НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

До захисту допущено
В.о. завідувача кафедри

_____ **Микола ГРАЙВОРОНСЬКИЙ**
(підпис)

«_____»_____2021 р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною программою «Системи, технології та математичні методи кібербезпеки»
спеціальності: 125 «Кібербезпека»

на тему: Виявлення центрів керування шкідливого ПЗ за допомогою аналізу DNS трафіку

Виконав (-ла): здобувач вищої освіти **IV** курсу, групи ФБ-73

Божко Анастасія Юріївна
(підпис)

Керівник доцент Барановський Олексій Миколайович
(підпис)

Рецензент _____ _____
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище, ім'я, по батькові) (підпис)

Засвідчую, що у цій дипломній роботі немає запозичень з праць інших авторів без відповідних посилань.
Здобувач вищої освіти _____
(підпис)

Київ - 2021 року

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

Рівень вищої освіти – перший (бакалаврський)
Спеціальність – 125 «Кібербезпека»
Освітньо-професійна програма «Системи, технології та математичні методи кібербезпеки»

ЗАТВЕРДЖУЮ
В.о. завідувача кафедри
_____ Микола ГРАЙВОРОНСЬКИЙ
(підпис)
«___»_____2021 р.

ЗАВДАННЯ
на дипломну роботу здобувачу вищої освіти

Божко Анастасії Юріївні
(прізвище, ім'я, по батькові)

1. Тема роботи Виявлення центрів керування шкідливого ПЗ за допомогою аналізу DNS трафіку,
керівник роботи Барановський Олексій Миколайович, доцент,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від « » _____ 2021 р. №

2. Термін подання здобувачем вищої освіти роботи 07 червня 2021 р.

3. Вихідні дані до роботи

1) Попередні дослідження.

2) Методи машинного начання.

4. Зміст роботи

1) Вивчити основні ознаки центрів керування та методи маскування в DNS трафіку, що характерні для даної проблеми.

2) Проаналізувати різні методи виявлення центрів керування та більш детально розглянуті методи виявлення алгоритмів генерації доменів.

3) Проаналізувати які характеристики та алгоритми більш доцільні для класифікації шкідливих доменних імен, які генеруються для центрів керування.

4) Вибрати комплекс характеристик та алгоритми для вирішення даної проблеми.

5) Провести дослідження та аналіз отриманих результатів.

6) Побудувати комплекс методів виявлення алгоритмів генерації доменів.

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо)

Презентація

6. Дата видачі завдання 5.10.2020

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів дипломної роботи	Примітка
1	Отримання завдання	05.10.2020	
2	Збір інформації	07.12.2020	
3	Дослідження предметної області	10.03.2021	
4	Дослідження існуючих проблем	22.03.2021	
5	Розробка плану роботи	12.04.2021	
6	Вибір методів порівняльного аналізу	19.04.2021	
7	Визначення основних ознак та вибір критеріїв порівняння	26.04.2021	
8	Проведення дослідження	04.05.2021	
9	Обґрунтовування результатів	18.05.2021	
10	Оформлення дипломної роботи	25.05.2021	
11	Отримання допуску до захисту	01.06.2021	

Здобувач вищої освіти _____ Анастасія БОЖКО
(підпис) (Власне ім'я, ПРІЗВИЩЕ)

Керівник роботи _____ Олексій БАРАНОВСЬКИЙ
(підпис) (Власне ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Обсяг роботи становить 72 сторінки, 31 рисунки, 7 таблиць та 43 джерел інформації.

Завданням роботи є вивчення основних ознаки центрів керування та методів маскування в DNS трафіку, що характерні для даної проблеми; аналіз різних методів виявлення центрів керування та більш детальний огляд методів виявлення алгоритмів генерації доменів; аналіз які характеристики та алгоритми більш доцільні для класифікації шкідливих доменних імен, які генеруються для центрів керування; дослідження та аналіз отриманих результатів; побудова комплексу методів виявлення алгоритмів генерації доменів.

Метою даної роботи є дослідження методів виявлення центрів керування в DNS трафіку, щоб запобігти зловмисне втручання в систему, яке має на меті викрадення особистих даних та поширення шкідливого програмного забезпечення.

Об’єктом дослідження є DNS трафік.

Предметом дослідження є методи машинного навчання класифікації шкідливих доменних імен в DNS трафіку.

Актуальність роботи полягає в тому, що на сьогоднішній день достатня кількість спеціалістів нехтує належним аналізом DNS трафіку, провокуючи зловмисників застосовувати центри керування для шкідливого ПЗ.

Методами дослідження є аналіз інформаційних джерел, сучасних публікацій, які мають відношення до теми роботи, а також експериментальні дослідження із використанням машинного навчання.

Новизна роботи – запропоновано нову модель виявлення центрів керування, яка містить в собі комплекс методів, з метою підвищення ефективності виявлення даної проблеми.

Практичне застосування дослідження полягає в можливості використання комплексу для виявлення алгоритмів генерації доменів в майбутньому, який може бути інтегрованим у антивірусні програми, з метою виявлення центрів керування ПЗ.

Ключові слова: dns, центри керування, шкідливе програмне забезпечення, швидкий потік, алгоритми генерації доменів, dns тунелювання.

ABSTRACT

The work consists of 72 pages, 31 figures, 7 tables and 43 references.

The task of the work is studying the main features of command centers and methods of masking in DNS-traffic, which are characteristic of select problem; different methods analysis of command centers detection and a more detailed review of domain generation algorithms detection methods; analysis of characteristics and algorithms to classificate malicious domain names generated for control centers; research and analysis of the obtained results; construction a methods combination for detecting generation domain algorithms.

The purpose of this work is the study the methods of detecting command centers in DNS traffic, to prevent malicious interference in the system, which aims to steal data and the spread of maltware.

The object of study is DNS traffic.

The subject of the research is the machine learning methods of malicious domain names classification in DNS-traffic.

The relevancy of the work is in the fact that today a sufficient number of specialists neglect of analysis DNS-traffic, provoking attackers use command centers for malware.

The research methods are the analysis of information sources, modern publications related to the topics of work, as well as experimental research using machine learning.

The scientific novelty is a new proposed model of command center detection, which includes a set of methods to increase the efficiency of detecting these problems.

The practical application of the research is in the possibilities of using the complex to detect generation domain algorithms in the future, which can be integrated into anti-virus programs for identification malware command centers.

Keywords: dns, command centers, malware, fast flux, generation domain algorithms, dns tunneling.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ASN – Autonomous System Number

BGP – Border Gateway Protocol

C&C – Command and Control

DDoS – Distributed Denial of Service

DGA – Domain Generation Algorithm

DNS – Domain Name System

DT – Decision Trees

FQDN – Fully Qualified Domain Name

GRU – Gated Recurrent Unit

IP – Internet Protocol

KNN – K-Nearest Neighbor

LR – Logical Regression

LSTM – Long Short-Term Memory

NB – Naïve Bayes

NXDomain – Non-existent domain

PRNG – Pseudorandom number generator

PSD – Power Spectral Density

RF – Random Forest

RNN – Recurrent Neural Networks

RR – Resource Record

SVM – Support Vector Machines

TLD – Top-level Domain

TTL – Time-to-live

P2P – Peer-to-peer

ВСТУП

Система доменних імен (DNS) є найважливішою ланкою Інтернет-інфраструктури, який в загальних випадках використовується для перетворення доменних імен на IP-адреси. Грубо кажучи, DNS являє собою «телефонну інтернет-книгу», без якої довелося б занотовувати у пам’яті випадкові числа у вигляді IP-адрес, а для цього людський мозок недостатньо підготовлений.

На сьогодні більшість мережевих служб та програм покладаються на DNS систему. Протокол DNS також є одним з небагатьох протоколів застосунків, яким дозволено перетинати периметри мереж організацій, а отже, трафік неможливо блокувати належним чином. Чимала кількість спеціалістів із безпеки не приділяють достатньої уваги поширеності DNS порушень злочинцями, тому не цензурують DNS трафік на існування загроз, затим що не вважають запити, які надіслані через протокол DNS і порт 53, зловмисним та загрозою вилучення даних, або ж затим що велика кількість DNS трафіку, тому дослідження даного трафіку вимагає чималих ресурсів. [1]

Отже у такий спосіб, зловмисники вдало використовують привілеї DNS системи, яка забезпечує канали для злочинців для ініціювання неприступних каналів зв'язку між зараженими машинами та центрами керування (C&C). Центри керування – це централізовані машини, які керують зараженими комп’ютерами або пристроями та віддалено вказують їм ініціювати шкідливі дії. Використовуючи сервери C&C, зловмисники можуть одночасно здійснювати широкомасштабні атаки на тисячі заражених комп'ютерів і приховувати свої сліди. [2]

До C&C входять: швидкий потік, алгоритми генерації доменів (DGA) та їх комбінації, а також використання протоколу DNS для тунелювання.

Існує декількома способами зараження комп’ютерів:

- Через фішингові електронні листи, які ззовні можуть майже не відрізняються від звичайного листа. Тому користувач може легко перейти за посиланням на шкідливий веб-сайт або відкрити вкладення, яке виконує зловмисний код.
- Через дірки в безпеці в плагінах браузера.
- Завантаження шкідливого програмне забезпечення.
- Зі шкідливим кодом, який надходить на зовнішні пристрої, наприклад, USB-флешка.
- Через інше заражене програмне забезпечення. [1]

Шкідливі ПЗ з центрами керування являють собою велику небезпеку для організацій з потенційно серйозними операційними, фінансовими та репутаційними ризиками. Як правило, експлуатуючи C&C, зловмисники мають на меті:

- Викрадення інформації: Конфіденційну інформацію, наприклад, облікові записи, можна дублікують або перемістити на сервер зловмисника.
- Вимкнення: Зловмисник може вимкнути один або декілька комп’ютерів, або навіть вимкнути всю цільову мережу.
- Перезавантаження: Машини, які використовуються, можуть раптово з’явитися або можуть перейти до стану виключення та перезавантаження, що врешті-решт призведе до порушення типових поточних завдань.
- Розповсюдження відмови в обслуговуванні: DDoS-атаки переповнюють сервер численними запитами. Встановивши бот-мережі, злочинець може надати команду ботам відправити запит на цільову IP-адресу, створюючи скупчення для запиту на цільовий сервер або цільову адресу. [3]

Метою роботи є дослідження методів виявлення центрів керування в DNS трафіку, щоб запобігти зловмисне втручання в систему, яке має на меті викрадення особистих даних та поширення шкідливого програмного забезпечення.

Завдання дослідження:

1) Вивчити основні ознаки центрів керування та методи маскування в DNS трафіку, що характерні для даної проблеми.

2) Проаналізувати різні методи виявлення центрів керування та більш детально розглянути методи виявлення алгоритмів генерації доменів.

3) Проаналізувати які характеристики та алгоритми більш доцільні для класифікації шкідливих доменних імен, які генеруються для центрів керування.

4) Вибрати комплекс характеристик та алгоритми для вирішення даної проблеми.

5) Провести дослідження та аналіз отриманих результатів.

6) Побудувати комплекс методів виявлення алгоритмів генерації доменів.

Об’єктом дослідження є DNS трафік.

Предметом дослідження є методи машинного навчання класифікації шкідливих доменних імен в DNS трафіку.

Актуальність роботи полягає в тому, що на сьогоднішній день достатня кількість спеціалістів нехтує належним аналізом DNS трафіку, провокуючи зловмисників застосовувати центри керування для шкідливого ПЗ.

Методами дослідження є аналіз інформаційних джерел, сучасних публікацій, які мають відношення до теми роботи, а також експериментальні дослідження із використанням машинного навчання.

Новизна роботи – запропоновану нову модель виявлення центрів керування, яка містить в собі комплекс методів, з метою підвищення ефективності виявлення даної проблеми.

Практичне застосування дослідження полягає в можливості використання комплексу для виявлення алгоритмів генерації доменів в майбутньому, який може бути інтегрованим у антивірусні програми, з метою виявлення центрів керування ПЗ.

Апробація результатів роботи та публікації: Робота була опублікована у збірнику матеріалів XIX Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених «Теоретичні і прикладні проблеми фізики, математики та інформатики».

1 ОСНОВНІ ПОНЯТТЯ C&C

1.1 Визначення та структура

Центр керування (C&C) – це централізована машина, що керується кіберзлочинцем, яка використовується для надсилання команд системам, скомпрометованим шкідливим програмним забезпеченням, та отримання викрадених даних із цільової мережі машин.[4]

За архітектурою центрів керування можна виділити чотири види:

1. Централізована модель – вид мережевої моделі, в якій всі клієнти взаємодіють з центральною системою, яка є центром керування.

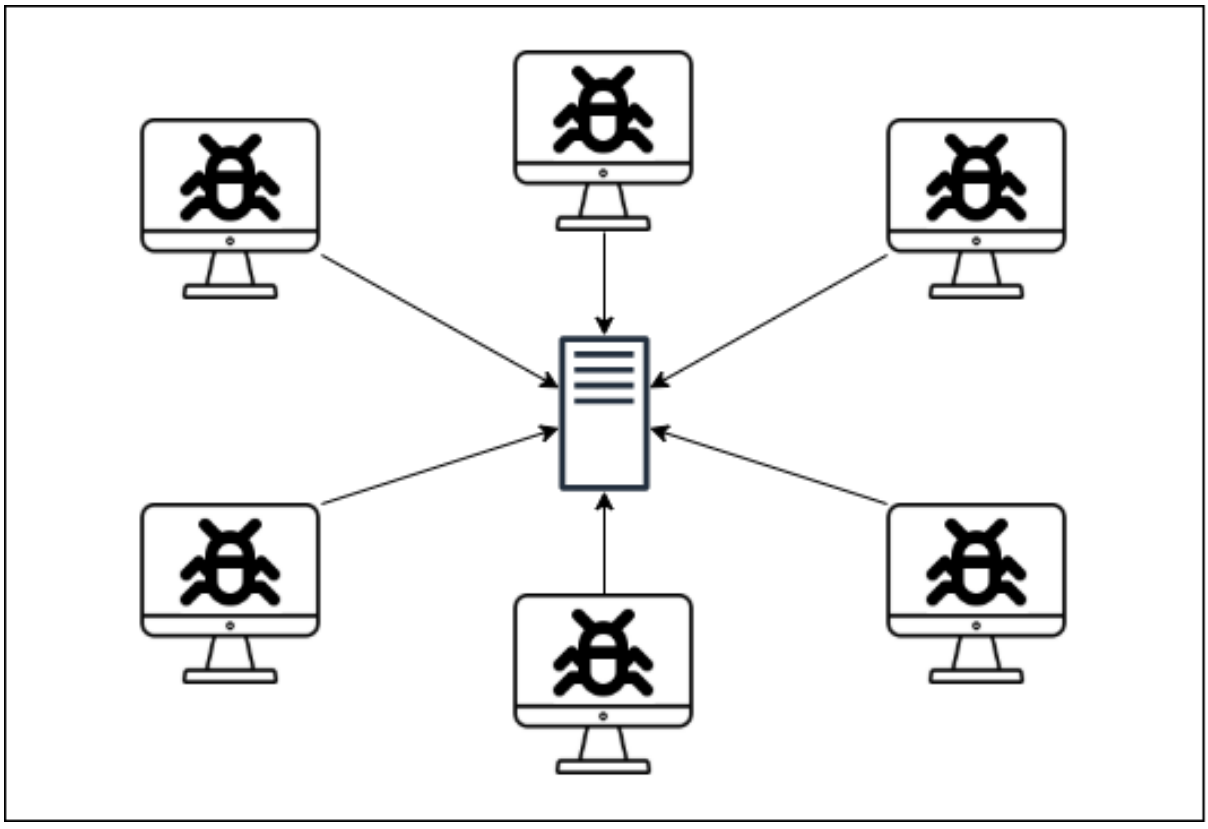

Рисунок 1.1 – Схема побудови централізованої моделі

Дана архітектура (див. рисунок 1.1) доволі проста і з’явилась чи не з найперших. Скомпрометовані машини звітують на центральний сервер періодично. Недоліком є те, що при зникненні центрального сервера, заражені комп’ютери не будуть мати сенсу. При виявленні та контролі центрального сервера захисниками, вони зможуть вимкнути всю мережу. Виявлення C&C сервера є легкою задачею для спеціалістів.

2. Модель «proxies» – модель, яка використовує проксі в архітектурі, задля ускладнення пошуку сервера C&C.

