БАКАЛАВРСЬКА РОБОТА

БР. ІП - 12.00.00.000 ПЗ

Група ІП-21-4

Сагайдак Денис

2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Сагайдак Денис Ігорович

(прізвище, ім'я, по батькові)

УДК 004

(індекс)

БАКАЛАВРСЬКА РОБОТА

Розробка методології запобігання поширення шкідливого програмного забезпечення на рівні маркетингових стратегій

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Здобувач освітнього рівня Сагайдак Д.І.

(підпис, ініціали та прізвище здобувача)

Науковий керівник Процюк Галина Ярославівна, асистент

(підпис, прізвище, ім'я, по батькові, науковий ступінь, вчене звання керівника)

Допущено до захисту
Завідувач кафедри

доц. Бандура В.В.

(посада) (підпис) (дата) (ініціали та прізвище)

Івано-Франківськ – 2025

Івано-Франківський національний технічний університет нафти і газу

Інститут, факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Ступінь вищої освіти бакалавр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:
Зав. кафедрою **ІПЗ**
доц. В.В. Бандура
“ ____ ” _____2025 р.

ЗАВДАННЯ

НА БАКАЛАВРСЬКУ РОБОТУ СТУДЕНТОВІ

Сагайдаку Денису Ігоровичу

(прізвище, ім’я, по-батькові)

1. Тема проекту (роботи) “ Розробка методології запобігання поширення шкідливого програмного забезпечення на рівні маркетингових стратегій”

керівник проекту (роботи) Процюк Г.Я. асистент

затверджені наказом закладу вищої освіти від “ 28 ” квітня 2025 р. № 264/7

2. Строк подання студентом проекту (роботи) 10 червня 2025 р.

3. Вихідні дані до проекту (роботи) Результати і матеріали отримані під час проходження переддипломної практики

4. Зміст розрахунково - пояснювальної записки (перелік питань, які потрібно розробити)

1. Аналіз проблематики поширення шкідливого ПЗ засобами онлайн-реклами
2. Розробка методології запобігання поширення шкідливого програмного забезпечення
3. Аналіз та оцінка зібраних рекламних оголошень
4. Оцінка результатів імплементації методології запобігання поширення шкідливого ПЗ
5. Ефективність використаних методів виявлення

5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)

1. Ілюстрація рекламного оголошення, що імітує фірмову ідентичність "Uber Eats" (рис. 1.1)
2. Зображення лендінг-сторінки за посиланням на яку перенаправляється користувач (рис. 1.2)
3. Модель Real-Time Bidding (RTB) (рис. 1.3)
4. VirusTotal - веб-ресурс перевірки на шкідливе програмне забезпечення (рис. 1.4)
5. Приклад зловмисного рекламного оголошення, що розповсюджується як веб-push (рис. 1.5)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання **28 квітня 2025 р.**

Керівник _____
(підпис)

Завдання прийняв до виконання _____
(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів проекту	Примітка
1	Аналіз проблематики поширення шкідливого ПЗ засобами онлайн-реклами	07.05.2025	виконано
2	Розробка методології запобігання поширення шкідливого програмного забезпечення	17.05.2025	виконано
3	Аналіз та оцінка зібраних рекламних оголошень	27.05.2025	виконано
4	Оцінка результатів імплементації методології запобігання поширення шкідливого ПЗ	01.06.2025	виконано
5	Ефективність використаних методів виявлення	15.06.2025	виконано
6	Оформлення пояснювальної записки дипломної роботи завідувачем кафедри	10.06.2025	виконано

Студент – дипломник _____
(підпис)

Керівник роботи _____
(підпис)

АНОТАЦІЯ

Бакалаврська робота містить 76 сторінок, 36 рисунків, список використаних джерел із 38 найменуваннями.

Мета роботи - розробити та оцінити методологію ідентифікації та запобігання поширенню шкідливого програмного забезпечення через онлайн-рекламу з використанням сучасних інструментів автоматизації та аналізу даних.

Об'єкт дослідження - онлайн-реклама як середовище поширення шкідливого програмного забезпечення.

Предмет дослідження - методи і засоби виявлення та запобігання зловмисній рекламній діяльності з використанням автоматизованих інструментів.

В першому розділі аналізується онлайн-реклама як потужний канал для поширення шкідливого ПЗ, що потребує нових підходів до її моніторингу та аналізу в контексті фішингових атак і кібератак

В другому розділі запропоновано методологію та створено інструмент на основі Playwright для збору, маркування та аналізу рекламних оголошень з метою виявлення шкідливих елементів

В третьому розділі розроблена методологія продемонструвала високу ефективність у виявленні шкідливих рекламних кампаній, зокрема в аспектах географії загроз, векторів атак і джерел небезпеки.

Висновок: розроблено інструмент на базі фреймворку Playwright, який дозволяє здійснювати автоматизований збір, аналіз та класифікацію рекламних оголошень.

КЛЮЧОВІ СЛОВА: ОНЛАЙН-РЕКЛАМА, ШКІДЛИВЕ ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ, КІБЕРБЕЗПЕКА, PLAYWRIGHT, ІДЕНТИФІКАЦІЯ ЗАГРОЗ, ВЕБ-СКРЕПІНГ, КЛАСИФІКАЦІЯ ДАНИХ, ІНФОРМАЦІЙНА БЕЗПЕКА

ANNOTATION

The bachelor's thesis contains 76 pages, 36 figures, a list of used sources with 38 names.

The purpose of the work is to develop and evaluate a methodology for identifying and preventing the spread of malicious software through online advertising using modern automation and data analysis tools.

The object of the study is online advertising as a medium for the spread of malicious software.

The subject of the study is methods and means of detecting and preventing malicious advertising activities using automated tools.

The first section analyzes online advertising as a powerful channel for the spread of malicious software, which requires new approaches to its monitoring and analysis in the context of phishing attacks and cyberattacks

The second section proposes a methodology and creates a tool based on Playwright for collecting, marking and analyzing advertisements in order to detect malicious elements

In the third section, the developed methodology demonstrated high efficiency in detecting malicious advertising campaigns, in particular in terms of the geography of threats, attack vectors and sources of danger.

Conclusion: a tool based on the Playwright framework has been developed that allows for automated collection, analysis, and classification of advertisements.

KEYWORDS: ONLINE ADVERTISING, MALICIOUS SOFTWARE, CYBERSECURITY, PLAYWRIGHT, THREAT IDENTIFICATION, WEB SCRAPING, DATA CLASSIFICATION, INFORMATION SECURITY

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

RTB - Real-Time Bidding

CDP - Chrome DevTools Protocol

CDP+ - Chrome DevTools Protocol Plus (розширена версія протоколу)

eTLD - Effective Top-Level Domain

eTLD+1 - Effective Top-Level Domain plus one

HAR - HTTP Archive (format)

IPQS - IPQS (Specific service name)

JSON - JavaScript Object Notation

VT - VirusTotal (often used as an abbreviation for the service name)

WPNs: - Web Push Notifications

ВСТУП

У сучасних умовах глобальної цифровізації дедалі більше процесів у суспільстві переходять у віртуальний простір. Це стосується не лише комунікації, торгівлі чи розваг, а й механізмів впливу на аудиторію через онлайн-рекламу, яка стала одним із наймасштабніших інструментів сучасного маркетингу. Водночас разом із розвитком рекламних технологій зростає і рівень кіберзагроз, пов’язаних із їх використанням у зловмисних цілях.

