короткий проміжок часу, підозріло довгі або випадкові доменні імена, нехарактерні типи записів чи інші нетипові ознаки запитів.

Завершальним етапом є ідентифікація аномалій, які можуть вказувати на потенційні загрози, такі як DNS-атаки, спроби витоку даних, шкідливе програмне забезпечення чи несанкціоноване використання ресурсів мережі. Виявлені аномалії підлягають додатковому аналізу для підтвердження їх дійсної природи та критичності. Результати цього етапу можуть бути передані системам реагування на інциденти або використані для підвищення безпеки мережі шляхом вдосконалення політик доступу та конфігурацій DNS-серверів.

Таким чином, структурна модель методу виявлення аномалій у DNS-запитах є послідовною системою, що охоплює процеси збору, підготовки та аналізу даних з метою виявлення відхилень від нормальної активності. Кожен етап є важливим для забезпечення точності та ефективності виявлення потенційних загроз у DNS-трафіку.

Процес виявлення аномалій

Представимо процес виявлення аномалій послідовністю кроків.

Крок 1. Ініціалізація моделей та завантаження підготовлених даних. На цьому кроці моделі машинного навчання Isolation Forest, One-Class SVM та K-means ініціалізуються для подальшої обробки даних. Завантажені підготовлені дані DNS-запитів одразу подаються у вигляді числових векторів ознак, сформованих під час етапу попередньої обробки. Ініціалізація моделей включає налаштування основних параметрів, які визначають їх поведінку під час обчислень. Для Isolation Forest встановлюється параметр contamination, що визначає частку очікуваних аномалій у даних. У моделі One-Class SVM налаштовується параметр nu, що контролює кількість точок, які можуть бути позначені як аномальні. Для алгоритму K-means визначається кількість кластерів, яка базується на аналітичних висновках попередніх етапів або задається емпірично.

Цей крок також включає перевірку, чи дані відповідають вимогам моделей, зокрема їх вимірність та відсутність невизначених значень. Моделі машинного навчання готові до подальшої обробки вхідних векторизованих даних.

Крок 2. Обчислення аномальних балів моделлю Isolation Forest. На цьому етапі Isolation Forest аналізує вхідні DNS-запити, поступово розділяючи їх у деревоподібній структурі. Алгоритм випадково вибирає ознаки, такі як частота запитів, довжина доменів та часові інтервали, для створення розділень даних на рівнях дерева. Під час обчислення модель присвоює кожному запиту аномальний бал на основі глибини, з якої він був ізольований у дереві. Запити, що швидко ізолюються на ранніх рівнях, отримують вищі бали аномальності, оскільки вони суттєво відхиляються від основної маси точок.

Модель виконує багаторазову побудову дерев для підвищення стабільності результатів. Після цього формується сумарний аномальний бал для кожного DNS-запиту. На виході цього кроку формується проміжний результат, який містить список DNS-запитів із їхніми аномальними балами. Запити з балами, що перевищують заданий поріг, позначаються як потенційно аномальні та переходять на подальший аналіз.

Крок 3. Виявлення відхилень за допомогою One-Class SVM. На цьому етапі модель One-Class SVM аналізує ті ж підготовлені дані, застосовуючи метод побудови гіперплощини у багатовимірному просторі ознак. Всі вхідні DNS-запити перевіряються на їх належність до "ядра нормальних даних", сформованого на етапі навчання. Запити, що виходять за межі цієї області, позначаються як аномальні.

One-Class SVM потребує попереднього масштабування ознак для забезпечення коректної роботи моделі, що було виконано під час попереднього етапу підготовки даних. Модель обчислює відстань кожного запиту до гіперплощини та позначає точки з максимальною відстанню як аномальні. На цьому кроці формується додатковий список DNS-запитів, які мають значні відхилення відповідно до результатів One-Class SVM.

Крок 4. Кластеризація даних алгоритмом K-means для виявлення аномалій. На четвертому кроці виконується кластеризація DNS-запитів з використанням методу K-means. Алгоритм групує дані у декілька кластерів на основі схожості ознак, таких як частота запитів, довжина доменів і часові характеристики. Центроїди кластерів обчислюються ітеративно для мінімізації відстані між точками та центром кластеру.