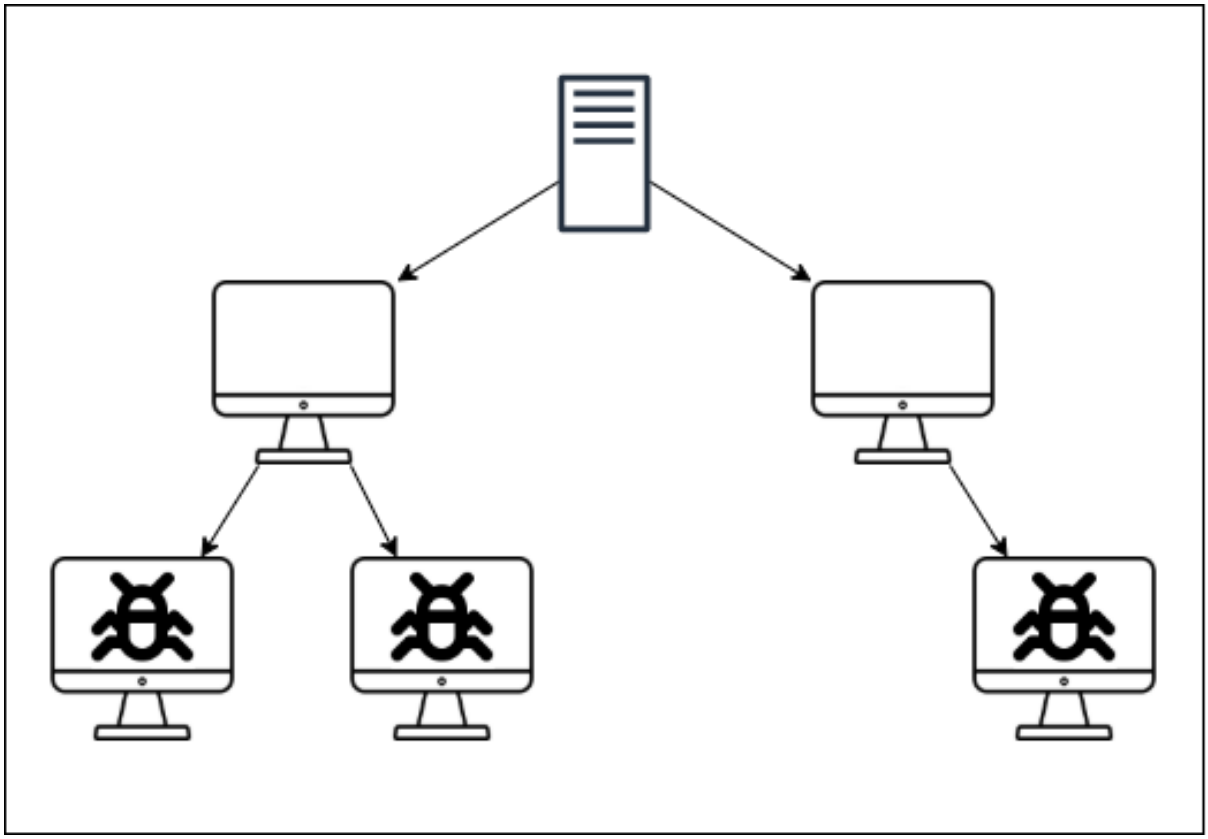

Рисунок 1.2 – Схема побудови «proxies» моделі

Архітектура (див. рисунок 1.2) містить в собі проміжні машини, які слугують як проксі, та скомпрометовані машини, що контактують з проміжними машинами, а не безпосередньо з сервером C&C. Слід зазначити, що проксі-серверами можуть виступати самі заражені машини, або ж C&C сервери.

Перевагами такої архітектури є те, що спеціалісти повинні проаналізувати проксі-сервер, щоб знайти сервер C&C, також при додаванні більшої кількості проксі, інфраструктура стає стійкішою. Але і недолік, як і в централізованій моделі, при відмові C&C сервера, безсиллі заражені машини.

3. Модель «peer-to-peer» (P2P) - однорангові мережі – це мережа, в якій скомпрометовані машини спілкуються між собою без центрального елемента, тобто без C&C.

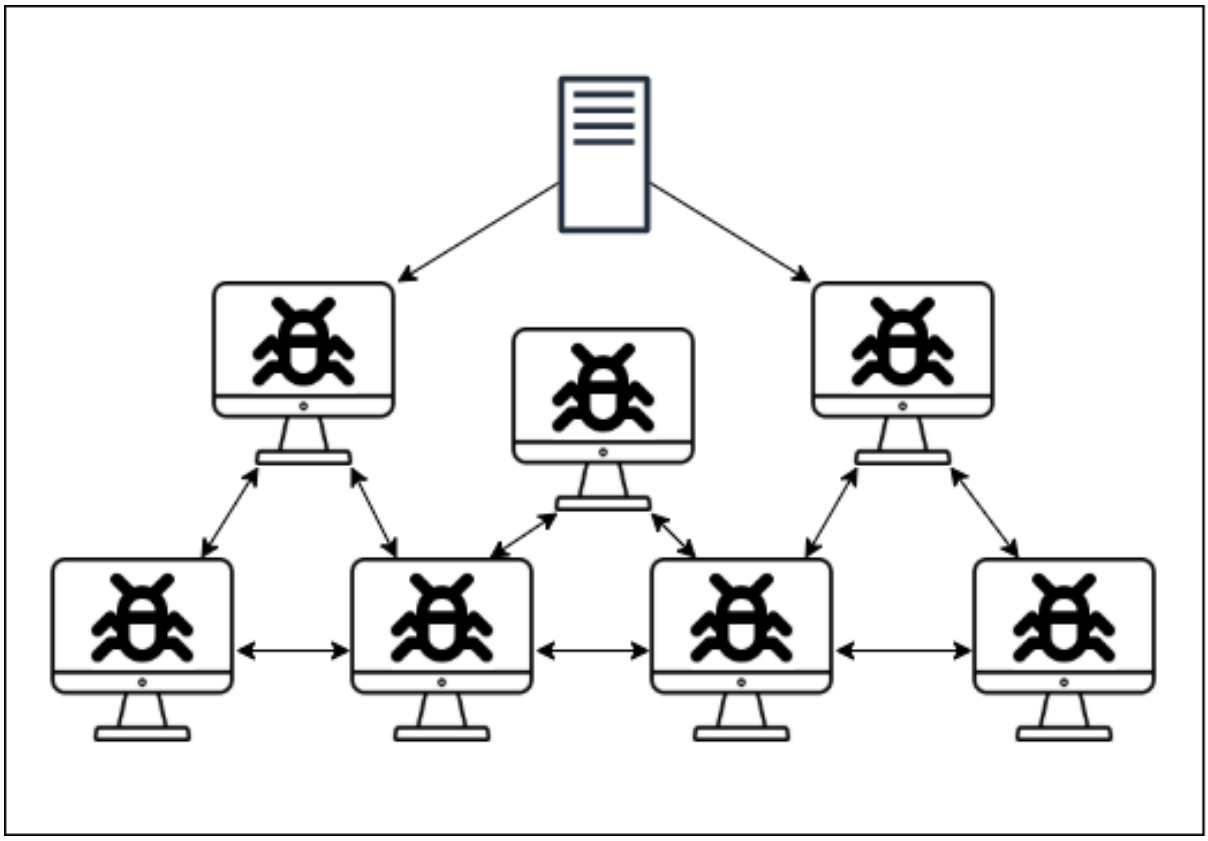

Рисунок 1.3 – Схема побудови P2P моделі

На рисунку 1.3 показано, що команди управління та інформація розповсюджуються в мережі від машини до машини. Зловмисник повинен мати зв’язок з будь-якою скомпрометованою машиною, з метою підтримки контролю над мережею. Тому ліквідація всієї інфікованої мережі являється нелегким завданням.

4. Модель «botnet ecomony» – зловмисники здають свої C&C іншим злочинцям для різних цілей:

- Надіслати спам;
- Виконувати DDoS-атаки;
- Викрасти банківську інформацію;
- Розміщувати нелегальні файли. [5]

1.2 Життєвий цикл атак, які базуються на C&C

Вивчення життєвого циклу атак (див. рисунок 1.4), які основані на C&C, дає змогу передбачати, виявляти та протидіяти їм. Звичайно подібні порушення можуть дещо відрізнятися, але можна представити основні примані риси як послідовність кроків.

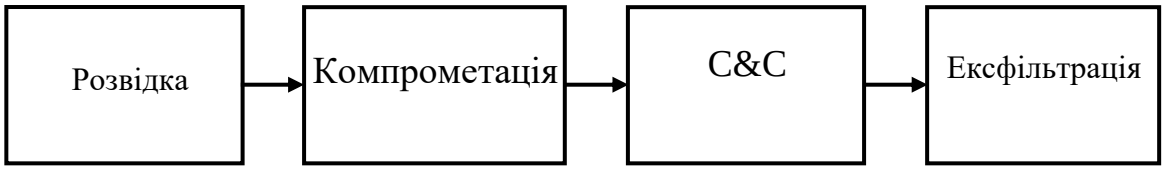

Рисунок 1.4 – Життєвий цикл атак основаних на C&C керуванні

1.2.1 Розвідка

Першим кроком життєвого цикл є розвідка. На цьому етапі зловмисник збирає інформацію про свою ціль та визначає вразливі місця, які будуть використані під час фактичної атаки. Розвідувальна діяльність охоплює не тільки комп'ютерні системи, а і людей. Злочинці досліджують мережі та системи своїх цілей, використовуючи звичні методики, наприклад, сканування портів та

перелік служб, у пошуках вразливостей та неправильної конфігурації, які можуть забезпечити точку входу в організацію. Зловмисники також збирають інформацію про вирішальних людей цільової організації, наприклад, перегляд даних, які можуть бути доступні в соціальних мереж: ця інформація буде використана для полегшення подальших стадій атаки.

1.2.2 Компрометація

Другим кроком циклу атаки є початкова компрометація. Ця фаза являє собою реальне підтвердження проникнення зловмисника в цільову мережу.

Найчастіше методом компрометації є фішинг. Фішингове повідомлення може містити зловмисне вкладення або посилання на шкідливий веб-сайт. Часто зміст фішингового повідомлення пристосовується на основі інформації, отриманої під час етапу розвідки, таким чином, щоб вона виглядала достовірною та законною.

Іншим поширеним варіантом проникнення є стратегічний компрометація веб-сайтів, що цікавлять ціль. Шкідливий код розміщується зловмисником на гіпотетично відвідуваних жертвою сайтах, для того, щоб вона зазнала експлойтів під час перегляду скомпрометований сайтів. Даний метод атаки представляють розвиток традиційних, прагматичних атак "завантаження", коли жертви різними способами залучаються до шкідливої веб-сторінки. У разі успішної операції експлойт завантажує шкідливе програмне забезпечення на машину жертви, що, як наслідок, стає повністю під контролем зловмисника.

1.2.3 Командування та управління

Третій крок атаки: командування та управління атакою – це фаза, коли встановлюється канал зв’язку між скомпрометованою системою та C&C сервером, який дозволяє зловмисникові керувати системою. Канал зв’язку C&C

дозволяє зловмисникові встановити додаткові спеціалізовані модулі шкідливого програмного забезпечення та виконувати додаткові шкідливі дії (поширювати на інші комп’ютери або розпочати DDoS атаку).

1.2.4 Ексфільтрація

Останньою фазою атаки є витягування, збір та шифрування даних, що були поцуплені з системи жертви. Слідом за цим, зашифровані дані передаються до C&C сервера зловмисникам. [6]

1.3 Техніки маскування центрів керування в DNS трафіку

Зловмисники достатньо часто використовують протокол DNS, це дозволяє їм гарантувати з'єднання та керування крадіжкою інформації і перенаправленням трафіку.

Поширеність DNS, а також не дуже ретельний контроль над ним, надає зловмисникам елегантні і тонкі методи підключення та обміну даними.

Існує три основні методи DNS для маскування C&C:

1. Швидкий потік.
2. Алгоритми генерації доменів (DGA).
3. DNS тунелювання.

1.3.1 Швидкий потік

Швидкий потік – це технологія DNS, яка використовує мережу скомпрометованих машин, що виступають в якості зворотних проксі, завбачає швидку та систематичну зміну IP-адрес, пов'язаних із повністю кваліфікованим доменним ім'ям (FQDN).

Основна ідея цієї технології полягає в тому, щоби мати кілька IP-адрес, які зв’язані з доменним іменем, та постійно обмінюватися IP-адресами у швидкій послідовності, змінюючи DNS A або AAAA записи з дуже малим значенням TTL. Пропоновані IP-адреси являються скомпрометованими машинами, яких ще називають агентами швидкого потоку. Зловмисник використовує технологію швидкого потоку в якості уникнення виявлення C&C серверів та чорного списку на основі IP, за допомогою маскування C&C серверів за скомпрометованими машинами, які залучені в ролі зворотних проксі. Технологія швидкого потоку гарантує підключення жертвою до агентів, а не до реальних C&C серверів. [7]

Мережа швидкого потоку поділяється на два типи:

1. Мережа з одним потоком.

2. Мережа подвійного потоку.

Першим кроком до втілення технології швидкого потоку є використання зловмисником скомпрометованих машин (бот-мереж). Бот-мережа містить заражені комп’ютери, які мають зв’язок із C&C серверами. У загальних випадках, агентами швидкого потоку працюють у якості зворотних проксі, які пересилають запити жертви на C&C сервери та відповідають на відповіді, що надходять зі сторони C&C серверів, жертві.

Далі зловмисник присвоює нові IP-адреси для доменного імені чи для сервера імен упродовж дуже короткого відрізку часу від агентів швидкого потоку. IP-адреси агентів виступають у ролі різних IP-адрес шкідливого доменного імені.

C&C сервер є базовою ланкою мереж швидкого потоку, він має багато серверів, що працюють у серверній системі для надання різноманітних послуг по мірі необхідності, наприклад, дозвіл зловмисних доменних імен. В технології швидкого потоку C&C сервери також називаються серверами материнства (mothership). Швидкий потік не обмежується лише будь-яким застосунком, який використовує DNS, але може взаємодіяти зі застосунками HTTP.

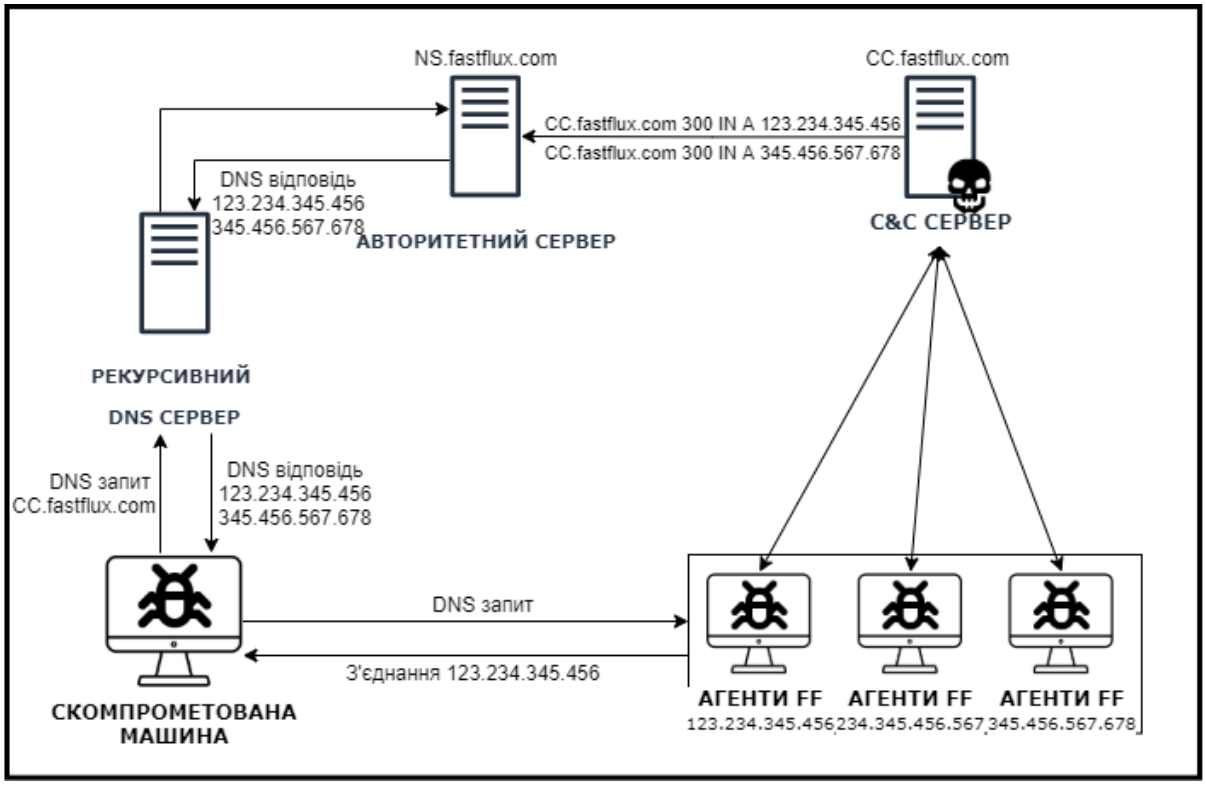

Рисунок 1.5 – Схема реалізації Single-flux

Якщо більш детально розглядати мережу з одним потоком (Single-flux), то потрібно зазначити, що цей тип швидкий потік належить до систематичної та швидкої зміни IP-адрес, що мають відношення до доменного імені.

Головна ідея полягає в систематичному відновлені доменних записів DNS типу A або AAAA адресою агентів швидкого потоку. В Single-flux (див. рисунок 1.5) зловмисник керує авторитетним сервером імен для дозволу імен шкідливого домену та динамічно відновлює DNS A з IP-адресами агентів швидкого потоку з дуже коротким значенням TTL. Авторитетний сервер імен розташовується на надійному хостинг-сервері.

Зміна старих на нові IP-адреси для записів DNS типу A у файлі зони DNS відбувається у зв’язку із закінченням терміну дії TTL. Тисячі IP-адрес агентів швидкого потоку циклічно використовуються для запису DNS типу A, які кожні

10 хвилин, а то й частіше, змінюються. Завдяки цій властивості, жертва кожного разу підключається до нового IP-адреса шкідливого доменного імені.

У ситуації, коли жертва хоче дозволити зловмисне доменне ім’я, надсилається DNS запит на рекурсивний DNS-сервер. А рекурсивний DNS-сервер, зі свого боку, дозволяє запитане доменне ім’я (FQDN) і повертає жертві набір IP-адрес, які є IP-адресами агентів швидкого потоку в ролі зворотних проксі. Далі ініціюється підключення жертвою до однієї з дозволених IP-адрес і відправляється туди запит. Потім за цією адресою пересилається агентом запит жертві на C&C-сервер і передається вміст, отриманий із C&C-сервера, назад жертві. Отже, жертва спілкується з C&C-сервером через агентів швидкого потоку, які виступають у ролі зворотних проксі, а не напряму з C&C-сервер.