Серед найбільш небезпечних тенденцій останніх років — використання онлайн-реклами як каналу для поширення шкідливого програмного забезпечення (ПЗ). Зловмисники використовують рекламні платформи для розміщення фішингових посилань, перенаправлення користувачів на небезпечні ресурси або навіть автоматичного завантаження ПЗ на пристрої. Особливо небезпечною є здатність таких атак маскуватися під легітимну рекламу, що ускладнює їхнє виявлення стандартними засобами захисту.

Особливу загрозу становить імітація фірмової ідентичності популярних брендів, через що користувачі легко втрачають пильність. Також спостерігається активне використання пошукової реклами для маніпулювання результатами видачі з метою скерування трафіку на підроблені сайти. Ці загрози не тільки ставлять під удар конфіденційність і безпеку користувачів, а й формують негативний імідж легітимних компаній, під які маскуються зловмисники.

Зважаючи на складність виявлення шкідливої реклами в масштабах реального часу та необхідність оперативного реагування, виникає потреба у створенні автоматизованих систем моніторингу, збору та аналізу рекламного контенту. Одним із перспективних напрямів є використання інструментів на кшталт Playwright — фреймворку для автоматизованої взаємодії з

вебсторінками, який дозволяє емулювати поведінку користувача, збирати контент і виконувати аналіз у контрольованому середовищі.

Актуальність роботи

Актуальність дослідження зумовлена потребою у створенні комплексних систем раннього виявлення зловмисної активності, що поєднують автоматизоване сканування, аналіз поведінки та інструменти обробки великих даних.

Ця робота присвячена розробці та впровадженню методології виявлення зловмисної онлайн-реклами із застосуванням сучасних засобів автоматизації, збору великих даних та їх подальшої класифікації. Результати дослідження дозволяють оцінити ефективність такого підходу в умовах динамічного ринку цифрової реклами та зростаючих кіберзагроз.

Проблематика поширення шкідливого програмного забезпечення через онлайн-рекламу є надзвичайно актуальною в умовах стрімкого зростання цифрових комунікацій. З розвитком рекламних технологій зловмисники отримали нові можливості для маскування шкідливого контенту під легітимну маркетингову активність. Традиційні методи фільтрації та реагування часто виявляються недостатніми для виявлення нових, складних форм загроз.

Мета роботи - розробити та оцінити методологію ідентифікації та запобігання поширенню шкідливого програмного забезпечення через онлайн-рекламу з використанням сучасних інструментів автоматизації та аналізу даних.

Завдання дослідження

- Проаналізувати сучасні загрози, пов’язані з використанням онлайн-реклами як каналу поширення шкідливого ПЗ.

- Дослідити існуючі методи виявлення зловмисної рекламної діяльності.

- Розробити методологію автоматизованого виявлення шкідливої реклами.

- Реалізувати інструмент збору та аналізу рекламного контенту.

- Оцінити ефективність запропонованого підходу на практичних даних.

- Визначити особливості векторів атак, цільової аудиторії, та джерел загроз.

Об'єкт дослідження - онлайн-реклама як середовище поширення шкідливого програмного забезпечення.

Предмет дослідження - методи і засоби виявлення та запобігання зловмисній рекламній діяльності з використанням автоматизованих інструментів.

Методи дослідження

- Методи веб-скрепінгу та автоматизованого збору даних;

- Статистичний аналіз рекламних повідомлень;

- Метод класифікації даних;

- Аналіз поведінкових патернів реклами;

- Візуалізація географічного та технічного розподілу загроз.

Наукова новизна

Удосконалено підхід до ідентифікації шкідливої онлайн-реклами шляхом розробки методології, що поєднує автоматизований веб-скрепінг, позначення та категоризацію даних з метою запобігання кібератакам у межах маркетингових стратегій.

Практичне застосування

Результати дослідження можуть бути впроваджені в системи моніторингу рекламного трафіку для захисту користувачів від фішингових та зловмисних вебресурсів, а також використовуватись у кібербезпековій діяльності маркетингових платформ.

Бакалаврська робота містить 76 сторінок, 36 рисунків, 3 розділи список використаних джерел із 38 найменуваннями.

РОЗДІЛ 1. АНАЛІЗ ПРОБЛЕМАТИКИ ПОШИРЕННЯ ШКІДЛИВОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ЗАСОБАМИ ОНЛАЙН-РЕКЛАМИ

1.1. Онлайн-реклама як область для кіберзагроз

Онлайн-реклама посідає центральне місце у формуванні дохідної бази для значної кількості брендів у глобальному масштабі. Процес взаємодії споживачів з продуктами та послугами нерозривно пов'язаний з асоційованими брендами та рекламними повідомленнями, які інтенсивно циркулюють у цифровому просторі. Часто проста згадка або візуальне представлення назви відомої компанії є достатнім чинником для мотивації потенційного клієнта до здійснення покупки або замовлення послуги, що ґрунтується на усталеній довірі до бренду.

Ця довіра, однак, створює сприятливе середовище для суб'єктів зі зловмисними намірами, які можуть експлуатувати репутацію бренду шляхом тактики, відомої як бренджекінг. Бренджекінг у контексті онлайн-реклами часто проявляється у використанні фірмової символіки, назв або візуального стилю легітимних компаній для маскування шкідливих рекламних оголошень – малвертайзингу. Такий підхід значно підвищує ймовірність того, що користувач, довіряючи візуально знайомому бренду, взаємодіятиме з небезпечним оголошенням, що може призвести до встановлення шкідливого програмного забезпечення, фішингових атак або інших кіберінцідентів.

Малвертайзинг (Malvertising) - це поєднання слів "malicious" (зловмисний) та "advertising" (реклама). Малвертайзинг — це практика використання онлайн-реклами для поширення шкідливого програмного забезпечення (малварі) або здійснення інших зловмисних дій.

Замість того, щоб просто показувати рекламу товарів чи послуг, зловмисники впроваджують шкідливий код у рекламні банери, оголошення або перенаправляють користувачів на шкідливі вебсайти. Це може

відбуватися навіть без активної взаємодії користувача з рекламою (так звані "drive-by downloads"), або шляхом обману користувача, щоб він клікнув на рекламу, яка імітує legitimate content або спонукає до певних дій.

Метою малвертайзингу є зараження пристроїв користувачів вірусами-вимагачами, троянами, шпигунським ПЗ, викрадення особистих даних, фінансової інформації тощо

Бренджекінг (Brandjacking) - термін походить від слів "brand" (бренд) та "hijacking" (захоплення). Бренджекінг — це вид шахрайства або зловмисної діяльності, коли хтось використовує ім'я, логотип, символіку або іншу інтелектуальну власність відомого бренду без дозволу з метою обману, введення в оману користувачів або отримання власної вигоди.