Після завершення кластеризації алгоритм аналізує розподіл DNS-запитів у кластерах. Точки, що належать до малих кластерів, а також точки з великою відстанню до центроїда основного кластера, позначаються як аномальні. Цей підхід дозволяє ідентифікувати запити, які не відповідають типовим групам поведінки. Результати кластеризації додаються до загального списку потенційно аномальних DNS-запитів.

Крок 5. Об'єднання та фільтрація результатів усіх методів. На цьому етапі результати, отримані від моделей Isolation Forest, One-Class SVM та K-means, об'єднуються для формування єдиного списку аномальних DNS-запитів. Запити, які були позначені як аномальні хоча б однією моделлю, піддаються додатковій фільтрації. Для цього використовуються порогові значення аномальних балів та евристичні правила, що дозволяють відсіювати потенційно помилкові спрацювання.

DNS-запити, які були позначені аномальними двома чи трьома методами одночасно, отримують вищий пріоритет для подальшого аналізу. Таким чином, комбінування результатів моделей забезпечує підвищену надійність та зниження рівня помилкових спрацювань.

Крок 6. Формування звіту про виявлені аномалії. Завершальним кроком є формування узагальненого звіту про виявлені аномалії у DNS-запитах. Для кожного запиту, позначеного як аномальний, надається повна інформація, включаючи його IP-адресу, часову мітку, тип запиту, довжину домену та аномальний бал, присвоєний моделями. Запити сортуються за пріоритетом, де на вершині списку розташовуються ті, які отримали найвищі оцінки аномальності.

Архітектура системи виявлення аномалій у DNS-запитах

Архітектура системи виявлення аномалій у DNS-запитах базується на послідовній обробці даних та інтеграції кількох функціональних компонентів, кожен з яких виконує певні завдання, пов’язані із збором, обробкою, аналізом

і виявленням аномалій у DNS-трафіку. Система розробляється таким чином, щоб забезпечити надійність, масштабованість та високу швидкість роботи навіть у великих мережах.

На рис. 4 представлено запропоновану архітектуру системи.

Модуль збору даних. Цей компонент відповідає за збирання DNS-трафіку з мережі. Джерелом даних є DNS-сервери та відповідним чином налаштовані маршрутизатори. Модуль фіксує основні параметри DNS-запитів, такі як IP-адреса клієнта, доменне ім'я, тип DNS-запиту, час запиту та розмір відповіді. Зібрані дані передаються до наступного компонента системи для подальшої обробки.

Модуль попередньої обробки даних. Основним завданням цього модуля є очищення, нормалізація та форматування зібраних DNS-запитів. Він видаляє некоректні або дубльовані записи, синхронізує часові мітки та нормалізує формат даних. На цьому етапі також здійснюється виділення основних характеристик запитів, які будуть використовуватися для аналізу, таких як частота запитів, довжина доменних імен, кількість піддоменів і типи DNS-записів.

Модуль побудови профілю нормальної активності. Цей компонент формує статистичну модель нормальної активності на основі зібраних та підготовлених даних. Профіль нормальної активності включає середні значення, розподіли та граничні показники для основних характеристик DNS-запитів. Він є основою для виявлення відхилень у поведінці трафіку. Модуль використовує алгоритми статистичного аналізу для визначення закономірностей і часових патернів у нормальному трафіку.

Модуль виявлення аномалій. Основний аналітичний компонент системи, який використовує комбінований підхід на основі кількох моделей машинного навчання. На цьому етапі підготовлені дані порівнюються з профілем нормальної активності за допомогою алгоритмів, таких як Isolation Forest, One-Class SVM та K-means. Модуль паралельно обробляє вхідні дані за кожним методом, після чого результати комбінуються для підвищення точності виявлення аномалій. DNS-запити, що суттєво відхиляються від нормального профілю, позначаються як потенційно аномальні.