Розглядаючи мережу подвійного потоку (Double-Flux) більш детально, потрібно зазначити, що цей тип швидкого потоку належить до динамічної та багаторазової зміни IP-адреси не лише доменного імені, як це було в Single-Flux, а ще і його авторитетних серверів імен із дуже малим значенням TTL.

Головна ідея полягає в тому, що відбувається часта зміна запису Glue DNS A та DNS NS у файлі DNS зони з IP-адресою агентів швидкого потоку. Тисячі агентів беруть участь у процесі та часто реєструють та скасовують реєстрацію своїх IP-адрес як частини запису Glue DNS A та DNS NS**,** для доменного імені та для авторитетного сервера імен відповідно.

Варто підкреслити, що Glue запис — це A-запис сервера імен у реєстрі доменних імен. Цей запис необхідний для уникнення нескінченного зациклення, коли сервери імен знаходяться в піддомені цього ж самого доменного імені. [8]

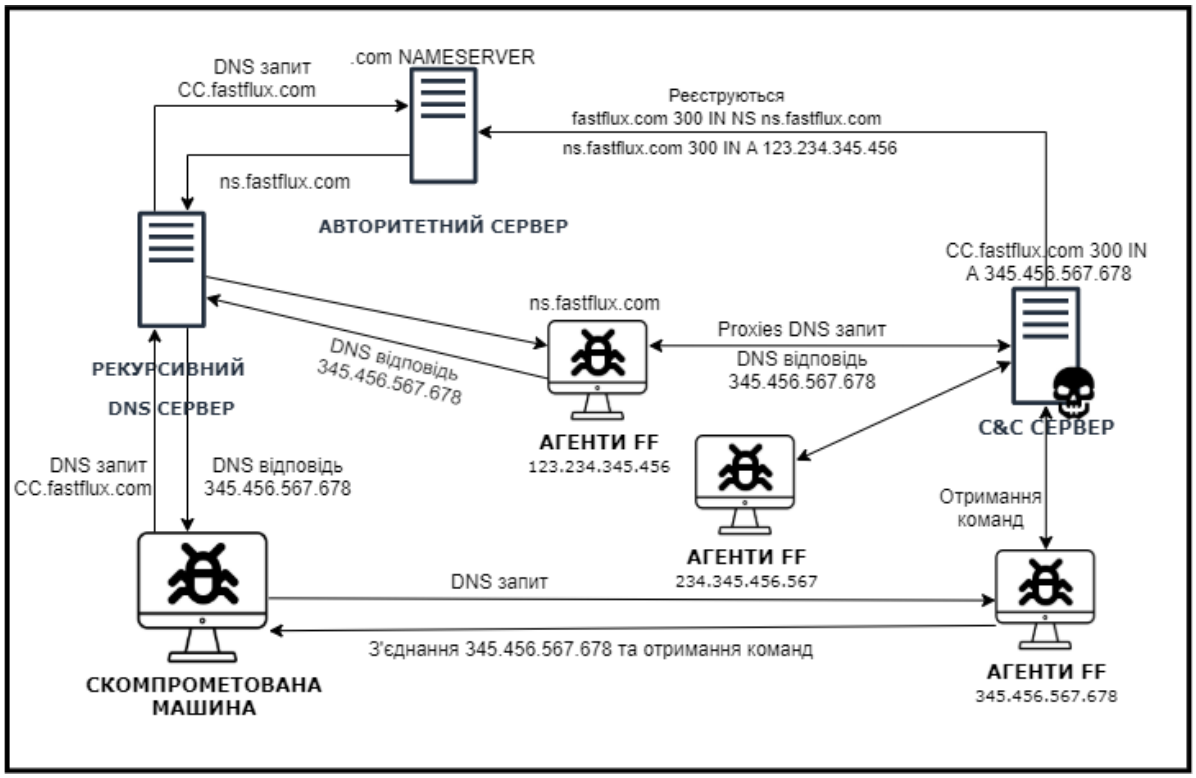

Рисунок 1.6 – Схема реалізації Double-flux

В Double-Flux (див. рисунок 1.6), у порівнянні з Single-Flux, IP-адреси агентів швидкого потоку виступають не лише в ролі IP-адреси шкідливого доменного імені, а й у ролі авторитетного сервера імен. З бот-мережі використовуються зловмисником тисячі агентів швидкого потоку та час від часу реєструється та скасовується реєстрація цих IP-адрес для доменного імені та авторитетного сервера імен.

Отже, основними індикаторами Fast-Flux може бути:

- Велика кількість IP-адрес у одного доменного імені з різних AS і їх часта зміна.
- Швидка/періодична зміна ip-адрес серверів доменних імен.
- Швидка/періодична зміна імен серверів доменних імен.

- IP-адреси належать безлічі різних блоків виділених різним провайдерам.
- Малі значення TTL.
- Невеликий вік домену.
- Наявність сервера ngnix в якості зворотнього проксі-сервера. [41]

1.3.2 Domain Generation Algorithm

Алгоритм генерації доменів (DGA) – це технологія, яка передбачає періодичну генерацію великої кількості псевдовипадкових неіснуючих доменних імен для C&C сервера. Згенерувавши псевдовипадкові доменні імена, скомпрометована машина надсилає DNS-запити, з метою дозволити щойно створені домени. Це відбувається доти, поки одне із доменних імен не дозволить IP-адресу C&C сервера. Також цю технологію називають Domain-Fluxing через невпинну зміну доменного імені для C&C сервера через реалізацію DGA. Доменні імена, які генеруються за даним алгоритмом, також називаються алгоритмічно згенеровані домени (AGD). Варто зауважити, що основною задачею використання DGA є обхід виявлення та відключення C&C-серверів та перешкоджання спробам внесення в чорний список.

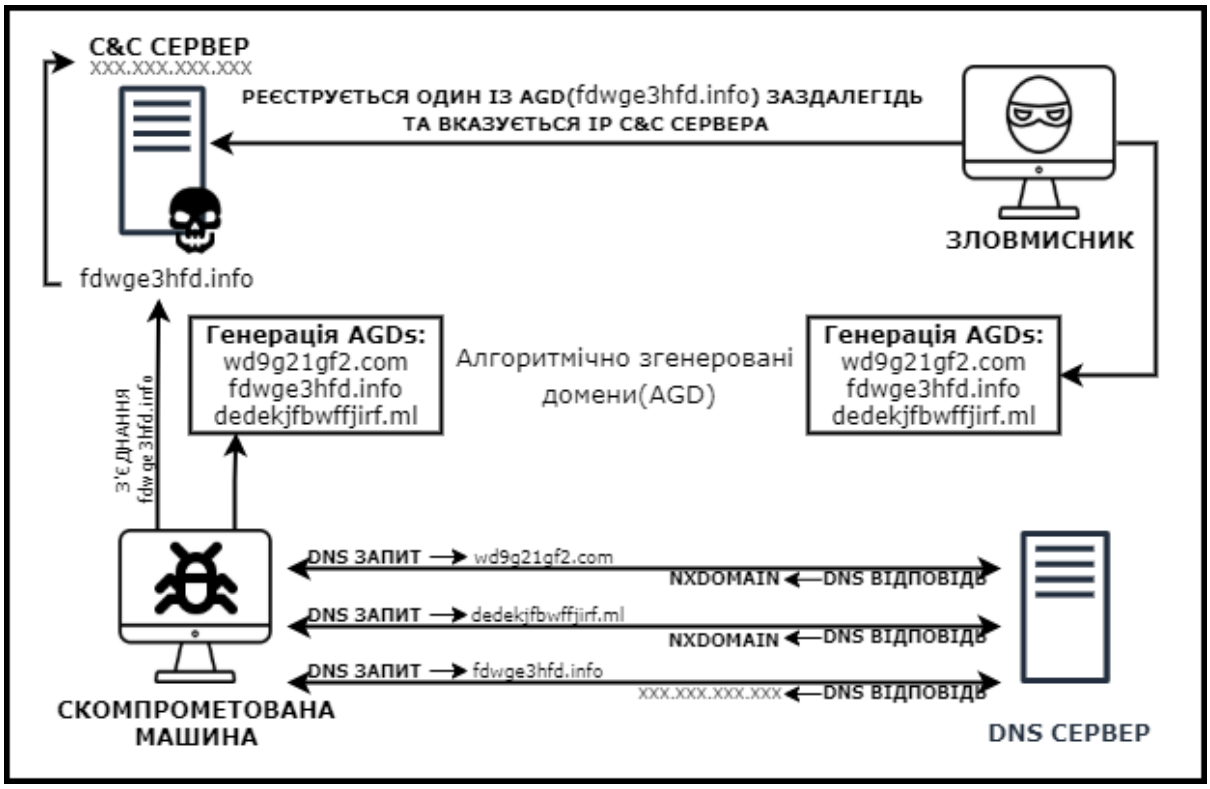

Рисунок 1.7 – Схема реалізації DGA

Завдяки існуванню DGA алгоритму (див. рисунок 1.7) та наявності початкового DGA значення, за допомогою якого генеруються псевдовипадкові доменні імена, зловмисник спроможний згенерувати такий самий список доменних імен, який вираховує шкідливе ПЗ на стороні скомпрометованої машини. При наявності цих знань, зловмисник знатиме з якими доменними іменами заражений комп’ютер намагатиметься зв’язатися через DNS-запити в певні дату та час, а це дає змогу правопорушнику заздалегідь зареєструвати одне із доменних імен, які будуть генеруватися шкідливим ПЗ на основі DGA.

Шкідливе ПЗ на основі DGA систематично генерує велику кількість потенційних доменних імен для C&C сервера та запитує всі ці AGD, щоби визначити IP-адресу центру керування, тобто формально питає який зі списку згенерованих доменних імен є зареєстрованим.

Після того, як зареєстроване доменне ім’я стало відоме для скомпрометованого комп’ютера, зловмисне ПЗ здійснює з’єднання з C&C

сервером та отримує нові вказівки. У ситуації, коли шкідливе ПЗ не в змозі з’єднатися з C&C сервера за попереднім доменним іменем, використовується наступний набір DGA згенерованих доменних імен, доки не буде знайдений зареєстрований домен. Зазвичай, реєстрація доменного імені зловмисником відбувається за 1 годину до атаки й дійсний він протягом 1 доби.

Насіння (seed) є базовим елементом DGA або, як ще можна назвати — початкове значення DGA, воно слугує для генерації псевдовипадкових доменних імен, що є основною вимогою алгоритму генерації доменів (DGA). Насіння відоме тільки зловмиснику і шкідливому ПЗ, це їхній спільний секрет.

Насіння можна розділити на два основні різновиди:

- Статичне насіння
- Динамічне насіння

DGA приймає значення seed як вхідний параметр для генерації псевдовипадкових рядків і алгоритмічно додає TLD (.com,.org,.info) з рядком для виведення можливих доменних імен.

Приклад алгоритмічно згенерований домен (AGD):

nrj4333bdw7039pojdwjex90a75b8f.com

Статичне насіння може базуватися на словниках слів, випадкових рядках чи числах або на будь-чому, що зловмисник може змінювати за будь-який час.

Динамічне насіння залежить від часу, тобто насіння змінні з часом. Це може бути будь-що, починаючи з трендових хештегів в Tik-Tok, показників курсу акцій чи ціни нафту. У загальних випадках, у ролі початкового значення DGA для генерації псевдовипадкових доменних імен використовується поточна дата та час.

Статичні та динамічні елементи насіння об’єднуються в алгоритм для генерації псевдовипадкових рядків, потім TLD, таких як.com,.info, додається з рядками для створення доменних імен. [10]

З метою підвищення рівня захищеності від виявлення C&C серверів, зловмисники вхитряються винаходити все більше й більше видів DGA, які будуть перелічені нижче.

Генератор псевдовипадкових (PRNG) — належить до найбільш вживаного та загального методу DGA. PRNG використовує детермінований генератор випадкових насінь для створення списків доменних імен, передбачуваних як для зловмисника, так і для шкідливого програмного забезпечення. Часто PRNG будуть використовувати системну дату та час як насіння.

DGA на основі символів — цей тип є найпростішим і використовує випадкове насіння для вибору алфавіту або цифр для генерування доменних імен. Оскільки вони найбільш примітивні, їх також найпростіше виявити.

DGA на основі словника – це тип, у якому використовуються слова, засновані на словниках, і випадковим чином поєднує їх, щоби генерувати доменні імена з випадковим, нерозбірливим виглядом. Доменні імена, засновані на базі цього виду, мають властивість бути дуже схожим на законні доменні імена, тому вони є більш складними для систем AI / ML. (Наприклад, Suppobox шкідливих програм.)

DGA з високим зіткненням — це тип, який має сильні зіткнення з іншими DGA, а також легітимними доменними іменами. DGA із високим зіткненням генерує тисячі можливих випадкових доменних імен, які представляють собою 6–15 символів у парі із загальними доменними іменами верхнього рівня (TLD), такими як.net,.org,.info тощо. Структура цих DGA збільшує ймовірність що вони зіткнуться із законними доменами. [9]

1.3.3 Tunneling

DNS тунелювання – це технологія, яка використовує DNS протокол для тунелювання даних за допомогою DNS-запитів та пакетів відповідей. Ця технологія потребує, аби скомпрометована машина запускала шкідливе ПЗ, яке буде надсилати закодовані дані в доменному імені DNS-запитах, а C&C сервер, у свою чергу, надсилає закодовані дані в ресурсних записах пакета DNS-відповіді. Також DNS-тунелювання використовується для зв'язку з C&C сервером, ексфільтрації даних та тунелювання будь-якого трафіку інтернет-протоколу (IP) за допомогою протоколу DNS. Скомпрометована машина надсилає DNS-запит для визначення IP-адреси C&C-сервера, з метою встановлення зв'язку із сервером.

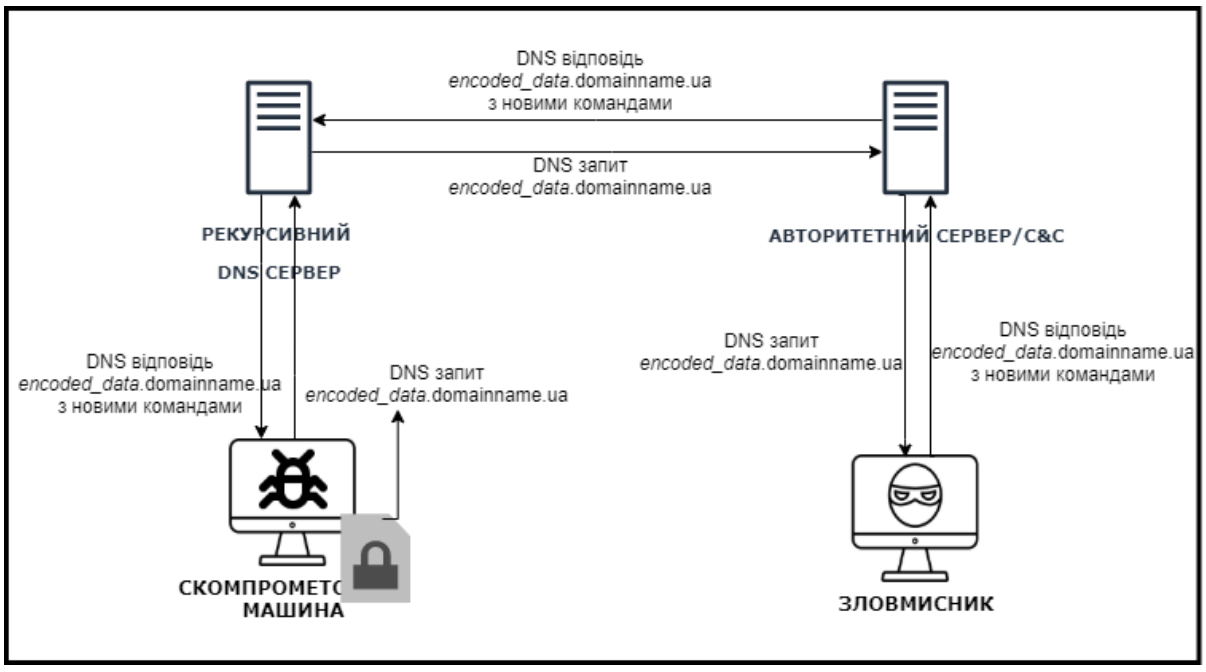

Рисунок 1.8 – Схема реалізації DNS тунелювання

Якщо розглядати більш детальніше (див. рисунок 1.8), то дані корисного навантаження кодуються шкідливим ПЗ в DNS-запитах, далі ці дані передаються як DNS-запит на центр керування. Слід зауважити, що дані зачасту кодуються у форматі base64. Дані корисного навантаження інкапсулюються до доменного

імені DNS-запита. Отримавши DNS-запити від зараженого комп’ютера, C&C надсилає свої закодовані дані, за таким же алгоритмом шифрування, у DNS-відповіді. В основному використовується поле RDATA різних типів ресурсних записів DNS. Частіше за все у DNS-тунелюванні використовується записи типу NULL, TXT та CNAME.

Для наглядності, можна привести приклад, де зловмисник використовує зареєстроване доменне ім’я domainname.ua задля передачі закодованих даних в DNS-запитах у вигляді:

encoded_data.domainname.ua

DNS-запит у вигляді A-записа може бути надісланим скомпрометованою машиною, у цьому записі дані кодуються:

DNS-запит: aGlldmVyeW9uZQ.domainname.ua

У відповідь на запит, центр керування надсилає інструкції CNAME записом:

DNS-відповідь: bWVzc2FnZWZyb21zZXJ2ZXI.domainname.ua

Отже, для реалізації DNS-тунелювання реєструється доменне ім'я, для прикладу, domainname.ua. Потім вказується записи DNS-серверів зловмисника на сервер, на якому запущено програму DNS-сервера тунелювання. Сервер діє як авторитетний сервер імен для цього доменного імені та його піддомену для полегшення тунелювання на стороні сервера та декапсуляції даних за допомогою фонової служби DNS-сервера.