Зловмисники можуть створювати фейкові вебсайти, сторінки в соцмережах, електронні листи або рекламу, які виглядають так, ніби вони належать легітимній компанії. Вони роблять це для того, щоб:

- викрасти облікові дані або фінансову інформацію користувачів (фішинг).

- поширювати шкідливе програмне забезпечення.

- продавати підроблені товари.

- завдати шкоди репутації бренду.

Представлена робота присвячена аналізу та вирішенню проблеми виявлення випадків використання зловмисної онлайн реклами, що базується на техніках бренджекінгу, з використанням методів веб-скрепінгу. Основною метою дослідження є розробка та оцінка ефективності веб-скрепера, реалізованого на базі фреймворку Playwright. Цей інструмент розроблено з метою автоматизованого сканування та аналізу рекламних оголошень у веб-середовищі. Ключовим аспектом методології є аналіз технічних атрибутів рекламних оголошень, зокрема цільових URL-адрес, ланцюжків перенаправлень, а також контенту та візуальних елементів самої реклами, для

ідентифікації аномалій та ознак, характерних для шкідливої активності та імітації бренду.

Отримані результати демонструють, що розроблений веб-скрепер успішно ідентифікує випадки виявлення зловмисної онлайн реклами, з особливою увагою до тих, що експлуатують відомі бренди. Це дослідження надає цінні інсайти щодо технічних підходів до автоматизованого виявлення шкідливих рекламних оголошень у динамічному онлайн-середовищі та закладає емпіричну основу для подальших наукових розробок. Подальші напрямки досліджень можуть включати інтеграцію методів машинного навчання для підвищення точності класифікації, розширення охоплення аналізованих рекламних платформ, а також вдосконалення механізмів розпізнавання складніших форм неправомірного використання бренду та обходу систем виявлення.

1.2. Ідентифікація зловмисної онлайн-реклами з використанням імітації фірмової ідентичності. Застосування веб-скрепінгу

У динамічному та постійно змінюваному ландшафті цифрової реклами, поширення шкідливої онлайн-реклами постає як значна загроза безпеці користувачів та підриває довіру до цифрових платформ. Шкідлива онлайн-реклама (малвертайзинг), передбачає інтеграцію шкідливого програмного коду або перенаправлень у легітимні рекламні мережі. Це може призводити до компрометації пристроїв користувачів шкідливим програмним забезпеченням, несанкціонованого доступу до даних та інших негативних наслідків при взаємодії з такими оголошеннями [1].

Особливо оманливою формою шкідливої онлайн-реклами є та, що базується на імітації фірмової ідентичності. У цьому випадку кіберзловмисники несанкціоновано використовують візуальні елементи, назви або символіку відомих брендів для створення рекламних оголошень,

що візуально імітують легітимні. Таким чином, вони експлуатують високий рівень довіри, яку користувачі мають до цих відомих брендів. Ця тактика не лише наражає користувачів на ризики безпеки, але й завдає репутаційної шкоди цільовим компаніям.

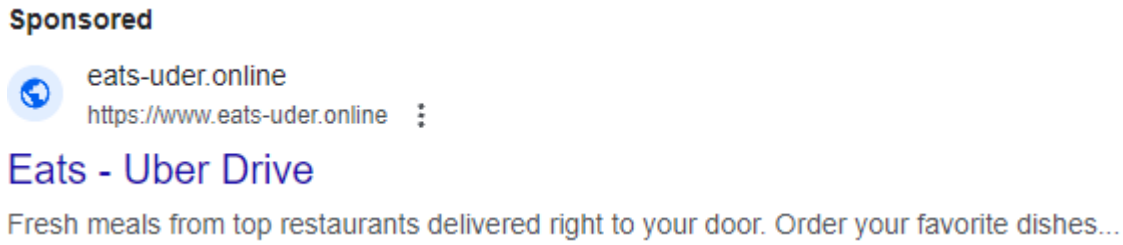

Рисунок 1.1 - Ілюстрація рекламного оголошення, що імітує фірмову ідентичність "Uber Eats"

Як приклад, на рисунку 1.1 представлено оголошення, яке візуально імітує фірмовий стиль платформи доставки їжі "Uber Eats". Після взаємодії з цим рекламним оголошенням користувач перенаправляється на лендінг-сторінку (рисунок 1.2), де йому пропонується потенційне отримання купону та запитується введення персональних даних для входу, таких як номер телефону/електронна пошта та пароль.

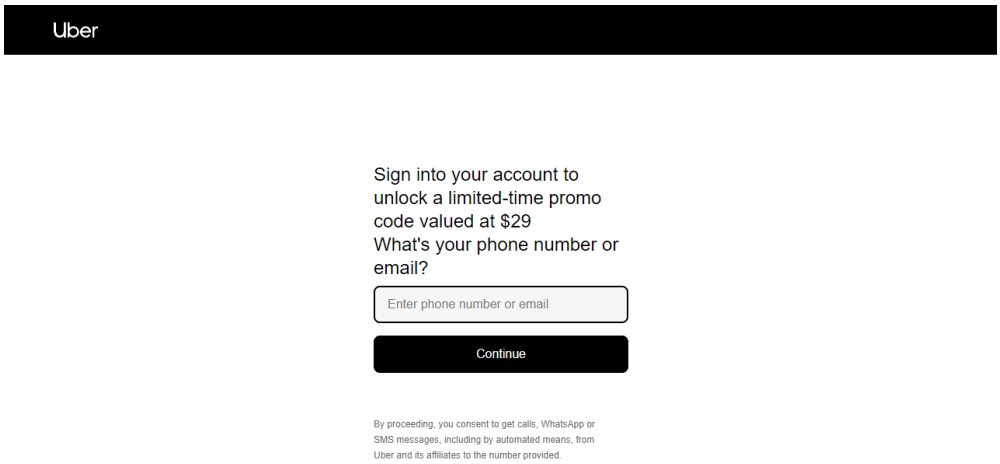

Рисунок 1.2 - Зображення лендінг-сторінки за посиланням eats-uder.online на яку перенаправляється користувач

Незважаючи на неможливість однозначно визначити кінцеві наміри зловмисників на основі лише візуального аналізу, наш інструмент виявив дане оголошення як випадок шкідливої реклами, що застосовує імітацію бренду.

З огляду на вищезазначене, дана робота спрямована на вирішення наступної проблеми: яким чином технології веб-скрепінгу у поєднанні з аналітичними інструментами можуть бути ефективно застосовані для ідентифікації випадків зловмисної онлайн-реклами, яка використовує імітацію фірмової ідентичності.

1.3. Передумови розробки методології ідентифікації випадків зловмисної онлайн-реклами в маркетингових стратегіях

Історичні передумови несанкціонованого використання брендів сягають початку 2000-х років. Однак, незважаючи на зусилля рекламних мереж у протидії зловмисній рекламній діяльності, існуючі методи виявлення часто виявляються неспроможними ефективно реагувати на еволюціонуючі тактики кіберзлочинців [8]. Про це свідчать численні звіти про виявлення такої реклами протягом 2023 року [4] та на початку 2024 року [6]. Дослідження за 2022 рік вказує, що приблизно 45% користувачів клікають на результат пошуку протягом 5 секунд, а майже 74% – протягом 15 секунд. Ця прискорена поведінка користувачів у поєднанні з постійними спробами кіберзлочинців уникнути систем виявлення підвищує вразливість користувачів до взаємодії з оманливою рекламою, яка імітує легітимні маркетингові комунікації брендів. Цей факт підкреслює нагальну потребу в розробці більш просунутих та проактивних механізмів виявлення для ефективного захисту користувачів від цих загроз, що швидко адаптуються.