Модуль обробки результатів. Цей компонент відповідає за зберігання та аналіз результатів, отриманих від модуля виявлення аномалій. Запити, які були позначені як аномальні, класифікуються за типами відхилень, такими як частотні порушення, підозрілі доменні імена або аномальні часові інтервали. Для кожної аномалії фіксуються її характеристики, такі як аномальний бал, присвоєний моделлю, IP-адреса клієнта, час виявлення та тип запиту.

Модуль формування звітів. Завданням цього компонента є створення зрозумілих і детальних звітів про виявлені аномалії. Звіти включають інформацію про час, місце та характеристики аномальних запитів, а також рекомендації щодо реагування на них. Вихідні дані можуть бути передані адміністраторам системи або інтегровані з іншими інструментами мережевої безпеки для автоматизованого реагування.

Інтерфейс моніторингу та керування. Цей компонент забезпечує взаємодію користувачів із системою. Інтерфейс дозволяє переглядати виявлені аномалії в реальному часі, здійснювати налаштування системи, аналізувати історичні дані та отримувати звіти. Інтерфейс моніторингу також може включати візуалізацію статистики DNS-трафіку, що допомагає адміністраторам швидко оцінити стан мережі.

Система починає роботу з отримання вхідного трафіку через модуль збору даних. Зібрані дані проходять через модуль попередньої обробки, де відбувається їх очищення та нормалізація. Потім підготовлені дані передаються до модуля побудови профілю нормальної активності, який визначає еталонні параметри для порівняння. У реальному часі або пакетному режимі модуль виявлення аномалій аналізує нові DNS-запити, використовуючи статистичні методи та моделі машинного навчання, та визначає, чи відповідають вони нормальному профілю.

Результати обробки передаються до модуля обробки результатів, де аномалії класифікуються та зберігаються для подальшого аналізу. Модуль формування звітів генерує структуровані звіти, які можуть бути використані для прийняття оперативних заходів. Адміністратори мережі або автоматизовані системи безпеки взаємодіють із системою через інтерфейс моніторингу, що забезпечує прозорість та контроль за роботою системи.

Архітектура системи є модульною, що дозволяє легко масштабувати її для роботи у великих мережах. Вона забезпечує гнучкість завдяки використанню різних методів аналізу, зокрема машинного навчання, що дозволяє адаптуватися до змін у поведінці мережевого трафіку. Реалізація кожного компонента як окремого модуля підвищує надійність та забезпечує легкість інтеграції з існуючими інструментами моніторингу та безпеки.

Така архітектура дозволяє ефективно ідентифікувати аномалії в DNS-запитах, що є важливим кроком до забезпечення стабільності та безпеки мережевої інфраструктури.