Також потрібно зазначити як відбувається ексфільтрація за допомогою DNS-тунелювання. На зараженому комп’ютері зчитуються дані. Ці дані розділяються на невеликі блоки та закодовуються, зазвичай за алгоритмом шифрування base64. Далі зашифровані дані вбудовуються в окремий піддомен зловмисного доменного імені у DNS-запиті, які надсилаються на C&C сервер зі зловмисним доменним іменем. Спочатку DNS-запит перевіряється рекурсивним DNS-сервером на наявність у кешуванні, так як він там відсутній, DNS-запит

передається через брандмауер на кореневий сервер, TLD-сервер, згодом DNS-запит направляється до C&C-сервер зловмисника, який виступає в ролі сервера імен, там же запущена програма сервера DNS-тунелювання. Правопорушник розшифровує піддомен і, таким чином, отримує викрадені дані. Після цього, зі сторони зловмисника відправляється DNS-відповідь жертві з наступними настановами, які закодовані у DNS-відповіді. Аби не потрапити в кешування, DNS-пакети мають мале значення TTL.

Довжина цільових даних залежить від доступної довжини піддоменів, які використовуються для інкапсуляції зашифрованих даних. 255 символів - максимальна довжина FQDN, з урахуванням крапок. Якщо говорити про піддомени, допустимою довжиною є 63 символи.

Що стосовно DNS-відповідей, то в загальних випадках DNS-тунелювання використовуються такі ресурсні записи як: NULL, TXT та CNAME. Поле даних (RDATA) зберігає дані, в залежності від типу запису, має різний формат/розмір. NULL запис допускає 65535 октетів, а записи CNAME та TXT - 255 октетів. Також DNS-пакети надсилаються через UDP, який обмежується 512 байтами, з метою збільшити об’єм дозволених даних, використовується механізми розширення для DNS - EDNS0. [11][12]

Отже, можна виділити основні ознаки DNS –тунелюваннння:

- **Незвичайні запити домену:** DNS-тунелювання зловмисного програмне забезпечення кодує дані в запитуваному імені домену. Перевірка запитуваних доменних імен у запитах DNS може дозволити жертві диференціювати законний трафік від спроб DNS-тунелювання.
- **Запити на незвичайні домени:** DNS- тунелювання може працює в тому випадку, коли зловмисник володіє цільовим доменним ім’ям, щоб запити DNS надходили на їхній DNS-сервер. Якщо в мережі жертви спостерігається раптовий сплеск запитів на незвичний домен,

це може сигналізувати про наявність DNS- тунелювання, особливо якщо цей домен був створений нещодавно.

- **Великий обсяг трафіку DNS:** максимальне розмір доменного імені в запиті DNS – 253 символи. Це є ознакою, що зловмисникові, скоріш за все, знадобиться велика кількість шкідливих запитів DNS для здійснення ексфільтрації даних або реалізації протоколу командування та управління. В результаті підвищення активності DNS-трафіку може бути показником DNS-тунелювання. [13]

Висновки з розділу 1

Аналіз основних понять центрів керування, які застосовують DNS трафік, та їх побудови показав, що дану проблему не слід недооцінювати і вона варта ретельної уваги зі сторони спеціалістів, оскільки існують методи маскування C&C серверів в DNS трафіку, такі як: fast-flux, DGA та DNS-тунелювання, їхні реалізації та особливості також було детально розглянуто в розділі 1.

Ігнорування аналізу DNS трафіку, який можливо є потенційно шкідливим, може призвести до вимкнення/перезавантаження/перевантаження цільової мережі, тобто порушення роботи системи жертви; крім того, є ймовірність викрадення конфіденційної інформації, а також, при наявності даного чинника, не виключаються шантаж, втрата репутації і т.д.

2 МЕТОДИ НАХОДЖЕННЯ C&C

2.1 Швидкий потік

На сьогоднішній день існує достатня кількість методів виявлення швидкого потоку. В роботі авторів [15] було представлено характеристики швидкого потоку та запропоновано перший метод виявлення в локальній мережі. У статті застосовується три характеристики: кількість унікальних записів A, кількість окремих номерів автономної системи (ASN) та кількість унікальних записів NS.

Команда авторів [17] пропонують пасивний моніторинг мережі за рахунок зменшення трафіку. В запропонованому методі використовуються такі характеристики: кількість унікальних адрес записів, доменне ім'я, TTL, різноманітність окремих префіксів мережі, кількість доменів у мережі, швидкість зростання унікальних мережевих адрес, різноманітність автономної системи, різноманітність протоколу маршрутизації префіксів (BGP), різноманітність організації, яка належить до мережі, різноманітність мережі у країні походження, кількість динамічних IP-адрес, доступність на мережі.

В роботі [16] розробили програмне забезпечення FLUXOR, в якому застосовуються для виявлення скомпрометованої мережі дев'ять характеристик, а саме: вік домену, місце реєстрації домену, кількість унікальних адрес записів, час життя (TTL), кількість різних мереж, кількість різних автономних систем, кількість різних повних імен доменів, отриманих за допомогою зворотного запиту, кількість власників різних організацій та кількість різних імен мережі, призначених реєстром. Недоліком даної техніки є те, що не всі функції застосовуються в режимі реального часу.

Автори [18] поєднали раніше встановлені методи, які базуються на кількісті унікальних адреси запису, різноманітністі мережевих записів та різноманітність автономної системи, з новими методами, де визначається час повторної спроби на авторитетному сервері в різних автономних системах.

Команда авторів [19] розробили утиліту, яка називається Digger. Інструмент визначає наявність в мережі скомпрометованих машин. У дослідженні автори розглядають такі характеристики: кількість унікальних адрес, швидкість зростання кількості унікальних адрес записів, час доступності мережевої адреси та перекриття між мережевою адресою між доменами. Недоліком є те, що аналіз темпів зростання та визначення дублюючих адрес у домені не дозволяють експлуатувати даний метод, як комплекс всіх функцій, в режимі реального часу.

Автори [20] в дослідженні розглядає такі характеристики: кількість унікальних адрес адреси, швидкість зростання кількості унікальних адрес, час життя записів, кількість унікальних записів адреси серверів імен, а також додаткові функції, пов'язані з доменне ім'я: схожість доменних імен зі словником, подібність певних елементів домену з дійсними доменними іменами. Недоліком є те, що повинен бути відомий словник для всіх мов, а також наявність нових технічних термінів.

Автори [21] пропонують інструмент, який називається Exposure. Утиліта використовує такі характеристики: доступність домену, щоденна схожість у поведінці, повторювані моделі поведінки та частота доступу, кількість різних адрес, різні країни для цих адрес, кількість різних доменів, з якими вони мають однакову адресу, середній час життя на сервері імен, стандартне відхилення часу життя між адресами, кількість чисел у доменному імені та відсоток довжини найдовшої частини доменного імені зі значенням. Недоліком є те, що неможливість реалізації в режимі реального часу через порівняння пасивними DNS даними.

Автори [22] пропонують метод, який виявляє аномальну поведінку в мережі. В роботі використовують такі характеристики: групи хостів має невеликий DNS TTL і виконує DNS-запити до не місцевих DNS-серверів. Техніка аналізує велику кількості порожніх DNS-відповідей з кодом помилки. Метод доволі вдалий, але недоліком є складність алгоритму, який несумісний з вбудованими системами.

Команда авторів [23] пропонують утиліту PsyBoG розроблена для виявлення зловмисної поведінки у великих обсягах трафіку DNS. PsyBoG використовує метод обробки сигналів, аналіз спектральної щільності потужності (PSD), з метою виявлення основних частот, що виникають внаслідок періодичних запитів DNS у скомпрометованих машинах. Недоліком є неможливість застосування в режимі реального часу, оскільки вимагає подальшої обробки. [38]

2.2 DGA

В роботі [24] розглядається чорний список як спосіб виявлення DGA доменних імен. Недоліком є те, що він не ефективний, оскільки постійно генерується велика кількість нових доменних імен.

В дослідженні [25] автори використовують машинне навчання. В роботі застосовуються такі характеристики: довжина, біграма та ентропія символів домену, а також спроба застосувати лексичний шаблон, витягнутий з доменів, що не є DGA, до алгоритмів машинного навчання, таких як випадкові ліси. Недоліком є те, що процес вилучення ознак також вимагає часу та досвіду, і зловмисники можуть визначити особливості, щоб уникнути цих методів виявлення, наприклад, DGA доменні імена, які базуються на словниках.

Автори роботи [26] виявив домени DGA, враховуючи морфему та лексичні особливості доменних імен. Також в роботі запропонували евристичний підхід для пошуку морфемних лексем у доменних іменах та розгляду особливостей цих лексем. Недоліком є витрачення немалого часу у вилучення функцій, а також легкість адаптації зловмисників до даного метода..

В дослідженнях [27] та [28] використовували дані WHOIS як додаткову інформацію, яка містить в собі доступність, дату створення та контактну інформацію доменних імен. Недоліком є неможливість застосування цього методу в деяких країнах.

Автори [29] виявляли скомпрометовані машини, центри керування яких базуються на DGA, за допомогою функцій на основі IP, чорного списку та інформації про запити DNS, а також мовних функцій.

В роботі [30] аналізують відповіді NXDOMAIN, які повідомляють про неіснування доменів. В загальних випадках, скомпрометовані машини в одній мережі генерують подібний трафік NXDOMAIN. Базуючись на цій особливості, автори кластеризували домени, розглядаючи схожість між доменними іменами та відповідями DNS. Потім визначали, генеровані вони DGA чи ні, за допомогою змінних дерев рішень.

Також в роботі [31] для виявлення шкідливих доменних імен пропонується застосовувати методи глибинного навчання, які покращують продуктивність. Перевагою є можливість виявляти DGA в режимі реального часу аналізуючи лише доменні імена. Непогані результати показують застосування повторювані нейронні мережі (RNN). Частіше за все використовується довгострокова LSTM та GRU, які достатньо вдало підходять для виявлення шкідливих доменних імен. [42]

2.3 DNS тунелювання

Існує достатня кількість видів методів виявлення DNS-тунелювання, такі як аналіз корисного навантаження DNS або аналіз трафіку, де вказується кількість та частота запитів. При аналізі корисного навантаження в одиночному запиті застосовуються такі характеристики: довжина домену, кількість байт та вміст, а якщо брати до уваги загальний трафік (групу запитів), то є можливість використання таких характеристик: обсяг трафіку DNS , кількість імен хостів для доменного імені, розташування та історії доменів. [43]

В роботах [32-34] автори використовують машинне навчання для виявлення DNS-тунелювання, де застосовують бінарну класифікацію з класом міток «законний» та «тунельний».

Розглядаючи роботу [34] дослідники запропонували метод, який базується машинному навчанні, в якому використовувались два типи класифікаторів, включаючи дерево рішень та випадковий ліс для виявлення DNS-тунелювання.

Запропоновані класифікатори проходили навчання із зашифрованих потоків шляхом проведення статистичного аналізу внутрішнього протоколу. Кожен потік був проаналізований з точки зору специфічних особливостей: розміру та затримки взаємодії пакетів у потоці.

Автори [35] запропонували класифікатор SVM, з метою ідентифікації DNS-тунелювання. Дослідники приділити увагу на змісті запитів та відповідей DNS, з метою отримання шкідливі дані, замасковані стандартним DNS, застосовуючи статистичні особливості запитів і відповідей DNS.

В роботі [33] запропонували метод машинного навчання для виявлення DNS-тунелювання, в якому аналізує прості статистичні особливості повідомлень протоколу: статистика часу взаємодії пакетів та розмірів пакетів. Вибрані характеристики мають на меті розділити трафік на законний та DNS-тунелювання.

Автори [36] запропонували метод випадкової класифікації лісів для ідентифікації DNS-тунелювання. В дослідженні розглядаються декілька характеристик: кількість відповідей, наданих у відповіді, час між двома послідовними пакетами для певного доменного імені та час між двома послідовними відповідями для конкретного доменного імені.

В дослідженні [37] запропонували метод виявлення DNS-тунелювання на основі класифікації ентропії. Автори дослідили внутрішню структуру пакетів методів DNS-тунелювання та описали інформаційну ентропію різних мережевих протоколів.

Автори [32] запропонували класифікатор SVM для ідентифікації DNS-тунелювання в мобільній мережі. В роботі застосовуються характеристики: час, джерело, адресат, протокол і тривалість запиту DNS.

2.4 Основні методи виявлення DGA

В цій роботі хочеться більше закцентувати увагу на основних методах виявлення такого типу маскування як DGA, який буде розглядатися далі в дослідженні більш детально.

2.4.1 Метод машинного навчання

Машинне навчання не вимагає трудомісткого процесу вилучення інформації. В основному використовується в якості класифікацій DGA та не-DGA, а також DNS-тунелювання та не-DNS-тунелювання. Частіше за все, на основі характеристик доменних імен проводиться навчання, для цього частіше за все використовують такі алгоритми:

• Naıve Bayes(NB)

• Decision Trees (DT)

• K-Nearest-Neighbors (KNN)

• Random Forest (RF)

• Logistic Regression (LR)

Також для виявлення використовують глибинне навчання для класифікації DGA доменних імен, в загальних випадках застосовуються LSTM та GRU.

Такий метод має більше 90% виявлення шкідливих доменних імен.

Недоліком даного методу є те, що достатньо складно відрізнити зловмисні доменні імена, які генеруються на основі словників та мають високу колізію із легітимними доменними іменами.

2.4.2 Зворотна інженерія

Зворотна інженерія часто використовується для деконструкції DGA, яка ідентифікує DGA зі зловмисних зразків доменів, а потім блокує домени DGA. Однак зворотне проектування запобігає лише невеликій частині DGA і вимагає значних витрат часу та зусиль від експертів доменів. Таким чином, метод зворотної інженерії є дуже повільним і непрактичним у порівнянні з іншими методами.

2.4.3 Метод чорного та білого списків

Чорний список – це список доменів, які були заблоковані через підозрілі характеристики. Чорний список має на меті знизити процент небажаних доменних імен.

Білий список – це список доменів, які мають надійну репутацію. Білий список має на меті знизити процент помилок першого роду.

Недоліками чорного списку є неспроможність ефективно виявити щойно створені шкідливі доменні імена. Близько 23% нових доменів встановлюється того ж дня чорними списками. Даний метод може бути ефективним, коли система не потребує швидкої реакції. В окремих випадках, щоб досягнути близько 80% виявлення, потрібно більше тижня після створення шкідливого доменного імені.

2.4.4 Пасивний DNS та WHOIS

Пасивні DNS дані – це збережені DNS записи, пошук IP-адрес, а також володіє статистикою, яка має відношення до серверів, доменних імен та IP-адрес, які взаємодіють із загальними комунікаціями DNS.

Завдяки інструментам, які базуються DNS, є можливість зробити аналіз щодо історій DNS записів та перевірити наявність старих доменних імен.

WHOIS надає інформацію про домени та дані, які можна застосувати для відстеження зловмисної діяльності доменного імені та запобігання онлайн-шахрайству.

В поєднанні цих двох методів, можна побачити більш повну картину інформації про домени та IP-адреси, а також достатньо значних розмірів списки асоційованих доменних імен.

2.4.5 Запропонований комплекс методів

Проаналізувавши вище перераховані методи, потрібно виділити те, що найбільш ефективним є машинне навчання. Його можна використовувати як і самостійно, так і в комплексі. Також метод списків, а саме: чорний та білий, цей метод краще використовувати в комплексі через недостатню ефективність у виявленні нових доменів. Всі інші методи потрібно використовувати в поєднані.

Так як головним недоліком машинного навчання є складність виявляти шкідливі доменні імена на основі словників, з метою покращення, пропоную (див. рисунок 2.1) застосовувати аналіз пасивного DNS трафіку, а саме повідомлення типу NXDOMAIN, які є ознакою того, що запитаний домен не існує.