Незважаючи на зусилля галузевих та академічних спільнот у боротьбі зі зловмисною рекламною діяльністю [1], постійна загроза шкідливої

рекламної діяльності, що використовує імітацію фірмової ідентичності, виявляє критичні прогалини у поточному розумінні та технічних можливостях. Традиційні методи виявлення, такі як фільтрація на основі "чорних списків", зазвичай є реактивними та не відповідають динаміці та складності тактик, які застосовують кіберзлочинці [8]. Дані за 2023 рік свідчать про створення нового фішингового веб-сайту кожні 11 секунд, що робить практично неможливим для реактивних "чорних списків" ефективно протидіяти масштабам та швидкості таких атак.

Існуючі наукові роботи часто розглядають ширший контекст виявлення зловмисної рекламної діяльності загалом [1, 10], але вони не повною мірою аналізують специфічні виклики, пов'язані з імітацією фірмової ідентичності. Періодичні спалахи інцидентів шкідливої реклами підкреслюють рецидивну природу цієї загрози [4 - 6]. Незважаючи на усвідомлення проблеми, спостерігається помітний дефіцит спеціалізованих рішень, спрямованих на ефективне виявлення зловмисної рекламної діяльності, що базується на імітації брендів. Зокрема, існує потреба в інструментах, які інтегрують можливості веб-скрепінгу в режимі реального часу для підвищення точності та оперативності виявлення.

Запропонований у рамках даного дослідження інструмент виконує автоматизовані пошукові запити в Google, генеруючи набір рекламних оголошень, націлених на заздалегідь визначені пошукові терміни. Зібрана інформація про кожне оголошення, включаючи скриншоти, відеозаписи сесії та HAR-файли, зберігається для подальшого аналізу. Цільові домени та відповідні рекламні оголошення проходять класифікацію як шкідливі або безпечні за допомогою інтеграції з чотирма різними платформами виявлення загроз. На завершальному етапі, наш інструмент надає статистичні дані щодо поширеності зловмисних оголошень у різних контекстах, таких як географічне розташування (країна), специфічні пошукові терміни або доменні імена.

1.4. Реклама на основі пошуку та впровадження зловмисної діяльності

Реклама на основі пошуку становить вагомий компонент стратегій цифрового маркетингу, інтегруючи оголошення, що з'являються на сторінках результатів пошукових систем, таких як Google, Bing та Yahoo. Після введення користувачем пошукового запиту, ці системи формують та відображають релевантні рекламні оголошення поруч з органічними результатами пошуку. Позиціонування цих оголошень визначається за допомогою процесу, відомого як торги в реальному часі (Real-Time Bidding, RTB).

Рисунок 1.3 – Модель Real-Time Bidding (RTB)

На рисунку 1.3 показано спрощену модель процесу торгів у реальному часі (Real-Time Bidding, RTB) у цифровій рекламі. Ця модель демонструє, як рекламний простір на веб-сайті або в мобільному додатку продається та купується через автоматизований аукціон за частки секунди.

На рисунку представлені три основні учасники процесу:

1. Publisher (Видавець). Представлений зображенням мобільного пристрою з місцем під рекламу ("AD"). Видавець — це власник веб-сайту або мобільного додатку, який має рекламний простір для продажу. Коли користувач завантажує сторінку або відкриває додаток, генерується Ad request (Запит на рекламу).

2. Real-time bidding (Торги в реальному часі). Представлені зображенням суддівського молотка, що символізує аукціон. Це центральний елемент системи, де відбувається процес торгів. Отримавши Ad request (Запит на рекламу) від Видавця, система RTB надсилає Bid request (Запит ставки) багатьом Рекламодавцям.

3. Advertisers (Рекламодавці). Представлені трьома щитами з написом "BID" (Ставка), кожен з яких відповідає окремому рекламодавцю, готовому купити рекламний простір. Отримавши Bid request (Запит ставки), Рекламодавці оцінюють цінність цього рекламного показу, виходячи з даних про користувача та контексту сторінки, і надсилають свої ставки. На рисунку показані три ставки: $1.15, $1 та $1.25.

Процес відбувається наступним чином:

А. Видавець надсилає запит на рекламу.

Б. Система RTB отримує запит і розсилає запити ставки рекламодавцям.

В. Рекламодавці надсилають свої ставки у відповідь.

Г. Система RTB проводить миттєвий аукціон, де перемагає найвища ставка. На рисунку ставка $1.25 позначена як Winner! (Переможець!).

Д. Після визначення переможця, реклама переможця (з його креативом) надсилається назад до Видавця як Ad delivery (Доставка реклами).

Е. Видавець відображає цю рекламу користувачеві на своєму ресурсі.

Таким чином, рисунок наочно демонструє, як система RTB дозволяє автоматизовано продавати та купувати рекламні покази в реальному часі між видавцями та рекламодавцями на основі аукціонної моделі.

У моделі RTB рекламодавці беруть участь у динамічних аукціонах за рекламний простір, конкуруючи за видимість своїх оголошень на основі їхньої релевантності до конкретного пошукового запиту. Рекламодавці використовують різні атрибути користувачів, включаючи демографічні показники, історію пошукової активності та дані перегляду веб-сторінок, для формування високонаціlених рекламних оголошень. Така таргетована реклама має вищу ймовірність зацікавити потенційних клієнтів, що сприяє підвищенню ефективності рекламних кампаній. Весь цикл, від ініціації пошуку користувачем до відображення персоналізованого рекламного оголошення, відбувається впродовж мілісекунд, що свідчить про надзвичайну динамічність та оперативність цієї екосистеми.

Проте, зловмисна рекламна діяльність порушує функціонування цієї рекламної екосистеми шляхом вбудовування шкідливого коду у візуально легітимні оголошення. Ці оголошення можуть експлуатувати вразливості в процесах доставки реклами, що потенційно призводить до зараження пристроїв шкідливим програмним забезпеченням, несанкціонованого доступу до даних або перенаправлення на фішингові веб-сайти при взаємодії користувачів з оголошеннями [10]. Оскільки пошукові системи інтегрують рекламні оголошення з органічними результатами, зловмисна рекламна діяльність становить значну загрозу для користувачів, які можуть ненавмисно клікнути на шкідливе оголошення, приймаючи його за легітимне.

Наявність масштабних рекламних мереж та високоавтоматизований характер розміщення реклами за допомогою RTB створюють широкі можливості для суб'єктів, що займаються зловмисною рекламною діяльністю, для проникнення в ці мережі. Використовуючи довіру користувачів до відомих брендів (часто шляхом імітації фірмової ідентичності), виконавці зловмисної рекламної діяльності можуть ефективніше поширювати свій

шкідливий контент, що робить це явище однією з ключових загроз у сфері цифрової реклами.

