DOI 10.31891/2307-5732-2024-345-6-22
УДК 004.77

КЛЬОЦ ЮРІЙ
Хмельницький національний університет
https://orcid.org/0000-0002-3914-0989
e-mail: klots@khmnu.edu.ua
МОСТОВИЙ СЕРГІЙ
Хмельницький національний університет
https://orcid.org/0000-0002-9505-3206
e-mail: sprmostovuy@gmail.com
СІКОРСЬКИЙ ПАВЛО
Хмельницький національний університет
e-mail: sikorskiyp@khmnu.edu.ua
ОСТАПЧУК ІГОР
Хмельницький національний університет
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.

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

ANOMALY DETECTION SYSTEM IN DNS QUERIES

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. Anomalous DNS queries pose a serious threat to network security and stability as they can be an indicator of cyber attacks or malicious activity. One of the most common threats is the use of DNS to carry out DDoS attacks, in particular through the DNS amplification mechanism. 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. Innovative methods based on deep learning and time series analysis open new horizons in automated detection of anomalies in DNS traffic. Deep neural networks, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), are used to detect complex patterns in large data sets, including the textual and temporal aspects of DNS queries.

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.

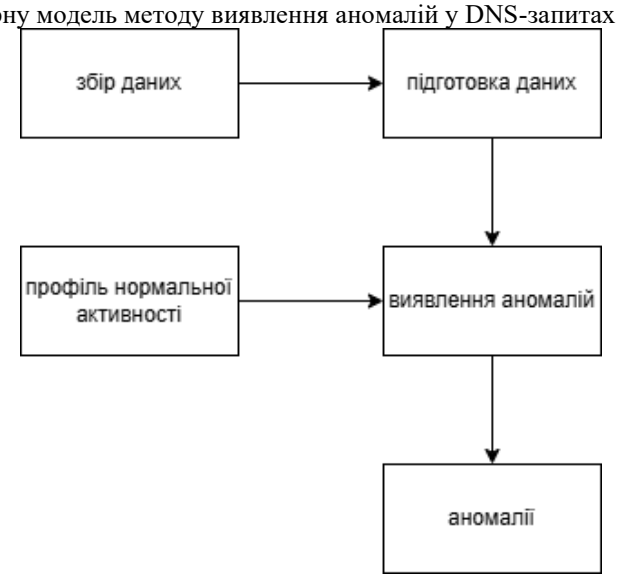

Рис.1 Структурна модель методу виявлення аномалій у DNS-запитах

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.

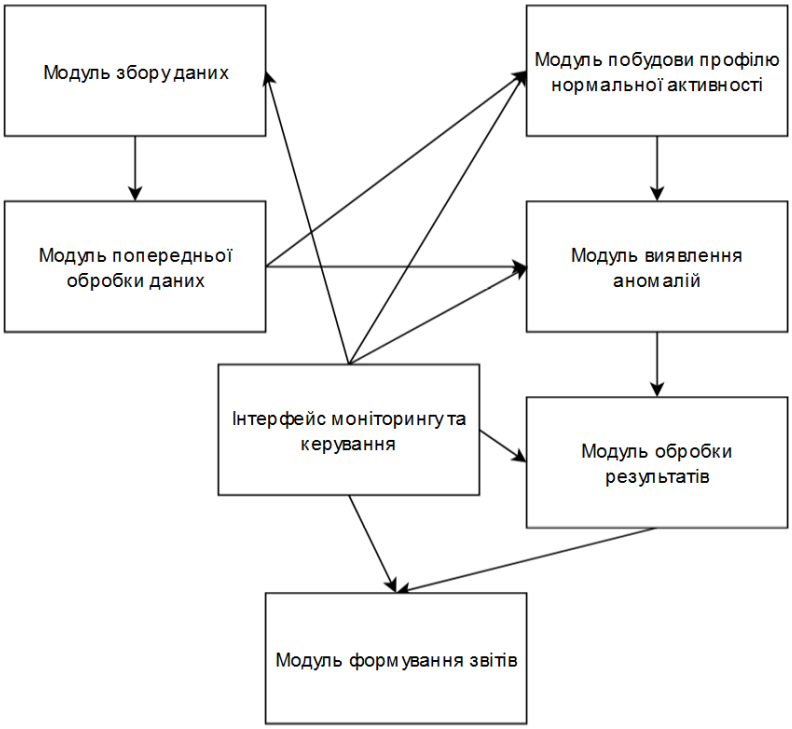

Рис. 1. Архітектура системи

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.

Літертура

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