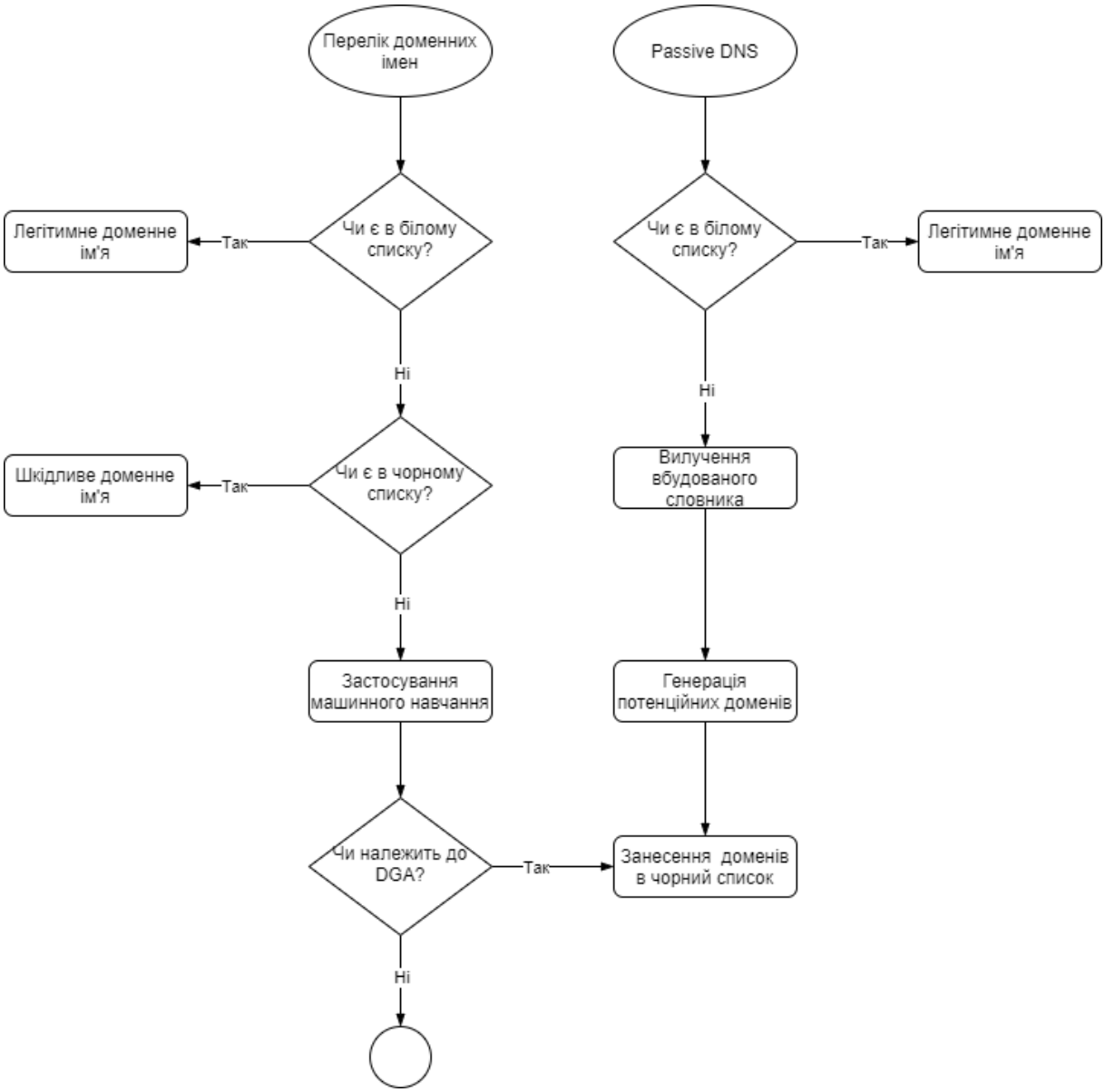

Рисунок 2.1 – Запропонований комплекс методів виявлення DGA

Висновки з розділу 2

У даному розділі були розглянуті методи виявляння, утиліти та характеристики, які використовуються, з метою викриття кожного з типів приховування центрів керування в DNS трафіку.

Також на таблиці 2.1 можна побачити можливість застосування тієї чи іншої характеристики у виявленні центрів керування конкретного типу маскування та можливість застосування функції в реальному часі.

Таблиця 2.1 – Порівняльний аналіз функцій виявлення

	Швидкий потік	DGA	DNS тунелювання	Режим реального часу
кількості унікальних записів [15-20, 29, 35, 36]	+	+	+	+
кількості окремих номерів автономної системи (ASN), різних організацій [15-19]	+	–	–	+
вік домену [16]	+	–	–	+
місце реєстрації домену, різноманітність мережі у країні походження [16, 17, 21, 32]	+	–	+	+
час життя (TTL) [16, 17, 20, 22, 32]	+	–	+	+
кількість різних повних імен доменів, отриманих за допомогою зворотного запиту [16]	+	–	–	+

Продовження таблиці 2.1

швидкість зростання унікальних мережевих адрес [17, 19]	+	–	–	–
кількість динамічних IP-адрес [17, 29]	+	+	–	+
доступність на мережі, час доступності мережевої адреси [17, 19, 21, 27, 28]	+	+	–	–
доменне ім'я та характеристики: довжина, біграма та ентропія символів домену, лексичний шаблон [17, 20, 21, 25, 26, 29, 37]	+	+	+	+
стандартне відхилення часу життя між адресами, аналіз спектральної щільності потужності (PSD) [21, 23]	+	+	–	–
чорний список [24, 29]	+	+	+	+
відповіді NXDomain, які мали місце при доступі до неіснуючих доменів [29, 30]	–	+	–	–
розмір та затримки взаємодії пакетів у потоці [33, 34, 36]	–	–	+	–

Також більш детально було розглянуто методи виявлення DGA, а саме машинне навчання, чорний та білий список, зворотна інженерія та пасивні DNS дані і WHOIS. Для нівелювання недоліків кожного окремого метода було запропоновано комплекс методів виявлення DGA (див. рисунок 2.1).

3 ПРАКТИЧНЕ ВИЯВЛЕННЯ C&C

3.1 Набір даних

Для проведення дослідження було використано щонайменше 430 тисяч доменних імен. В якості навчальних даних було проаналізовано приблизно 200 тисяч легітимних доменних імен та скільки ж шкідливих доменів характеру DGA. В якості тестових даних було використано близько 15 тисяч надійних доменних імен, і таку ж саму кількість шкідливих доменів. Дані для дослідження були взяті з таких ресурсів, як: Netlab OpenData Project By Network Security Research Lab та The Majestic Million [39][40].

Для класифікації даних, кожне доменне ім’я має характеристики, які представляються у векторному вигляді. Дані з характеристиками обробляються за допомогою мови Python на базі Scikit-learn та Keras. Scikit-learn являє собою безкоштовну бібліотеку машинного навчання, яка використовується в якості створення та тренування різних алгоритмів. Keras – це відкрита нейромережна бібліотека.

В роботі порівнюються алгоритми машинного навчання, які були вказані вище (див. розділ 2.4.1), що базуються на певних методах ансамблеювання. А саме: bagging - навчає декілька моделей та комбінує, з метою отримання ефективної моделі; boosting – складові моделі навчаються послідовно, кожна наступна покращує помилки попередньої, отже моделі мають залежність; stacking – використовує окрему модель, аби поєднати результати базових моделей.

Також порівнюються результати класифікацій, що базуються на глибинному навчанні RNN, а саме GRU та LSTM. LSTM-модулі розроблені з метою уникнення проблеми довготривалої залежності, запам'ятовуючи значення як на короткі, так і на довгі проміжки часу. GRU здатні ефективно зберігати

довгострокові залежності в послідовних даних. Крім того, вони можуть вирішити проблему "короткочасної пам'яті", яка мучить звичайну RNN.

3.2 Вибрані характеристики

Для класифікації доменних імен було проаналізовано та виділено декілька характеристик, які притаманні легітимним та шкідливим доменним іменам. Також в дослідженні були виявлені найефективніші комбінації характеристик, з метою якісної класифікації на основі бібліотек Scikit-learn та Keras.

Першою характеристикою виступає довжина доменних імен.

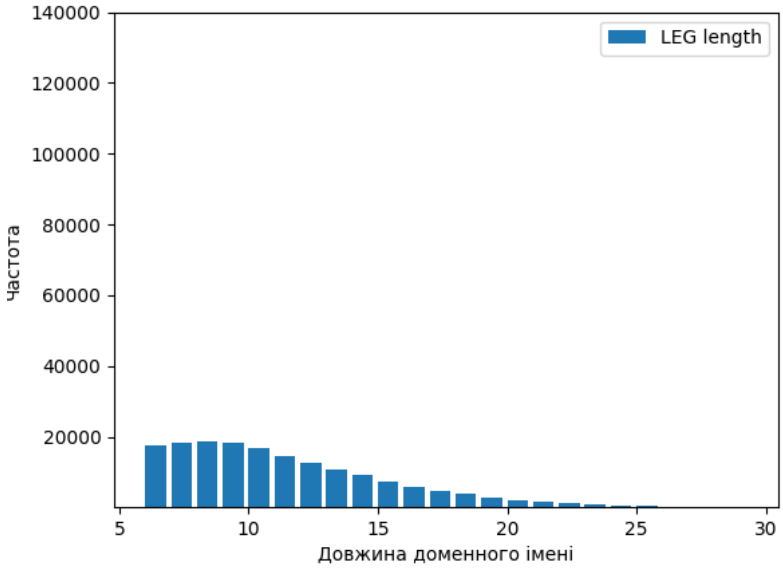

Рисунок 3.1 – Графік довжин легітимних доменних імен

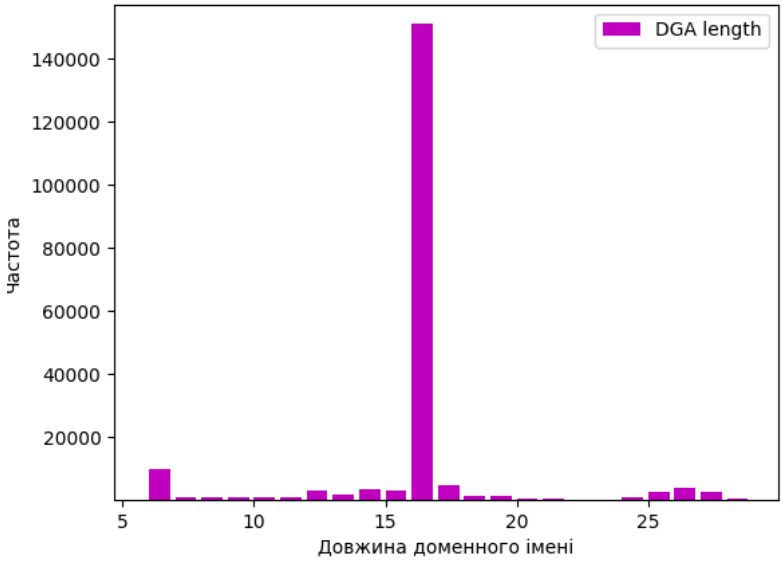

Рисунок 3.2 – Графік довжин шкідливих доменних імен

За рисунками 3.1 та 3.2 можна побачити, що в легітимних доменних іменах розподіл довжин більш нормальний, ніж в доменах, які базуються на алгоритмах.

Слід зазначити, що в середньому для довірених доменів притаманна довжина в 7-9 символів без урахування TLD. Натомість в шкідливих доменних іменах довжина становить 16 символів, і це значення доволі характерне для DGA, судячи за наведеними даними на рисунку 3.2.

Друга характеристика, яка береться до уваги, - піддомени або субдомени. Максимальна кількість рівнів субдоменів - 127, кожний з субдоменів може містити в собі 63 символи, але при цьому треба брати до уваги, що максимальна довжина не перевищує 255 символів.

Третьою характеристикою є домен верхнього рівня(TLD). TLD – це найвищий рівень в ієрархії системи доменних імен після кореневого домена. Також можна сказати, що це начальна точка відліку, з якої починається доменне ім’я в Інтернеті.

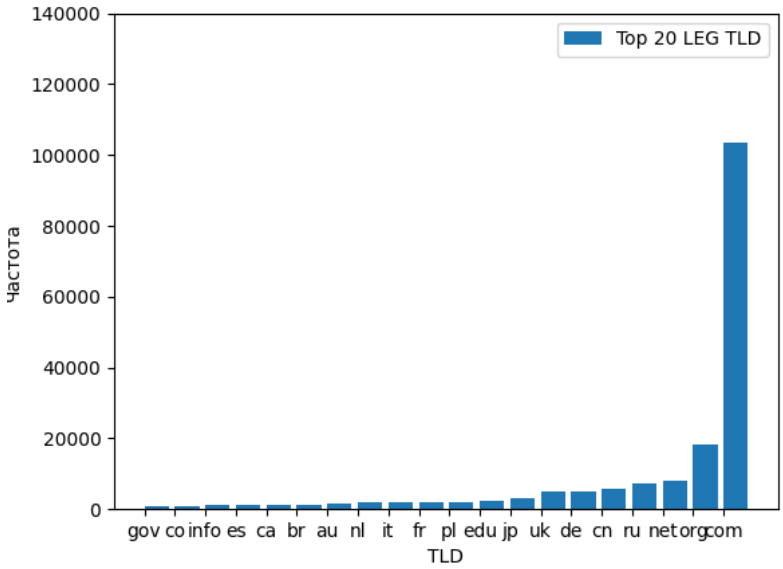

Рисунок 3.3 – Графік частоти TLD в легітимних доменах

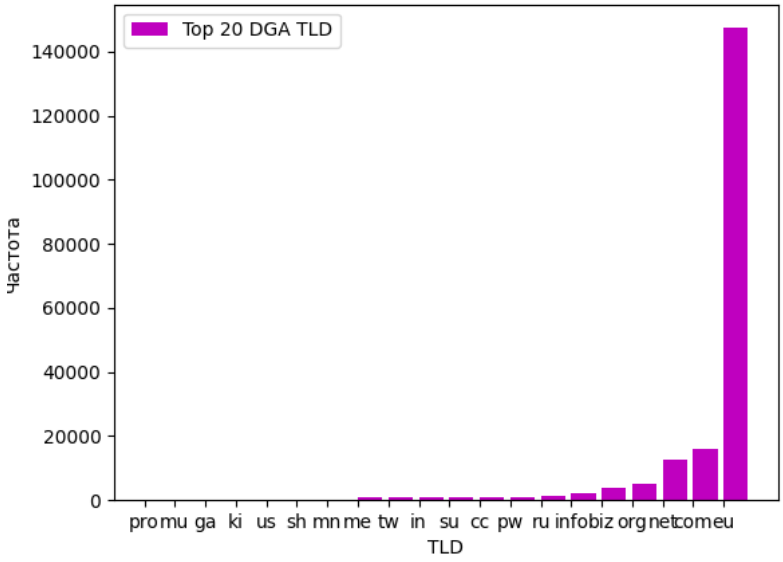

Рисунок 3.4 – Графік частоти TLD в шкідливих доменах

Опираючи на рисунки 3.3 та 3.4, можна побачити велику кількість легітимних доменних імен мають .com TLD, а шкідливі - .eu.

Четверта та п’ята характеристики кількість приголосних та голосних в доменних іменах.

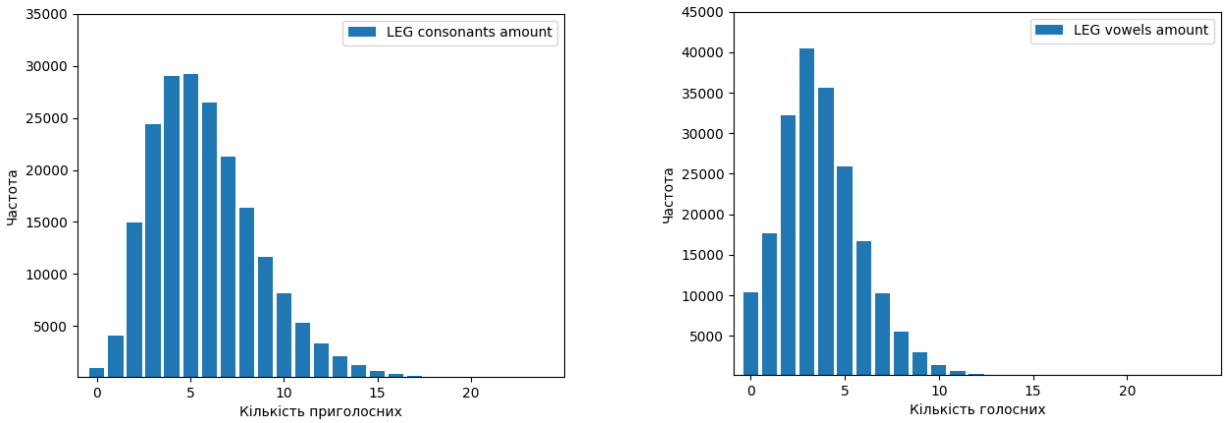

Рисунок 3.5 – Графіки частот букв в легітимних доменних іменах

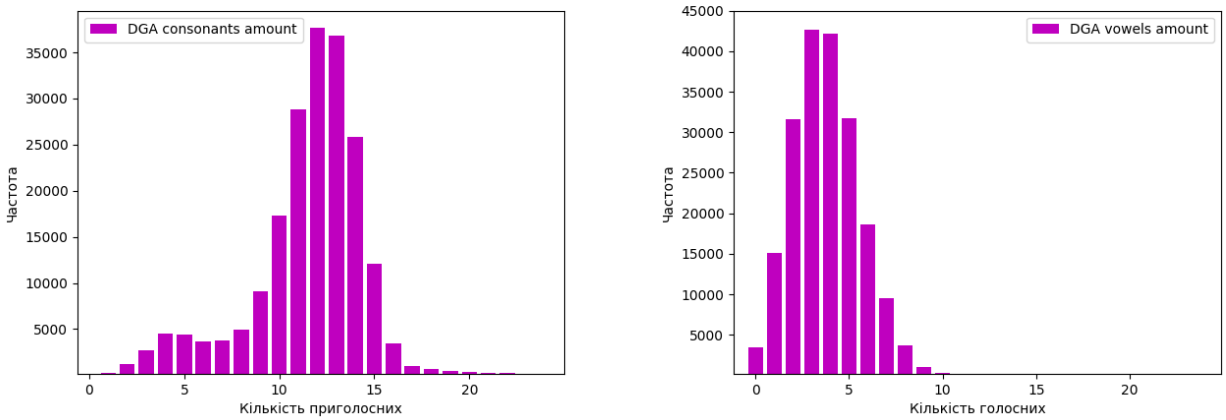

Рисунок 3.6 – Графіки частот букв в шкідливих доменних іменах

За рисунками 3.5 та 3.6 видно, що кількість приголосних більша в шкідливих доменних іменах, аніж у легітимних. Але за кількістю голосних майже не відрізняються, тому немає сенсу надалі розглядати кількість голосних.