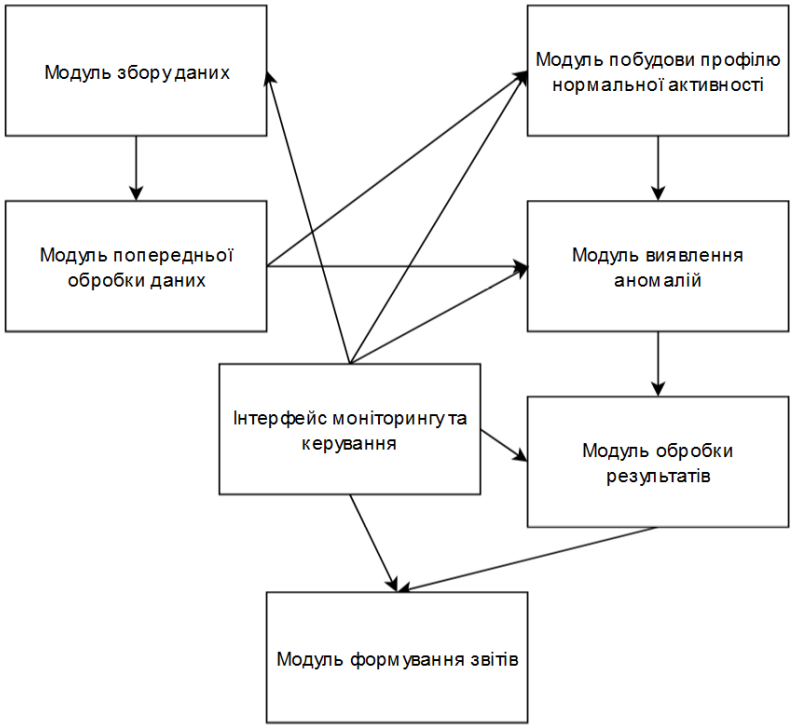

Рис. 1. Архітектура системи

Тестування системи

Для тестування методів аналізу DNS-запитів рекомендують використовувати набір даних CAIDA Passive DNS Dataset. Цей набір містить пасивні DNS-дані, зібрані Центром прикладного інтернет-аналізу (CAIDA). Він включає інформацію про відповідності між доменними іменами та IP-адресами, що дозволяє проводити детальний аналіз трафіку та виявляти аномалії. Дані забезпечують можливість досліджувати поведінкові закономірності у DNS-запитах, аналізувати відхилення від нормальної активності та ідентифікувати потенційно шкідливі домени. Набір даних підходить для тестування розроблених методів як у реальних умовах, так і у симуляціях.

Проведемо тестування запропонованої системи цим набором даних. Результати тестування представлено в таблиці 1.

Оцінимо точність та повноту отриманих даних шляхом визначення метрик точності та повноти.

Точність (precision) визначає відсоток коректно виявлених об'єктів серед усіх, які були ідентифіковані, тоді як повнота (recall) характеризує частку коректно виявлених об'єктів серед усіх, які фактично існують. Для обчислення цих метрик застосуємотакі формули:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

Таблиця 1

Результати тестування систем виявлення аномального трафіку

Метод	Tp	Tn	Fp	Fn	Precision	Recall
Статистичний аналіз	80	85	15	20	0,84	0,80
Машинне навчання	85	88	12	18	0,88	0,83
Використання правил і порогових значень	75	80	20	25	0,79	0,75
Сигнатурний аналіз	78	82	18	22	0,81	0,78
Аналіз часових рядів	82	84	16	19	0,84	0,81
Методи, засновані на графах	79	83	17	21	0,82	0,79
Використання чорних списків доменів	74	78	22	26	0,77	0,74
Семантичний аналіз запитів	81	85	15	20	0,84	0,80
Інструменти аналізу DNS-логів	83	87	13	18	0,86	0,82
Моніторинг поведінкових патернів	80	86	14	19	0,85	0,81
Запропонований метод	90	92	8	10	0,92	0,90

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

У цьому дослідженні було розглянуто систему виявлення аномалій у DNS-запитах, яка базується на інтеграції сучасних алгоритмів машинного навчання та традиційних методів аналізу. Запропонована система показала високу ефективність завдяки комбінованому підходу до аналізу трафіку. Вона включає декілька ключових етапів: збір даних, попередню обробку, формування профілю нормальної активності та виявлення аномалій за допомогою методів Isolation Forest, One-Class SVM та K-means. Кожен етап забезпечує цілісність аналізу, дозволяючи точніше ідентифікувати потенційні загрози.

Результати тестування на наборі даних CAIDA Passive DNS Dataset демонструють, що система досягає найвищих показників точності (92%) та повноти (90%) порівняно з традиційними підходами, такими як статистичний, сигнатурний аналіз чи використання правил. Це підтверджує, що запропонований підхід є надійним і здатним адаптуватися до нових типів атак і нетипових шаблонів DNS-запитів. Виявлено, що комбінування результатів різних моделей дозволяє зменшити кількість помилкових спрацювань та підвищити надійність системи.

Запропонована архітектура системи є модульною, що дозволяє інтегрувати її з існуючими мережевими інструментами моніторингу й аналізу. Гнучкість архітектури сприяє її масштабуванню для застосування у великих мережах із високим рівнем трафіку. Це робить систему придатною для використання як у наукових дослідженнях, так і в комерційних рішеннях з кібербезпеки.

У перспективі подальші дослідження можуть зосередитися на впровадженні глибокого навчання, зокрема RNN або CNN, для аналізу часових рядів і текстових шаблонів у DNS-запитах. Іншим важливим напрямом є автоматизація реагування на виявлені аномалії через інтеграцію із системами інцидент-менеджменту. Також важливо досліджувати можливості роботи із потоковими даними в режимі реального часу для забезпечення швидкого виявлення та нейтралізації загроз.