Протидія зловмисній рекламній діяльності є нетривіальним завданням. З метою уникнення виявлення та максимізації ефективності своїх кампаній, виконавці зловмисної рекламної діяльності застосовують різноманітні техніки. Одним із поширених методів є маскування (cloaking), коли контент, що відображається, варіюється залежно від характеристик користувача або системи-перевірки. Це дозволяє шкідливим веб-сайтам залишатися непоміченими для автоматизованих систем модерації рекламних мереж. Цей метод передбачає таргетування певних категорій користувачів або профілів, яким демонструється шкідливий вміст, тоді як перевіряючим системам або іншим користувачам показується безпечна версія. Таким чином, зловмисна рекламна діяльність може успішно проходити початкові перевірки та потрапляти у легітимні рекламні мережі.

Крім того, суб'єкти, що стоять за зловмисною рекламною діяльністю, часто змінюють свої тактики та інфраструктуру, щоб випереджати методи виявлення. Це включає регулярну зміну доменних імен, використання короткострокових кампаній та застосування методів рекламного шахрайства, таких як клік-фрод та імпресіон-фрод. Ці адаптивні техніки значно ускладнюють завдання рекламних мереж щодо оперативного виявлення та блокування зловмисної реклами.

1.5. Аналіз існуючих досліджень і методологій виявлення зловмисної рекламної діяльності

У цьому розділі акцент зроблено на аналізі існуючих досліджень, присвячених виявленню та пом'якшенню зловмисної рекламної діяльності. Шляхом розгляду методологій, отриманих результатів та обмежень цих наукових праць, ставиться за мету осмислити поточний стан досліджень у

широкому контексті галузі та підкреслити унікальні внески підходу, представленого в даній роботі, зокрема стосовно виявлення зловмисної рекламної діяльності, що базується на імітації фірмової ідентичності.

Галузь виявлення та нейтралізації зловмисної рекламної діяльності стала об'єктом значних досліджень та розробок в останні роки, значною мірою з використанням інструментарію веб-скрепінгу. Багато наукових робіт використовували технології веб-скрепінгу для збору та аналізу даних з різноманітних онлайн-джерел, що має схожість із підходом, прийнятим у цій роботі. Однак, інструмент веб-скрепінгу, розроблений у межах даного дослідження, відрізняється своєю унікальною архітектурою та спрямованістю, оскільки зосереджується на аналізі рекламної мережі Google, а не на скануванні окремих веб-сайтів. Цей вибір зумовлений наявністю значної кількості досліджень, присвячених другій категорії. Таке сфокусоване спрямування дозволяє поточному дослідженню внести нові аналітичні інсайти та стратегії в досліджувану галузь.

У роботі [1] було досліджено тактики, які застосовуються кіберзлочинцями для розповсюдження зловмисної рекламної діяльності в інтернеті, зокрема використання обфускації та часті зміни для уникнення виявлення [1]. Автори розробили фреймворк, що інтегрує статичний та динамічний аналіз для дослідження рекламних оголошень та поведінки відповідних цільових сторінок. Їхня методологія включала скрепінг топ-90 000 веб-сайтів за рейтингом Alexa, відтворення ланцюжків перенаправлень оголошень, ідентифікацію шляхів доставки реклами, анотування вузлів з різними атрибутами та застосування підходу машинного навчання для генерації правил виявлення зловмисної рекламної діяльності. Розроблений ними інструмент MadTracer продемонстрував високий рівень виявлення при збереженні низького рівня хибнопозитивних спрацьовувань. Це дослідження яскраво проілюструвало складну та постійно еволюціонуючу природу

зловмисної рекламної діяльності та підкреслило необхідність використання багатоаспектних підходів для ефективної боротьби з цією загрозою.

В [2] здійснили поглиблене дослідження сфери зловмисної рекламної діяльності, маючи на меті розкриття прийомів та тактик, що використовуються кіберзлочинцями. На початковому етапі було сформовано набір даних шляхом проведення масштабного веб-скрепінгу для збору понад 600 000 реальних рекламних оголошень. Веб-сайти для скрепінгу були отримані з двох різних джерел: одне надане антивірусною компанією, а друге – список топ-1 мільйона сайтів Alexa. Відвідування цих сайтів здійснювалося за допомогою автоматизованого браузера Selenium, при цьому фіксувався контент оголошень та HTTP-трафік. Дослідники також розробили оракул, що складається з трьох компонентів, який виконував автоматичну класифікацію зібраних оголошень. Це дослідження виявило, що у більшості випадків взаємодія між видавцями та рекламодавцями ґрунтується на довірі, що означає рідкісне застосування видавцями додаткових механізмів для фільтрації шкідливих оголошень. Також було показано, що деякі рекламні біржі обслуговують більшу кількість зловмисних оголошень порівняно з іншими, що пояснюється недостатньою ефективністю систем виявлення та процесами арбітражу реклами, які спрощують проникнення шкідливих оголошень на рекламні біржі.

В роботі [3] провели дослідження поширеності та характеру технологій трекінгу та зловмисної рекламної діяльності на веб-сайтах, призначених для дитячої аудиторії. Вони виявили значну кількість таких веб-сайтів, що містять технології трекінгу, які збирають дані без явної згоди, часто порушуючи таким чином нормативні акти про конфіденційність, розроблені для захисту неповнолітніх. Дані були зібрані шляхом скрепінгу списку з 2000 дитячих веб-сайтів. Збиралася інформація про самі оголошення, а також про рекламодавців (за допомогою розділу "Чому це оголошення?"). Отримані дані аналізувалися за допомогою багатомовних мовних моделей для

класифікації оголошень, які могли бути проблематичними для дітей. Дослідження показало, що 27% веб-сайтів містили оголошення, для відображення яких потрібна була явна згода батьків.

В [4] представили систему, розроблену для виявлення зловмисної рекламної діяльності шляхом комбінування функціональності трьох онлайн-інструментів для виявлення шкідливого програмного забезпечення: VirusTotal, URLVoid та TrendMicro [12].