Наступною характеристикою виступає – кількість символів (цифри та «-»).

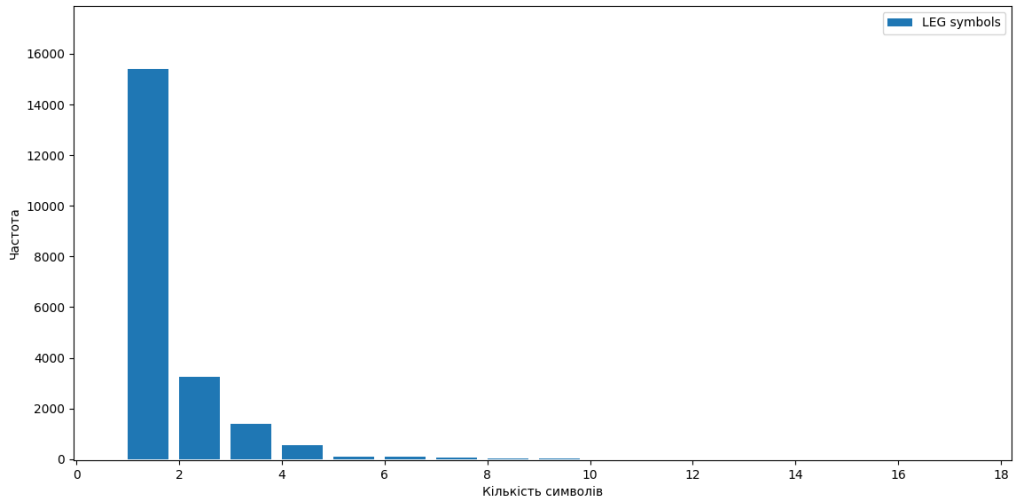

Рисунок 3.7 – Графік кількості символів в легітимних доменних іменах

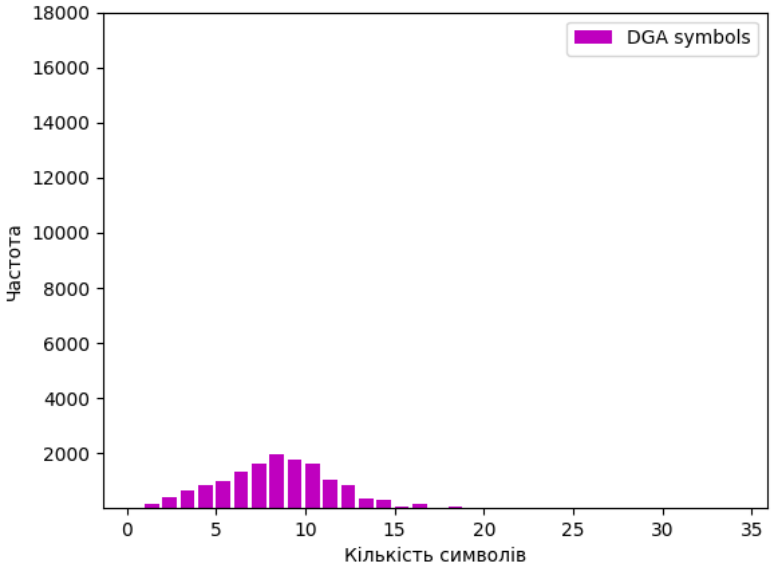

Рисунок 3.8 – Графік кількості символів в шкідливих доменних іменах

В порівняні з шкідливими доменами (див. рисунок 3.8), легітимні (див. рисунок 3.7) мають в загальних випадках незначну кількість цифр та «-» в цілому.

Наступна характеристика, яка бралася до уваги, - ентропія Шеннона (див. формула 3.1).

$$H = -\sum p * log_2 p \quad (3.1)$$

Де p – ймовірність появи n-грами в англійській мові.

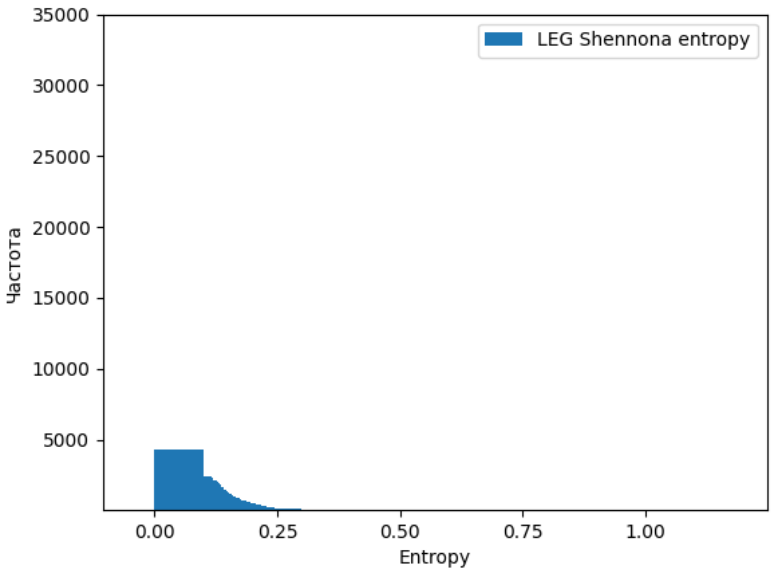

Рисунок 3.9 – Графіки ентропії Шеннона в легітимних доменних іменах

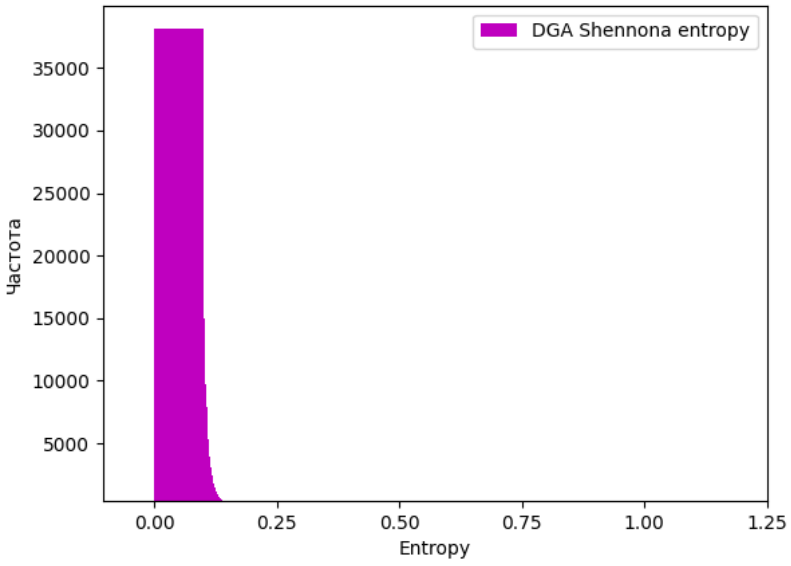

Рисунок 3.10 – Графіки ентропії Шеннона в шкідливих доменних іменах

Якщо порівнювати рисунки 3.9 та 3.10, то можна побачити, що шкідливі домени більше мають значення 0, це через те, що ймовірність n-грам дуже низька або дорівнює 0.

Восьма характеристика – розходження Кульбака-Лейблера. Цей показник – є мірою того, наскільки один розподіл імовірності відрізняється від іншого, який прийняти як еталонний розподіл ймовірностей.

$$D_{KL} = \sum p * log_2 \, p/q \quad (3.2)$$

Де p – ймовірність n-грами в «шкідливому» ймовірнісному розподілі, q – ймовірність n-грами в доменному імені.

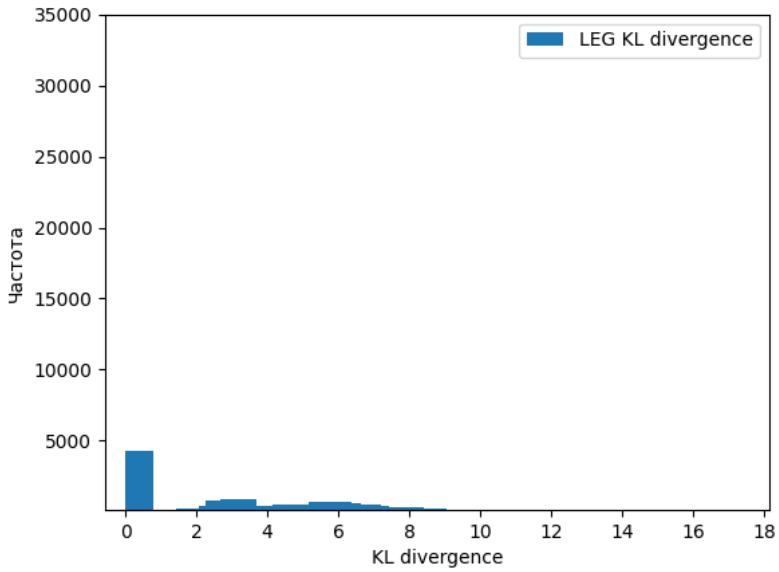

Рисунок 3.11 – Графік розходження Кульбака-Лейблера легітимному доменного імені

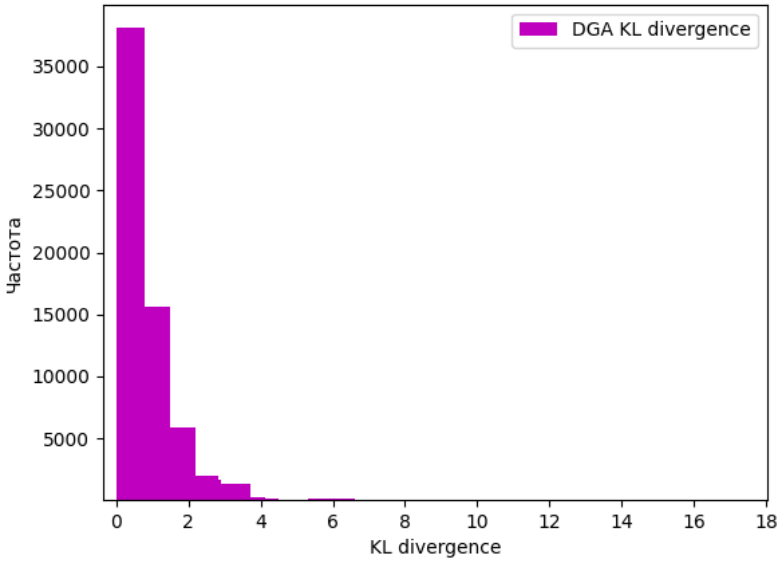

Рисунок 3.12 – Графік розходження Кульбака-Лейблера шкідливого доменного імені

Чим менший показник, тим більша схожість ймовірностей. На рисунку 3.11 видно, що значення ймовірностей домена розхожі із ймовірностями «шкідливого» розподілу. Натомість, показник шкідливих доменних імен (див. рисунок 3.12), в порівняні з легітимними, нижчий, що означає подібність ймовірностей.

Далі розглядається співвідношення значущих слів (див. формулу 3.3), яке визначає співвідношення символів в доменному імені, які містять в значущому слові. Ця характеристика показує наскільки домен покривається існуючими словами.

$$Mean\ character = \sum \frac{l_w}{l_d} \quad (3.3)$$

Де l_w – довжина слова, l_d – довжина доменного імені.

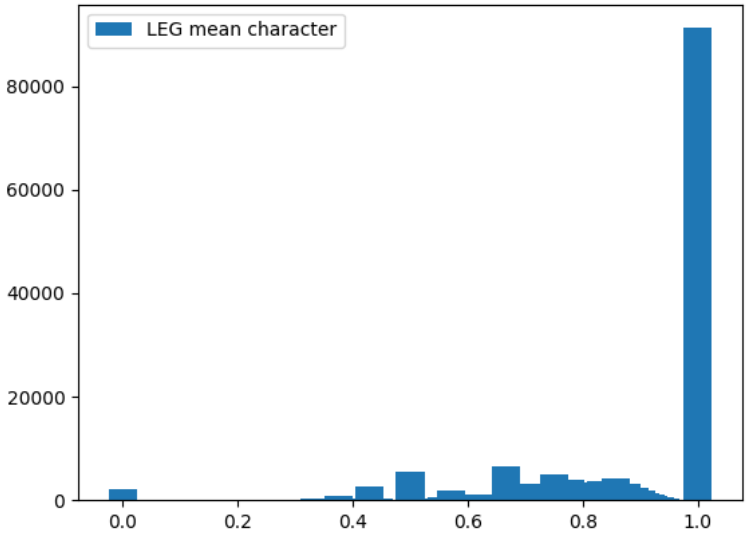

Рисунок 3.13 – Графік співвідношення значущих слів в легітимному домені

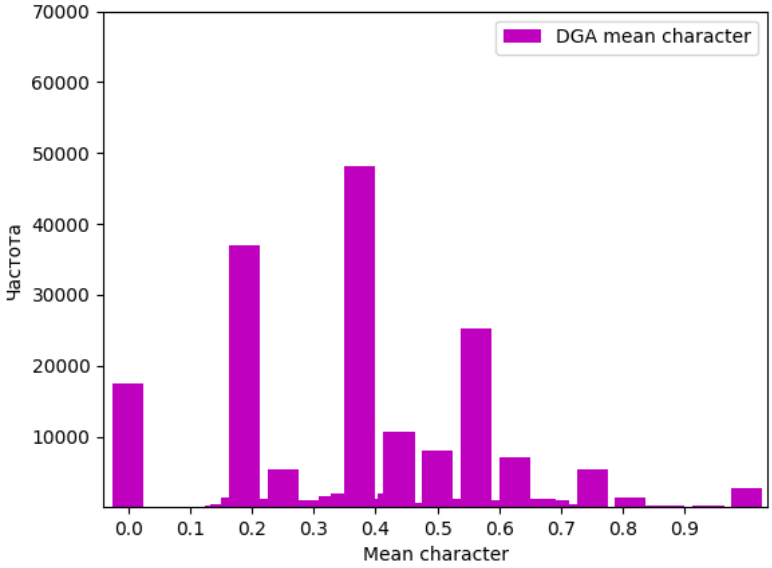

Рисунок 3.14 – Графік співвідношення значущих слів в DGA домені

За рисунком 3.13 можна побачити, що більшість законних доменних імен мають показник 1, це означає повне покриття словами домен, що не можна за шкідливі доменні імена (див. рисунок 3.14), де показники більш випадкові.

Остання характеристика, яка розглядалась, це показник нормальності n-грам (див. формулу 3.4).

$$Normality\ score = \sum \frac{p}{l_d - n + 1} \quad (3.4)$$

Де p – частота n-грами в англійській мові, l_d – довжина доменного імені.

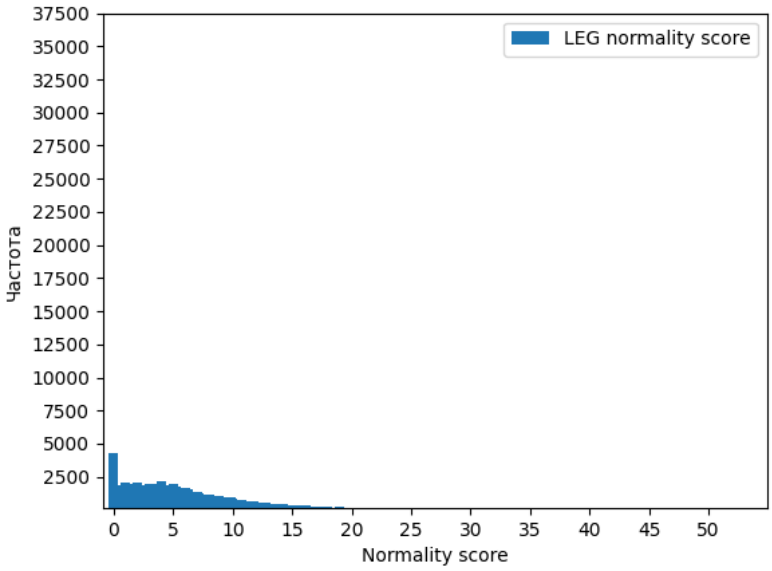

Рисунок 3.15 – Графік показника нормальності n-грам легітимного домену

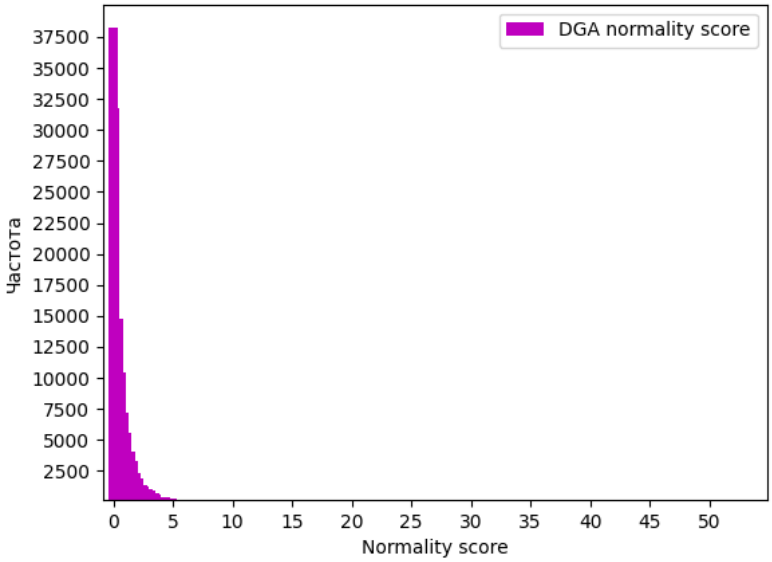

Рисунок 3.16 – Графік показника нормальності n-грам шкідливого домену

Чим нижчий показник, тим вища ймовірність ненадійності доменного імені. Результати, які проілюстровані на рисунках 3.15 та 3.16, вдало показують, що ненадійність законного домену нижча, аніж у DGA.