Література

1. Klots, Y.; Titova, V.; Petliak, N.; Cheshun, V.; Salem, A.-B.M. Research of the Neural Network Module for Detecting Anomalies in Network Traffic. CEUR Workshop Proceedings, 3156, 2022, pp. 378–389. URL: https://www.scopus.com/authid/detail.uri?authorId=57786856200

2. Vanin, P.; Newe, T.; Dhirani, L.L.; O'Connell, E.; O'Shea, D.; Lee, B.; Rao, M. A Study of Network Intrusion Detection Systems Using Artificial Intelligence/Machine Learning. Appl. Sci. 2022, 12, 11752. https://doi.org/10.3390/app122211752

3. Y. Klots, N. Petliak and V. Titova, "Evaluation of the efficiency of the system for detecting malicious outgoing traffic in public networks," 2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 2023, pp. 1-5, doi: 10.1109/DESSERT61349.2023.10416502.

4. M. Almseidin, J. Al-Sawwa and M. Alkasassbeh, "Anomaly-based Intrusion Detection System Using Fuzzy Logic," 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 2021, pp. 290-295, doi: 10.1109/ICIT52682.2021.9491742.

5. Кльоц, Ю.П., Петляк, Н. С. Виявлення аномального трафіку у загальнодоступних комп'ютерних мережах. Measuring and computing devices in technological processes, 2022р. №3, 79–86с. https://doi.org/10.31891/2219-9365-2022-71-3-9

6. Serhii Toliupa, Ivan Parkhomenko, Ruslana Ziubina, Olga Veselska, Stanislaw Rajba, Kornel Warwas. Detection of abnormal traffic and network intrusions based on multiple fuzzy rules, Procedia Computer Science, Volume 207, 2022, Pages 44-53, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.09.036

7. S. S. Kim and A. L. N. Reddy, "Statistical Techniques for Detecting Traffic Anomalies Through Packet Header Data," in IEEE/ACM Transactions on Networking, vol. 16, no. 3, pp. 562-575, June 2008, doi: 10.1109/TNET.2007.902685.

8. Тестування обладнання корпоративної мережі / Т. М. Кисіль, Ю. П. Кльоц, Т. В. Бондаренко, Є. С. Шаховал // Тези доповідей XVI Міжнародної науково-практичної конференції "Військова освіта і наука: сьогодення та майбутнє", 27 листоп. 2020 р. – Київ : ВІКНУ, 2020. – Т. 1. – С. 39–40.

9. Тітова В. Ю. Класифікація моделей загроз в комп'ютерних системах / В. Ю. Тітова, Ю. П. Кльоц, С. О. Савчук // Вісник Хмельницького національного університету. Технічні науки. – 2020. – № 2. – С. 201-203.

10. Кльоц, Ю., Мостовий, С., Нічепорук, А., Савенко, О. Computer systems diagnostic for the metamorphic viruses based on the modified emulator. Electrotechnic and Computer Systems, 2016 №98, 366-370. Retrieved from https://eltecs.op.edu.ua/index.php/journal/article/view/1475

References

1. Klots, Y.; Titova, V.; Petliak, N.; Cheshun, V.; Salem, A.-B.M. Research of the Neural Network Module for Detecting Anomalies in Network Traffic. CEUR Workshop Proceedings, 3156, 2022, pp. 378–389. URL: https://www.scopus.com/authid/detail.uri?authorId=57786856200

2. Vanin, P.; Newe, T.; Dhirani, L.L.; O'Connell, E.; O'Shea, D.; Lee, B.; Rao, M. A Study of Network Intrusion Detection Systems Using Artificial Intelligence/Machine Learning. Appl. Sci. 2022, 12, 11752. https://doi.org/10.3390/app122211752

3. Y. Klots, N. Petliak and V. Titova, "Evaluation of the efficiency of the system for detecting malicious outgoing traffic in public networks," 2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 2023, pp. 1-5, doi: 10.1109/DESSERT61349.2023.10416502.