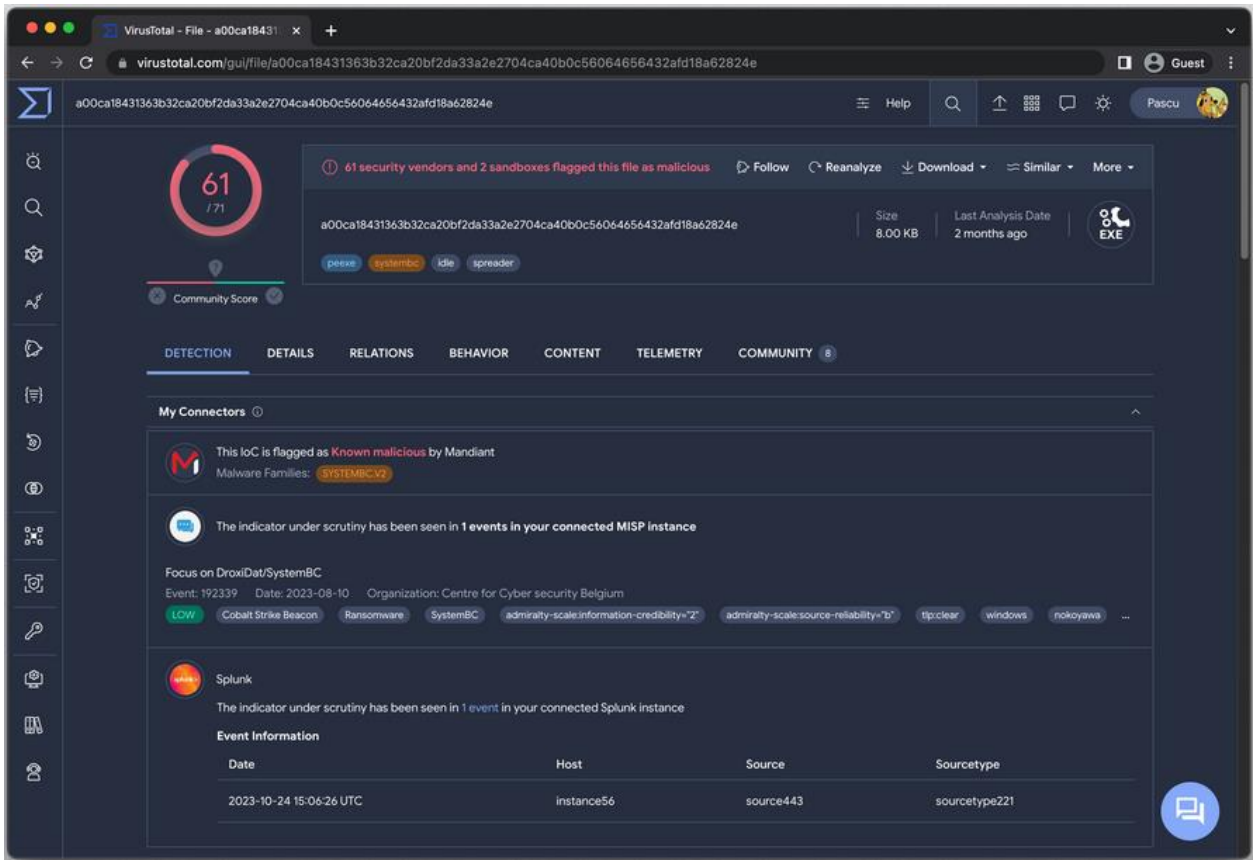

Рисунок 1.4 – VirusTotal - веб-ресурс перевірки на шкідливе програмне забезпечення

Їхня методологія передбачала вибір веб-сайтів з двох різних джерел даних (топ-1 мільйон сайтів Alexa та "чорний список") та використання Selenium для автоматизації веб-браузера. Інструмент витягував URL-адреси рекламних оголошень та надсилав їх на аналіз до трьох зазначених платформ. Кожне оголошення класифікувалося на основі отриманих результатів. Це

дослідження виявило URLVoid як більш надійний інструмент порівняно з двома іншими. Водночас, було підкреслено, що жоден окремий інструмент не є універсально "найкращим", і що досягнення вищої точності можливе лише шляхом комбінування функціональності різних платформ.

Рисунок 1.5 - Приклад зловмисного рекламного оголошення, що розповсюджується через веб-push сповіщення

В [5] дослідили зростаючу проблему використання push-сповіщень як значного вектора для доставки шкідливого контенту. У своїй роботі вони представили PushAdMiner, систему для автоматизованого збору та аналізу веб-push сповіщень (Web Push Notifications, WPNs). Вони розширили існуючі браузерні скрепери на базі Chromium для моніторингу Service Workers та WPNs. Скрепер відвідував веб-сайти та надавав дозволи на отримання сповіщень.

В результаті було зібрано WPNs та відповідні цільові сторінки, які потім піддавалися аналізу. Subramani et al. ідентифікували загалом 5143 WPN-оголошення та класифікували понад 50% з них як зловмисні. Це дослідження стало першим систематичним дослідженням цього конкретного вектора атаки в контексті зловмисної рекламної діяльності. Висока неефективність традиційних блокувальників реклами та URL-фільтрів у поєднанні зі значним обсягом зібраних даних продемонструвала гостру необхідність у вдосконаленні методів виявлення.

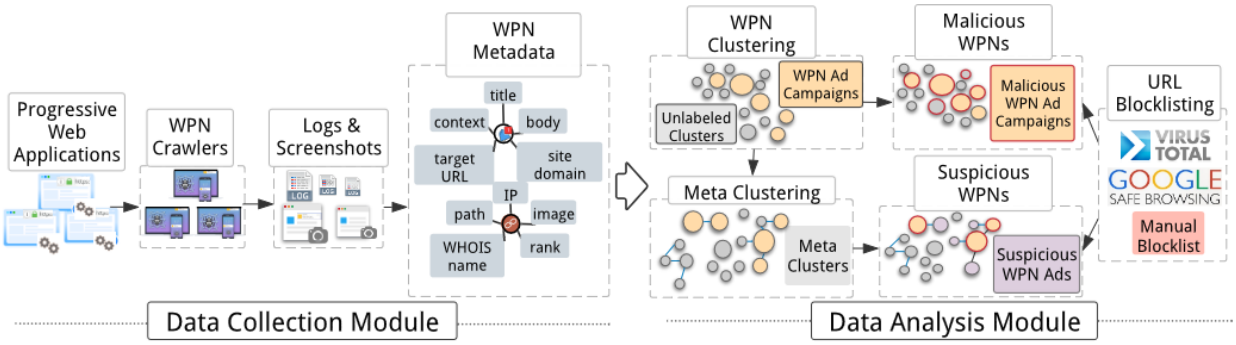

Рисунок 1.6 - Огляд PushAdMiner системи

Дослідження [6] оцінювало ефективність різних інтернет-сервісів у виявленні та категоризації зловмисної рекламної діяльності. Дослідники модифікували свій інструмент Katti для скрепінгу веб-сайтів та захоплення HTTP-запитів, пов'язаних з онлайн-рекламою.

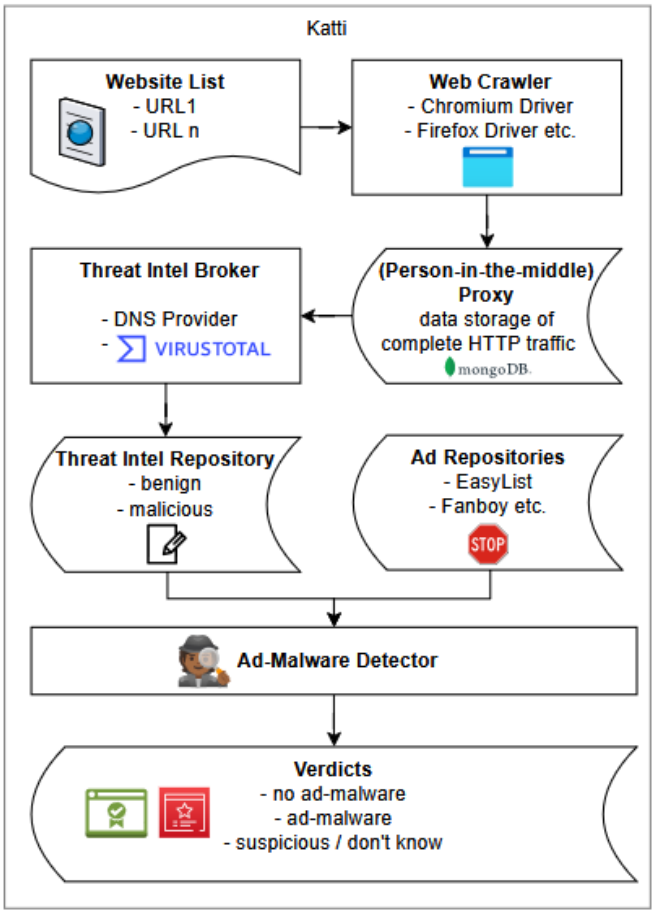

Рисунок 1.7 - Огляд підходу до виявлення рекламного шкідливого програмного забезпечення