3.3 Показники оцінювання

В даній роботі проведено певну кількість досліджень, які відрізнялися за своїм планом виконання. Ефективність того чи іншого дослідження та їх результати були оцінені за показниками оцінювання, результати були порівнянні між собою.

Для оцінки якості виявлення шкідливого доменного імен використовуються 3 критерія: accuracy, precision та recall.

Перший показник accuracy вказує на ефективність моделі (див. формулу 3.5).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (3.5)$$

Де TP - правильно визначене DGA доменне ім’я, FP - ідентифіковано як DGA, хоча легітимне, FN - ідентифіковано як легітимне, насправді DGA, TN - правильно визначене легітимне доменне ім’я. Також ці значення містяться в таблиці невідповідностей (див. таблицю 3.1).

Таблиця 3.1. Таблиця невідповідностей

	Визначено як DGA	Визначено як
Дійсний DGA	TP	FN(помилка 2-го роду)
Дійсний легітимний	FP(помилка 1-го роду)	TN

Другий критерій precision (див. формулу 3.6) вказує на точність правильно передбачених позитивних спостережень до загальних позитивно передбачуваних.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (3.6)$$

Третій показник recall – це відношення правильно передбачуваних позитивних спостережень до насправді позитивних даних (див. формулу 3.7).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.7)$$

3.4 Ефективність комбінацій характеристик

В дослідженні порівнюються результати комбінацій характеристик на базі вибраних показників та інструментів. Незмінними характеристиками доменних імен в комбінації виступають довжина, кількість символів (цифр та «-»), кількість приголосних, кількість піддоменів та тип TLD. До змінних характеристик в комбінації належать: ентропія Шеннона, розходження Кульбака-Лейблера, співвідношення значущих слів та показник нормальності n-грам.

Таблиця 3.2 – Результати методу машинного навчання bagging

	Ентропія Шеннона			Розходження Кульбака-Лейблера		
	Accuracy,%	Precision	Recall	Accuracy,%	Precision	Recall
NB	90.33	0.977	0.86	**96.75**	**0.979**	**0.961**
DT	93.29	0.991	0.895	95.21	0.996	0.921
LR	94.33	0.982	0.918	93.39	0.986	0.899
RF	94.62	0.992	0.915	95.26	0.996	0.921
KN	**95.49**	**0.993**	**0.928**	96.59	0.996	0.944
	Співвідношення значущих слів			Показник нормальності n-грам		
	Accuracy,%	Precision	Recall	Accuracy,%	Precision	Recall
NB	93.56	0.936	0.904	90.4	0.985	0.857
DT	95.64	0.998	0.926	94.59	0.992	0.914
LR	92.23	0.998	0.874	93.08	0.979	0.9
RF	95.9	0.998	0.931	95.44	0.992	0.928
KN	**96.13**	**0.998**	**0.934**	**95.9**	**0.992**	**0.935**

Аналізуючи результати метода bagging (див. таблицю 3.2), слід зазначити, що для кожної змінної характеристики можна виділити найвдаліший алгоритм. Для ентропії Шеннона найефективнішим показав себе KN – 95.49%, для розходження Кульбака-Лейблера – NB – 96.75%, для співвідношення значущих

слів – KN – 96.13% та для показника нормальності n-грам – KN – 95.9%. Найменш ефективним себе показав NB з ентропією Шеннона 90.33%.

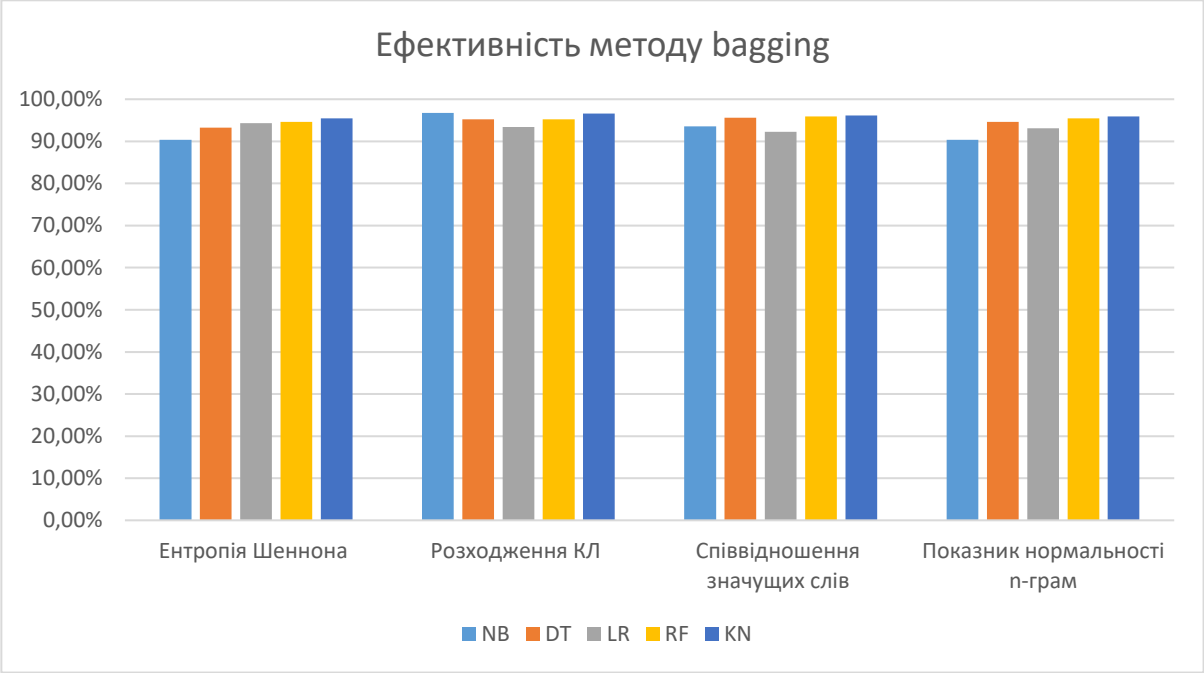

Рисунок 3.17 – Accuracy для методу bagging

За рисунком 3.17 можна виділити найвдаліші три результати: розходження Кульбака-Лейблера – NB та KN – 96.75% та 96.59% відповідно, співвідношення значущих слів – KN – 96.13%.

Таблиця 3.3 – Результати методу машинного навчання boosting

	Ентропія Шеннона			Розходження Кульбака-Лейблера		
	Accuracy,%	Precision	Recall	Accuracy,%	Precision	Recall
NB	39.09	0.041	0.182	75.36	0.992	0.686
DT	**91.18**	**0.989**	**0.865**	**94.41**	**0.996**	**0.907**
LR	82.43	0.997	0.753	91.57	0.994	0.867
	Співвідношення значущих слів			Показник нормальності n-грам		
	Accuracy,%	Precision	Recall	Accuracy,%	Precision	Recall
NB	28.1	0.281	0.005	39.73	0.002	0.012
DT	**94.09**	**0.998**	**0.902**	**94.87**	**0.991**	**0.919**
LR	89.05	0.999	0.83	89.2	0.993	0.835

Аналізуючи результати метода boosting (див. таблицю 3.3), слід зазначити, що в усіх випадках найефективнішим показав себе DT, ентропія Шеннона – 91.18%, розходження Кульбака-Лейблера – 94.41%, співвідношення значущих слів – 94.09%, показник нормальності n-грам – 94.87%. Найменш ефективним себе показав NB з ентропією Шеннона, співвідношенням значущих слів та показником нормальності n-грам – 39.09%, 28.1% та 39.73% відповідно – це дуже погані показники.

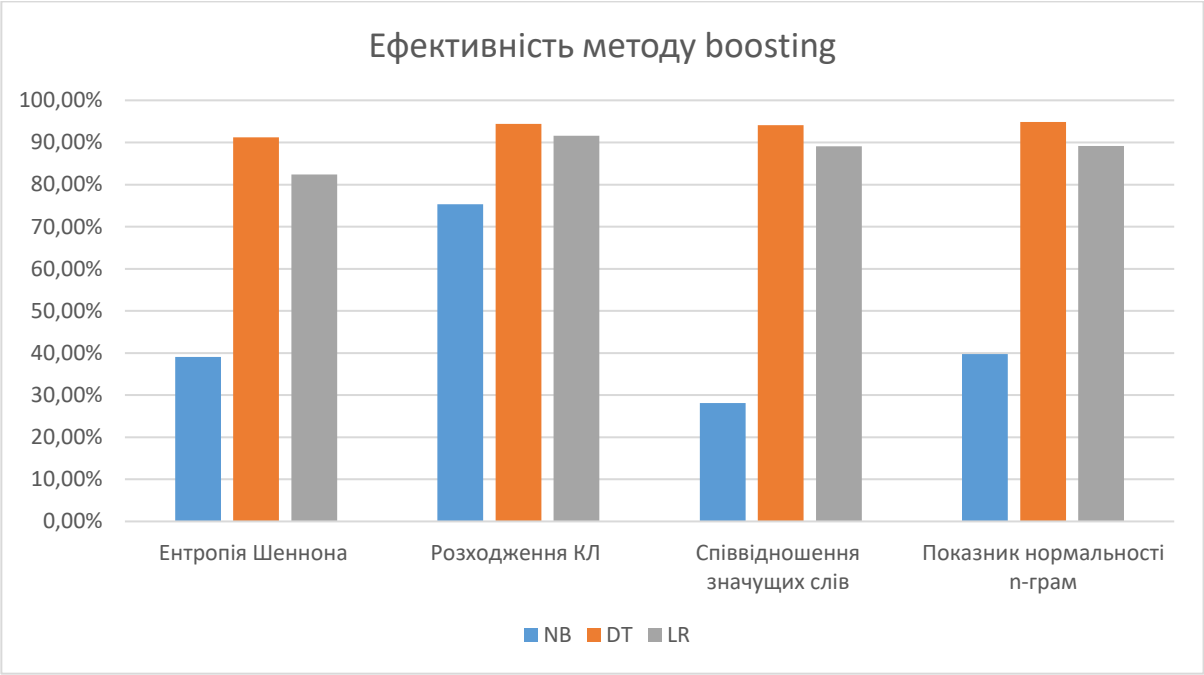

Рисунок 3.18 – Accuracy для методу boosting

За рисунком 3.18 можна побачити різницю результатів алгоритмів, також можна виділити три найвдаліші результати: розходження Кульбака-Лейблера – 94.41%, співвідношення значущих слів – 94.09%, показник нормальності n-грам – 94.87%. В усіх варіантах застосовується алгоритм DT. Результати NB виявились найменш вдалими.

В таблиці 3.4 вказані три вдаліших результати застосування кожної із протестованих характеристик. (Продовження результатів див. додаток А.)

Таблиця 3.4 – Результати методу машинного навчання stacking

	Ентропія Шеннона				Розходження Кульбака-Лейблера		
	Accuracy,%	Precision	Recall		Accuracy,%	Precision	Recall
NB&LR	95.35	0.994	0.924	LR&RF	92.47	0.998	0.878
NB&KN	**96.24**	**0.996**	**0.938**	NB&KN	93.47	0.998	0.892
LR&KN	96.22	0.96	0.937	LR&KN	**93.64**	**0.998**	**0.895**
	Співвідношення значущих слів				Показник нормальності n-грам		
	Accuracy,%	Precision	Recall		Accuracy,%	Precision	Recall
NB&DT	95.32	0.998	0.921	NB&KN	96.17	0.994	0.938
LR&DT	95.32	0.998	0.921	LR&KN	**96.2**	**0.994**	**0.938**
LR&RF	**95.54**	**0.998**	**0.925**	RF&LR	94.68	0.993	0.914

Аналізуючи результати метода stacking (див. таблицю 3.4), слід зазначити, що у випадку з ентропією Шеннона найефективнішим показав себе – NB&KN – 96.24%, розходження Кульбака-Лейблера – LR&KN – 93.64%, співвідношення значущих слів – LR&RF – 95.54%, показник нормальності n-грам – LR&KN – 96.2%.

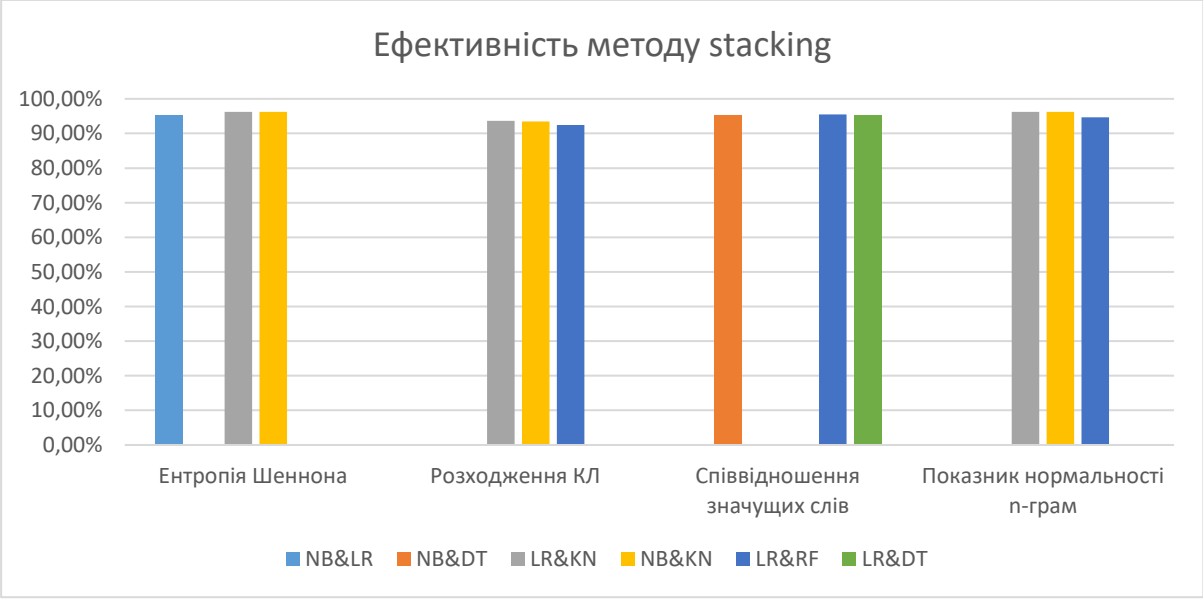

Рисунок 3.19 – Accuracy для методу stacking

За рисунком 3.19 можна виділити чотири найвдаліші результати: ентропія Шеннона – NB&KN та LR&NK – 96.24% та 96.22% відповідно, показник нормальності n-грам – NB&KN та LR&NK – 96.2% та 96.2%. Результати показника розходження Кульбака-Лейблера виявились найменш ефективними.

Таблиця 3.5 – Результати глибинного навчання

	Ентропія Шеннона			Розходження Кульбака-Лейблера		
	Accuracy,%	Precision	Recall	Accuracy,%	Precision	Recall
LSTM	**97.19**	**0.998**	**0.96**	**67.33**	**0.423**	**0.922**
BiLSTM	95.44	0.998	0.923	67.01	0.423	0.91
GRU	93.44	0.997	0.892	66.13	0.423	0.88
BiGRU	91.97	0.998	0.871	65.24	0.423	0.849
	Співвідношення значущих слів			Показник нормальності n-грам		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
LSTM	93.44	0.998	0.892	82.99	0.739	0.928
BiLSTM	**94.44**	**0.998**	**0.91**	83.27	0.739	0.933
GRU	93.36	0.998	0.891	**83.67**	**0.738**	**0.943**
BiGRU	93.31	0.998	0.89	82.37	0.739	0.914

Аналізуючи результати глибинного навчання, слід зазначити, що у випадку з ентропією Шеннона найефективнішим показав себе – LSTM – 97.19%, розходження Кульбака-Лейблера – LSTM – 67.33%, співвідношення значущих слів – BiLSTM – 94.44%, показник нормальності n-грам – GRU – 83.67%.

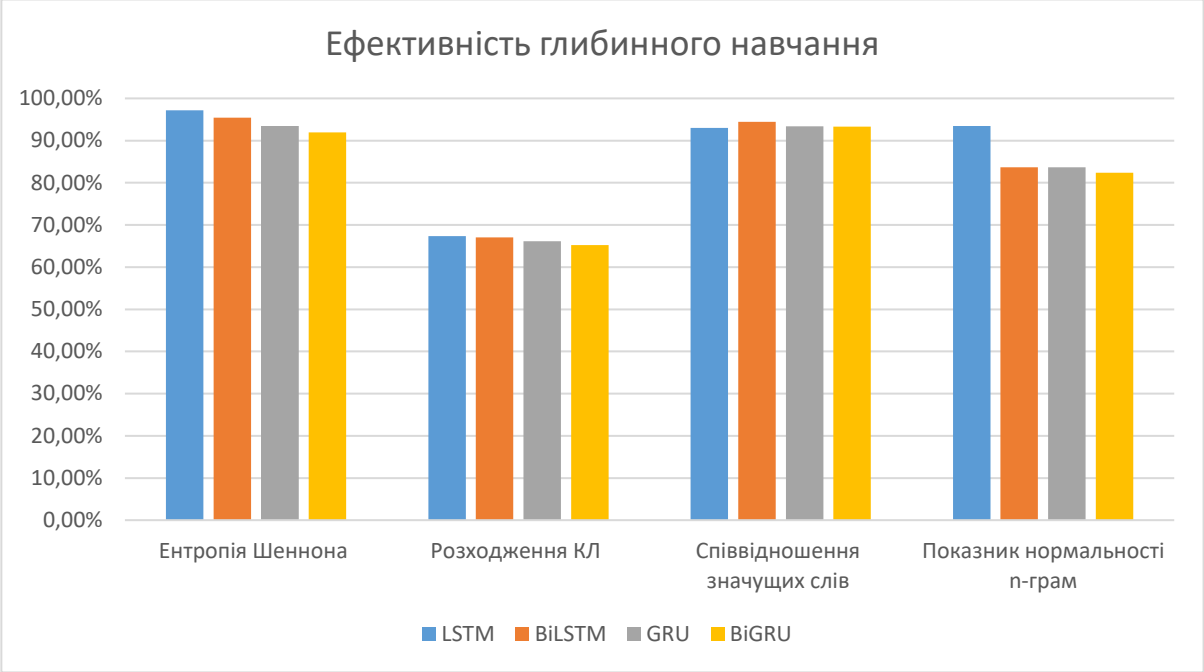

Рисунок 3.20 – Accuracy для глибинного навчання

За рисунком 3.20 можна виділити чотири найвдаліші результати: ентропія Шеннона – LSTM та BiLSTM – 97.19% та 95.44% відповідно, показник нормальності n-грам – LSTM – 93.44%. Результати показника розходження Кульбака-Лейблера виявились найменш ефективними.