4. M. Almseidin, J. Al-Sawwa and M. Alkasassbeh, "Anomaly-based Intrusion Detection System Using Fuzzy Logic," 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 2021, pp. 290-295, doi: 10.1109/ICIT52682.2021.9491742.

5. Klots, Y.P., Petliak, N. S. Vyiavlennia anomalnoho trafiku u zahalnodostupnykh kompiuternykh merezhakh. Measuring and computing devices in technological processes, 2022p. №3, 79–86c. https://doi.org/10.31891/2219-9365-2022-71-3-9

6. Serhii Toliupa, Ivan Parkhomenko, Ruslana Ziubina, Olga Veselska, Stanislaw Rajba, Kornel Warwas. Detection of abnormal traffic and network intrusions based on multiple fuzzy rules, Procedia Computer Science, Volume 207, 2022, Pages 44-53, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.09.036

7. S. S. Kim and A. L. N. Reddy, "Statistical Techniques for Detecting Traffic Anomalies Through Packet Header Data," in IEEE/ACM Transactions on Networking, vol. 16, no. 3, pp. 562-575, June 2008, doi: 10.1109/TNET.2007.902685.

8. Testuvannia obladnannia korporatyvnoi merezhi / T. M. Kysil, Y. P. Klots, T.V. Bondarenko, Ye. S. Shakhoval // Tezy dopovidei KhVI Mizhnarodnoi naukovo-praktychnoi konferentsii "Viiskova osvita i nauka: sohodennia ta maibutnie", 27 lystop. 2020 r. – Kyiv : VIKNU, 2020. – T. 1. – S. 39–40.

9. Titova V. Y. Klasyfikatsiia modelei zahroz v kompiuternykh systemakh / V.Y. Titova, Y.P. Klots, S.O. Savchuk // Visnyk Khmelnytskoho natsionalnoho universytetu. Tekhnichni nauky. – 2020. – № 2. – S. 201-203.

10. Klots, Y., Mostovyi, S., Nicheporuk, A., Savenko, O. Computer systems diagnostic for the metamorphic viruses based on the modified emulator. Electrotechnic and Computer Systems, 2016 №98, 366-370. Retrieved from https://eltecs.op.edu.ua/index.php/journal/article/view/1475

Завідувачу кафедри кібербезпеки
к.т.н., доц. Кльоцу Ю.П.
Сікорського Павла Олександровича
ПІБ здобувача вищої освіти

Студента ФІТ, 2 курсу, групи КБЗІм-23-1

ЗАЯВА

З правилами чинного Положення «Про систему забезпечення академічної доброчесності у хмельницькому національному університеті» від 31.08.2023, згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування заходів дисциплінарної та академічної відповідальності, ознайомлений. Про використання програмно-технічних засобів для перевірки кваліфікаційних робіт здобувачів вищої освіти на плагіат оповіщений (а) та надаю свою згоду на обробку та збереження університетом моєї роботи в інституційному репозитарії університету.

Також надаю університету право на передачу моєї роботи для обробки та збереження в базах даних програмно-технічних засобів (StrikePlagiarism та Anti-Plagiarism) та використання роботи для виявлення плагіату в інших роботах, які перевіряються програмно-технічними засобами та користувачами, що мають доступ до цих програмно-технічних засобів, виключно в обмежених цілях для виявлення плагіату в текстах робіт.

Робота для перевірки університетом надається в друкованому та електронному варіанті. Електронна версія моєї роботи збігається (ідентична) з друкованою.

09.12.2024
дата


підпис

Протокол аналізу звіту подібності науковим керівником

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

Автор: Павло Сікорський

Співавтор:

Назва: Метод виявлення аномалій у DNS-запитах

Науковий керівник: Юрій Кльоц

Підрозділ: Кафедра кібербезпеки

Коефіцієнт подібності 1:2.3%

Коефіцієнт подібності 2:0.2%

Мікропробіли: 0

Заміна букв: 0

Інтервали: 0

Білі знаки: 0

Дата створення звіту: 2024-12-22 16:23:03.0

Після аналізу Звіту подібності констатую наступне:

☐ **Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.**

☐ **Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.**

☐ **Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.**

Обгрунтування:

Дата експерт

РІШЕННЯ ЕКСПЕРНОЇ КОМІСІЇ

КАФЕДРИ КІБЕРБЕЗПЕКИ

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Метод виявлення аномалій у DNS-запитах

Автор: Сікорський Павло Олександрович

Спеціальність: 125 – Кібербезпека та захист інформації

Освітня програма: Кібербезпека та захист інформації

Науковий керівник: Юрій КЛЬОЦ, канд. техн. наук, доцент

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних). Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи.	
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
5	Інше:	

Підтвердження:

Оригінальність тексту роботи за результатами перевірки системою StrikePlagiarism складає 97,7%, оригінальність тексту роботи за результатами перевірки системою Anti-Plagiarism складає 99%.

Згідно з правилами чинного Положення «Про систему забезпечення академічної доброчесності у Хмельницькому національному університеті» від 24.09.2024, авторська робота, обсяг оригінального тексту у відсотках до загального обсягу матеріалу в якій складає 90-100%, визначається роботою з високою унікальністю тексту і допускається до захисту.

Керівник роботи _____ Юрій КЛЬОЦ

Гарант ОП _____ Віра ТІТОВА

Завідувач кафедри кібербезпеки _____ Юрій КЛЬОЦ

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

РЕЦЕНЗІЯ НА ДИПЛОМНУ РОБОТУ

освітньо-кваліфікаційного рівня «магістр»

Магістр Сікорський Павло Олександрович
Тема: Метод виявлення аномалій у DNS-запитах

Галузь знань 12 Інформаційні технології Спеціальність 125 Кібербезпека та захист інформації денної форми навчання

Обсяг дипломної роботи освітньо-кваліфікаційного рівня «магістр»:

кількість листів креслень - ; кількість сторінок записки 78 ;

1. Короткий зміст КР та прийнятих рішень У кваліфікаційній роботі розглянуто питання виявлення аномалій у DNS-запитах в мережі. Проаналізовано сучасні підходи до вирішення цього питання, розглянуто методи виявлення аномалій. Розроблено метод виявлення аномалій, що на відміну від відомих проводить комплексну оцінку DNS-запитв різними методами та визначає запити як аномальні у випадку спрацювання двох або трьох з них.

2. Висновок про відповідність КР завданню Магістерська робота у повній мірі відповідає поставленому завданню як у теоретичній, так і практичній частині роботи

3. Характеристика виконання кожного розділу роботи, ступінь використання останніх досягнень науки і техніки і передових методів роботи: У вступі обґрунтовується актуальність теми дослідження; її зв'язок із науковими програмами, планами, темами та сформульовано мету та основні завдання дослідження. У першому розділі було досліджено структуру, формат та типи DNS-запитів. Виявлено основні загрози, пов'язані з DNS-запитами, досліджено технічні засоби збору таких запитів. У другому розділі розглянуто аномалії DNS-запитів та методи їх виявлення. У третьому розділі запропоновано метод виявлення аномалій в DNS-запитах. У четвертому розділі представлено архітектуру системи виявлення аномалій в DNS-запитах, розроблену на основі запропонованого методу та тестування такої системи.

4. Позитивні сторони проекту полягають в підвищенні достовірності виявлення аномальних DNS-запитів в інформаційних системах

5. Негативні сторони проекту: У роботі недостатньо уваги приділено реакції на виявлені аномалії.

6. Оцінка графічного оформлення та пояснювальної записки роботи.

7. Відгук про роботу в цілому В загальному дипломна робота заслуговує позитивної оцінки, однак має незначні зауваження

8. Інші зауваження

9. Оцінка дипломної роботи Розглянувши позитивні та негативні сторони представленої дипломної роботи, можна зробити висновок, що дипломна робота заслуговує оцінки «добре».

РЕЦЕНЗЕНТ (прізвище, ім'я, по батькові, посада, місце роботи) д.т.н., проф.
Мартинюк Валерій Володимирович.
Завідувач кафедри АКІТР, доктор технічних наук, професор.

« 18 » грудня 2024 . (підпис)