Ці запити потім перевірялися за допомогою фільтруючих DNS-провайдерів та VirusTotal, а відповіді різних інтернет-сервісів порівнювалися для оцінки їхньої здатності визначати зловмисну рекламну діяльність.

Шляхом використання зазначених сервісів було досліджено, як вони маркують підозрілий контент. Результати виявили значні розбіжності у класифікації зловмисної рекламної діяльності різними сервісами, причому деякі сервіси маркували більшу кількість доменів, ніж інші. Дослідження підкреслило необхідність стандартизованих визначень та більш прозорих критеріїв у виявленні зловмисної рекламної діяльності, а також важливість врахування повної структури URL у подальших дослідженнях.

Хоча ці дослідження роблять значний внесок у галузь виявлення та пом'якшення зловмисної рекламної діяльності, в існуючій літературі все ще залишаються помітні прогалини, які прагне заповнити поточне дослідження. Зосереджуючись на розробці унікального інструменту веб-скрепінгу та оригінальних методологій виявлення, це дослідження має на меті надати нові аналітичні інсайти та ефективні стратегії для боротьби, зокрема, зі зловмисною рекламною діяльністю, що використовує імітацію фірмової ідентичності.

Незважаючи на огляд загальних методів виявлення зловмисної рекламної діяльності, існує недостатньо досліджень, що сфокусовані саме на виявленні цієї загрози в контексті великих рекламних мереж, де активно використовуються техніки імітації брендів для обману користувачів.

На основі обговорених у цій главі пов'язаних робіт, наступна глава більш детально представить методологію, використану в даному дослідженні. Буде надано детальний опис архітектури та функціональності розробленого інструменту веб-скрепінгу, його процесу збору даних та застосованих технік аналізу для виявлення зловмисних рекламних оголошень.

Висновки до першого розділу

У першому розділі було здійснено ґрунтовний аналіз проблематики поширення шкідливого програмного забезпечення через онлайн-рекламу, що набуває все більшої актуальності в умовах цифрової трансформації суспільства. Розглянуто ключові аспекти, які роблять онлайн-рекламу привабливим інструментом для кіберзлочинців, зокрема широке охоплення аудиторії, динамічність контенту та складність контролю джерел походження рекламних матеріалів.

Підрозділ 1.1 висвітлює онлайн-рекламу як потенційне джерело кіберзагроз, зокрема через використання рекламних мереж для поширення шкідливих посилань або фішингових повідомлень. У підрозділі 1.2 розглянуто техніки ідентифікації зловмисної реклами, що маскується під легітимну за допомогою фірмового стилю брендів, а також можливості веб-скрепінгу для збору релевантних даних.

У підрозділі 1.3 визначено передумови для створення ефективної методології виявлення шкідливої реклами в межах маркетингових стратегій, що враховує як технічні, так і соціальні аспекти впливу. Розділ 1.4 проаналізував загрози, пов’язані з рекламою на основі пошуку, яка використовується для маніпуляції результатами видачі з метою перенаправлення користувачів на небезпечні ресурси. У підрозділі 1.5 було здійснено огляд існуючих наукових і практичних підходів до виявлення зловмисної рекламної діяльності, зокрема методів машинного навчання, аналізу поведінки користувачів і сигнатурних моделей.

Таким чином, розділ сформував цілісне уявлення про масштаби та складність проблеми поширення шкідливого ПЗ через онлайн-рекламу, а також обґрунтував необхідність розробки нових, більш адаптивних методів ідентифікації таких загроз у динамічному веб-середовищі.

РОЗДІЛ 2. РОЗРОБКА МЕТОДОЛОГІЇ ЗАПОБІГАННЯ ПОШИРЕННЯ ШКІДЛИВОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

2.1. Опис пропонованої методології

Цей розділ описує методичний підхід, застосований для розробки, впровадження та оцінки веб-скрепера, призначеного для виявлення зловмисної реклами, що використовує імітацію бренду. Методологія охоплює етапи проєктування та розробки веб-скрепера, процедуру збору даних, а також методи аналізу зібраної інформації. Кожен етап детально описаний з метою забезпечення повного розуміння застосованих методів та гарантування надійності й ефективності інструменту в ідентифікації шкідливих оголошень. Загальний огляд функціональності інструменту представлений на рисунку 2.1.

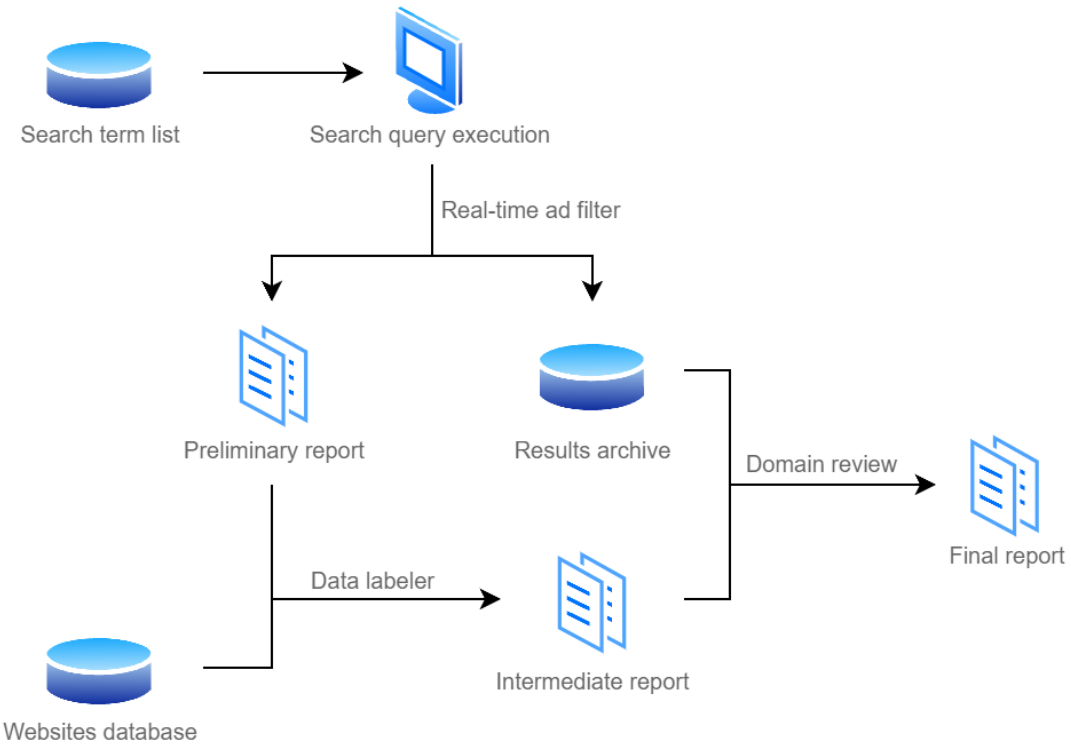

Рисунок 2.1 - Схема функціональності інструменту виявлення зловмисної реклами

На рисунку 2.1 представлена схема функціональності інструменту для виявлення зловмисної реклами. Процес розпочинається з переліку пошукових термінів (Search term list), який слугує вхідними даними для етапу виконання пошукових запитів (Search query execution). На цьому етапі інструмент виконує пошук, ймовірно, в пошукових системах, використовуючи надані терміни.