Даний метод класифікації достатньо ефективний, але не з доменними іменами, які базуються на словниках. Тому для підвищення ефективності, в даній роботі я використовую аналіз законних та шкідливих, створених на словниках, доменних імен, з метою вивчення їх особливостей для подальшого аналізу пасивних DNS даних, а саме DNS повідомлень типу NXDOMAIN. Цей тип повідомляє про неіснування запитаного доменного імені.

Для отримання словників, на основі яких генеруються зловмисні доменні імена, кожний домен розбивається на існуючі слова та будується неорієнтовний граф, де вершина графа – це слово, а степінь вершини – кількість слів, які разом зустрічались в тому чи іншому доменному імені.

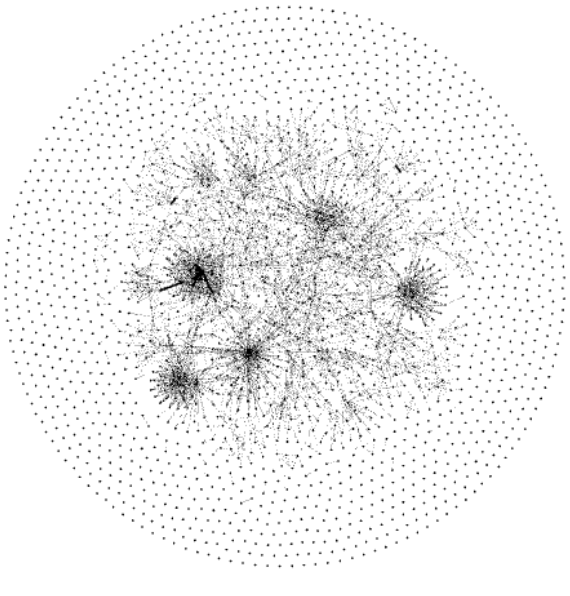

Рисунок 3.21 – Граф зв’язку слів легітимних доменних імен

За проаналізованими легітимними даними, в середньому степінь вершин дорівнює 3. А якщо степінь і перевищує дане значення (як можна побачити на рисунку 3.21 існують чорні плями), то скоріш за все це корпораційні доменні імена, наприклад, google, microsoft і т.д, які повинні міститись в білому списку.

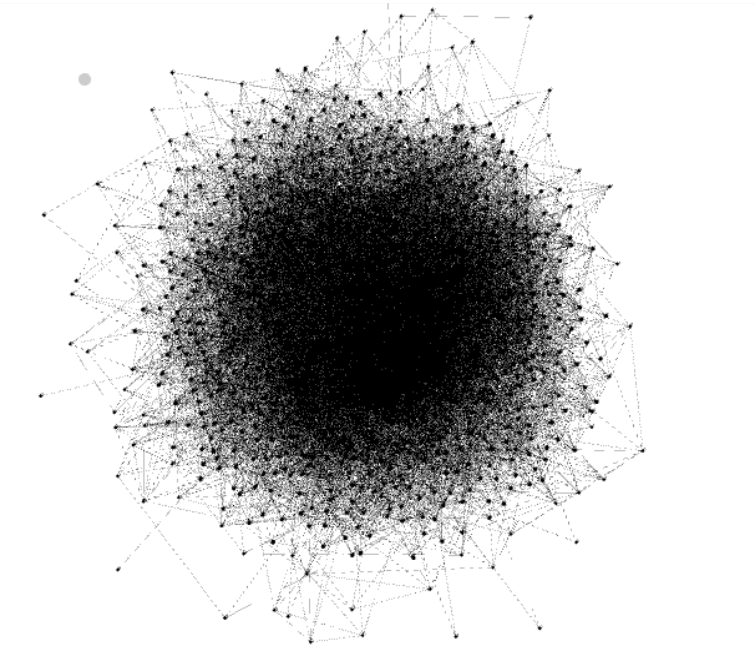

Рисунок 3.22 – Граф зв’язку слів зловмисних доменних імен

За рисунком 3.22 видно, що кожне слово, тобто вершина має велику степінь. Це свідчить про те, що більшість слів мають зв’язки між собою і створюють велику кількість доменних імен. Слова, які мають степінь більшу за 3, заносяться до словника, для подальшої генерації потенційних шкідливих доменних імен, які будуть заноситися до чорного списку.

Висновки з розділу 3

В розділі 3 були проведені дослідження з метою виявленя найефективніший метод класифікації шкідливих доменних імен, також виявити найвдалішу комбінацію характеристик. Для оцінки якості виявлення шкідливих доменних імен застосовувалися 3 критерія: accuracy, precision та recall.

Результати дослідження показали, що в методі bagging найефективнішим показав себе NB з розходженням Кульбака-Лейблера – 96,75%; в методі boosting – DT з показником нормальності n-грам – 94,87%; в методі stacking – NB&KN з ентропії Шеннона – 96,24%; в глибинному навчанні – LSTM з ентропією Шеннона – 97,19%. Найвдаліші результати мають доволі невелику розбіжність поміж собою. Також слід вказати, найгірші випадки: в методі bagging NB з ентропії Шеннона – 90,33%; в методі boosting – NB з співвідношення значущих слів – 28,1%; в глибинному навчанні – BiGRU з розходженням Кульбака-Лейблера – 65,24%. Аналізуючи результати, можна сказати, що метод bagging виглядає більш стабільним. Якщо опиратися на всі значення, то найефективнішим з усіх інструментів виявились глибинне навчання та метод bagging, де використовуються характеристики ентропія Шеннона та розходження Кульбака-Лейблера відповідно.

Також в розділі 3 було порівняно графи зв’язку слів легітимних та шкідливих доменних імен, з метою нівелювання недоліків машинного навчання у виявленні доменів, які генеруються словниках.

ВИСНОВКИ

DNS є важливою ланкою мережі і не слід нехтувати аналізом даного трафіку, бо в противному випадку актуальна загроза конфіденційної інформації через існування центрів керування шкідливого ПЗ, які є централізованими машинами, що керуються кіберзлочинцем, які використовуються для надсилання команд системам, скомпрометованим шкідливим програмним забезпеченням, та отримання викрадених даних із цільової мережі машин. Наразі є достатня кількість виявлення центрів керування, але не все є досконалими.

У цій роботі:

1) Вивчено основні ознаки центрів керування, також було проаналізовано методи маскування в DNS трафіку, що характерні для даної проблеми.
2) Проаналізовано різні методи виявлення центрів керування, більш детально розглянуті методи виявлення алгоритмів генерації доменів та їх особливості, недоліки.
3) Проаналізовано, які характеристики та алгоритми більш доцільні для класифікації шкідливих доменних імен, які генеруються для центрів керування.
4) Вибрано комплекс характеристик та алгоритми для вирішення даної проблеми, проведено дослідження та аналіз найбільш вдалого алгоритму. А саме: незмінними характеристиками доменних імен в комбінації виступають довжина, кількість символів, кількість приголосних, кількість піддоменів та тип TLD. До змінних характеристик в комбінації належать: ентропія Шеннона з використанням глибинного навчання, розходження Кульбака-Лейблера з використанням методу bagging. А також написано алгоритм для вилучення словника.

5) Запропоновано комплекс методів виявлення алгоритмів генерації доменів (див. рисунок 2.1).

Практична цінність дослідження полягає в можливості використання комплексу для виявлення алгоритмів генерації доменів в майбутньому, який може бути інтегрованим у антивірусні програми, з метою виявлення центрів керування ПЗ.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Stop attackers from using dns against you. [Текст] — Режим доступу:https://cdw-prod.adobecqms.net/content/dam/cdw/on-domain-cdw/brands/palo-alto-networks/stop-attackers-from-using-dns-against-you.pdf.
2. Malware hunter powered by Shodan. [Електронний ресурс] — Режим доступу:https://hackersterminal.com/malware-hunter-powered-by-shodan/.
3. Cyber security attacking through command and control. [Електронний ресурс] — Режим доступу:https://www.geeksforgeeks.org/cyber-security-attacking-through-command-and-control/.
4. Command and control server. [Електронний ресурс] — Режим доступу:https://www.trendmicro.com/vinfo/us/security/definition/command-and-control-server.
5. Botnets. [Електронний ресурс] Режим доступу:https://www.enisa.europa.eu/topics/csirts-in-europe/glossary/botnets)
6. Gardiner J. Cova M. Nagaraja S. CommandControl. Understanding, Denying and Detecting. [Текст] — Режим доступу:https://arxiv.org/ftp/arxiv/papers/1408/1408.1136.pdf.
7. Malicious uses of Fast-Flux Service Networks(FFSN) - Hackers Terminal. [Електронний ресурс] — Режим доступу:https://hackersterminal.com/fast-flux-service-networks-ffsn-technique/.
8. DNS glue records – что такое и почему они так важны? [Електронний ресурс] Режим доступу:https://habr.com/ru/company/webo/blog/328188/.
9. Domain generation algorithms dgas. [Електронний ресурс] — Режим доступу:https://zvelo.com/domain-generation-algorithms-dgas/.
10. Domain Generation Algorithm – DGA in Malware- Hackers Terminal. [Електронний ресурс] — Режим

доступу:https://hackersterminal.com/domain-generation-algorithm-dga-in-malware/.

11. Extension Mechanisms for DNS (EDNS(0)). [Електронний ресурс] — Ре-жим доступу:https://tools.ietf.org/id/draft-ietf-dnsext-rfc2671bis-edns0-09.html#rfc.section.3.
12. How DNS Tunneling Works as CC CommunicationChannel for Botnet - Hackers Terminal. [Електронний ресурс] — Режимдоступу:https://hackersterminal.com/dns-tunneling/.
13. What is dns tunneling. [Електронний ресурс] — Режим доступу:https://www.checkpoint.com/cyber-hub/network-security/what-is-dns-tunneling/.
14. Всё про техногию Fast-Flux сетей. [Електронний ресурс] — Режим доступу:https://www.imena.ua/blog/.
15. Holz T. Gorecki C. Rieck F. Measuring and detecting fast-flux service networks. [Текст]
16. Passerini E. Paleari R. Martignoni L. Bruschi D.FluXOR: Detecting and Monitoring Fast-FluxService Networks. [Текст]
17. Perdisci R. Corona I. Dagon D. Lee W. DetectingMalicious Flux Service Networks through PassiveAnalysis of Recursive DNS Traces. [Текст]
18. Nazario J. Holz T. As the net churns: Fast-fluxBotnet observations.
19. Hu X. Knysz M. Shin K. Measurement and analysisof global ipusage patterns of fast-flux Botnets. [Текст]
20. Marchal S. Francois J. Wagner C. State R.Dulaunoy A. Engel T. Festor O. DNSSM A largescale passive DNS security monitoring framework. [Текст]
21. Bilge L. Kirda E. Kruegel C. Balduzzi M.EXPOSURE: Finding Malicious Domains UsingPassive DNS Analysis. [Текст]
22. Pomorova O. Savenko O. Lysenko S. Kryshcuk A. Bobrovnikova K. A Technique for the BotnetDetection Based on DNS-Traffic Analysis. [Текст]
23. Kwon J. Lee K. Lee H. Perrig A. PsyBoG: Ascalable Botnet detection method for large-scaleDNS traffic. [Текст]

24. K¨uhrer M. Rossow C. Holz T. Paint it black:evaluating the effectiveness of malware blacklists. [Текст]
25. Zhang Y., Xiao J. Detecting the DGA-based malicious domain names. [Текст]
26. Wei-wei Z. Jian G., Qian L. Detecting machine generated domain names based on morpheme features. [Текст]
27. Curtin R. Gardner A. Grzonkowski S.Kleymenov A., Mosquera A. Detecting DGAdomains with recurrent neural networks and sideinformation. [Текст]
28. Ferrante A. The impact of GDPR on WHOIS: implications for businesses facing cybercrime. [Текст]
29. Schiavoni S. Maggi F. Cavallaro L., Zanero S.Phoenix: DGA-based botnet tracking and intelli-gence. [Текст]
30. Antonakakis M. Perdisci R. Nadji Y. From throw-away traffic to bots: detecting the rise of DGA-based malware. [Текст]
31. Tran D. Mac H. Tong V. Tran H., Nguyen L. A LSTM based framework for handling multiclassimbalance in DGA botnet detection. [Текст]
32. Do V. Engelstad P. Feng B. Do T. Detection of DNS tunneling in mobile networks using machinelearning. [Текст]
33. Aiello M. Mongelli M. Papaleo G. DNS tunnelingdetection through statistical fingerprints of protocolmessages and machine learning. [Текст]
34. Allard F. Dubois R. Gompel P. Morel M. Tunnelingactivities detection using machine learning techni-ques. [Текст]
35. Aiello M. Mongelli M. Papaleo G. Basic classifiersfor DNS tunneling detection. [Текст]
36. Buczak A. Hanke P. Cancro G. Toma M. WatkinsL. Chavis J. Detection of tunnels in PCAP data byrandom forests. [Текст]
37. Homem I. Papapetrou P. Dosis S. Entropy-basedPrediction of Network Protocols in the ForensicAnalysis of DNS Tunnels. [Текст]
38. Cafuta D. Sruk V. Dodig I. Fast-Flux Botnet Detection Based on Traffic Response and Search Engines Credit Worthiness. [Текст]

39. Списки dga доменних імен. [Електронний ресурс] — Режим доступа: https://data.netlab.360.com/dga/.
40. Списки dga доменних імен. [Електронний ресурс] — Режим доступа: https://majestic.com/reports/majestic-million/
41. Техники использования DNS в атаках вредоносных программ. [Електронний ресурс] – Режим доступа:https://cyberart.ru/poleznoe/post/tehniki-ispolzovaniya-dns-v-atakah-vredonosnyh-programm
42. Namgung J. Son S. Moon Y. Efficient Deep Learning Models for DGA Domain Detection. [Текст]
43. Almusawi A. Amintoosi H. DNS Tunneling Detection Method Based on Multilabel Support Vector Machine. [Текст]

ДОДАТОК А. ПРОДОВДЖЕННЯ РЕЗУЛЬТАТІВ МЕТОДУ STACKING

Розходження К-Л	Ентропія Шеннона	Співвідношення ЗС	Показник норм.
NB & DT	NB & DT	NB & LR	NB & DT
accuracy 92.36	accuracy: 93.77	accuracy: 91.73	accuracy: 93.81
NB & LR	NB & RF	NB & KN	NB & LR
accuracy 81.89	accuracy: 94.75	accuracy: 94.75	accuracy: 94.23
NB & RF	DT & NB	NB & RF	NB & RF
accuracy 92.12	accuracy: 93.77	accuracy: 95.08	accuracy: 94.28
DT & NB	DT & LR	DT & LR	DT & NB
accuracy 92.36	accuracy: 93.77	accuracy: 95.31	accuracy: 93.809
DT & LR	DT & KN	DT & KN	DT & LR
accuracy 92.38	accuracy: 93.78	accuracy: 95.21	accuracy: 93.71
DT & KN	DT & RF	DT & RF	DT & KN
accuracy 92.34	accuracy: 93.77	accuracy: 95.31	accuracy: 93.63
DT & RF	LR & NB	LR & NB	DT & RF
accuracy 92.41	accuracy: 95.34	accuracy: 91.73	accuracy: 93.81
LR & NB	LR & DT	LR & KN	LR & NB
accuracy 81.89	accuracy: 93.75	accuracy: 94.53	accuracy: 94.43
LR & DT	LR & RF	KN & NB	LR & DT
accuracy 92.39	accuracy: 94.37	accuracy: 94.75	accuracy: 93.706
KN & NB	KN & NB	KN & DT	LR & RF
accuracy 93.46	accuracy: 95.24	accuracy: 95.21	accuracy: 94.43
KN & DT	KN & DT	KN & LR	KN & NB
accuracy 92.35	accuracy: 93.78	accuracy: 94.53	accuracy: 96.16
KN & LR	KN & LR		KN & DT
accuracy 93.64	accuracy: 96.22		accuracy: 93.65
KN & RF	KN & RF		KN & LR
accuracy 91.4	accuracy: 94.06		accuracy: 96.19
RF & NB	RF & NB		KN & RF
accuracy 92.31	accuracy: 94.7		accuracy: 94.49
RF & DT	RF & DT		RF & NB
accuracy 92.44	accuracy: 93.78		accuracy: 94.38
RF & LR	RF & LR		RF & DT
accuracy 92.45	accuracy: 94.69		accuracy: 93.78
RF & KN	RF & KN		RF & KN
accuracy 92.09	accuracy: 94.29		accuracy: 94.64