Результати пошуку далі надходять до фільтру оголошень у реальному часі (Real-time ad filter). Цей компонент обробляє отримані рекламні оголошення. Відфільтровані дані розгалужуються на два потоки: формування попереднього звіту (Preliminary report) та збереження до архіву результатів (Results archive), який, є базою даних для довгострокового зберігання зібраної інформації.

Наступним кроком є обробка даних з попереднього звіту компонентом мітки даних (Data labeler). Цей компонент також використовує інформацію з бази даних веб-сайтів (Websites database), яка містить перелік легітимних доменів. Компонент мітки даних аналізує попередній звіт, порівнюючи домени оголошень з базою легітимних сайтів, і формує проміжний звіт (Intermediate report), який містить позначки або класифікацію відповідних оголошень.

Фінальний етап, перегляд доменів (Domain review), обробляє дані як з архіву результатів, так і з проміжного звіту. На цьому етапі відбувається остаточний аналіз та верифікація виявлених оголошень. Результатом цього етапу є кінцевий звіт (Final report), що містить підсумкові дані про виявлену зловмисну рекламу.

Таким чином, схема ілюструє послідовний процес від ініціації пошуку за заданими термінами до формування фінального звіту про виявлені шкідливі оголошення, включаючи фільтрацію, архівування та аналіз даних з використанням бази легітимних сайтів.

2.2. Проектування інструменту

Веб-скрепер був розроблений з використанням Playwright, потужного фреймворку для автоматизації веб-взаємодії, що імітує дії користувача. Його функціональні можливості подібні до інших інструментів, таких як Selenium або Cypress.

2.2.1. Опис фреймворку Playwright

Playwright — це сучасний фреймворк для автоматизації браузерів, розроблений компанією Microsoft. Його основне призначення — написання end-to-end (E2E) тестів для вебзастосунків.

Основні особливості Playwright:

- працює з Chromium (Chrome, Edge), Firefox та WebKit (Safari), що дозволяє тестувати кросбраузерну сумісність.

- підтримка мов програмування JavaScript, TypeScript, Python, Java.

- дозволяє імітувати мобільні пристрої, геолокацію, зміну мережевого з’єднання, часовий пояс тощо.

- висока продуктивність завдяки одночасному запуску кількох тестів.

- playwright автоматично чекає на завантаження елементів або завершення анімацій, що зменшує потребу у «sleep» або ручних таймерах.

- має інструменти для автоматичного запису дій користувача та генерації коду тестів.

Лістинг 2.1. Приклад простого тесту на JavaScript

```
const { test, expect } = require('@playwright/test');

test('перевірка заголовка сторінки', async ({ page }) => {
  await page.goto('https://example.com');
  await expect(page).toHaveTitle(/Example Domain/);
});
```

Основні переваги полягають в легкій інтеграції в CI/CD-процеси, інтуїтивному API, підтримці тестування SPA (Single Page Applications).

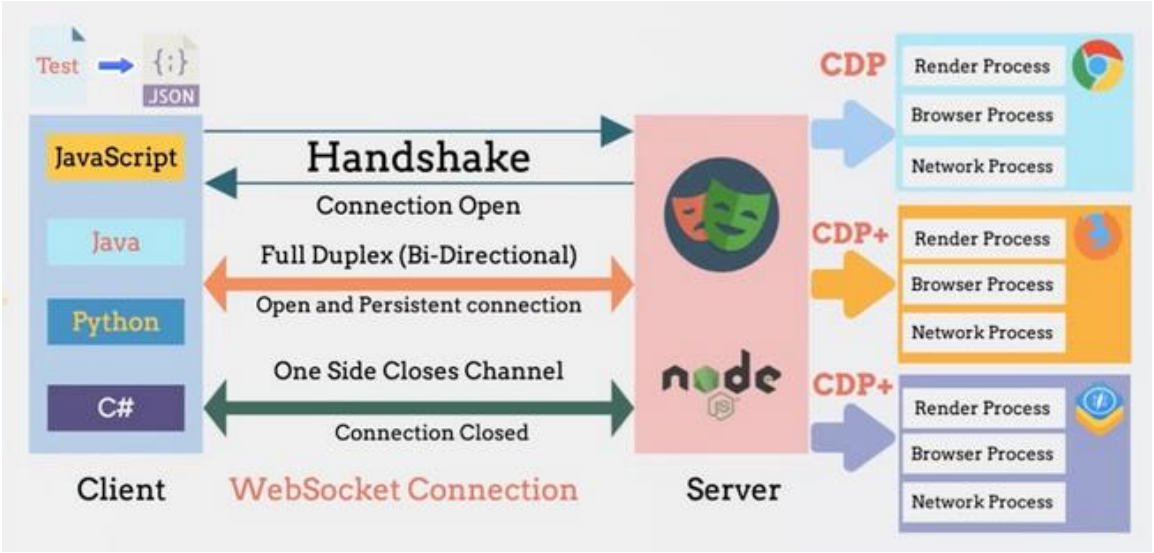

Рисунок 2.2 – Playwright архітектура

Рисунок 2.2 візуалізує ключові архітектурні принципи роботи Playwright. Він демонструє, як користувацький код взаємодіє з браузерами через проміжний шар.

Розглянемо основні елементи рисунка та їх відповідність компонентам Playwright.

Client (Клієнт) - цей блок представляє програму або тест-скрипт, написаний з використанням Playwright API. Як показано, Playwright підтримує множину мов програмування, що відповідає різним "каналам" від клієнта: JavaScript/TypeScript (для Node.js), Python, Java, та C# (для .NET). Ваші команди автоматизації (наприклад, перейти на сторінку, клікнути елемент, ввести текст), які можна розглядати як "Test" або дані у форматі "JSON", формуються на цьому рівні.

WebSocket Connection (З'єднання WebSocket) - це критично важливий елемент архітектури Playwright. Він встановлює постійне (Persistent) та двонаправлене (Full Duplex / Bi-Directional) з'єднання між клієнтською частиною (вашим скриптом) та серверним процесом Playwright. На відміну

від старіших протоколів автоматизації (наприклад, WebDriver до стандарту W3C), Playwright використовує це постійне з'єднання для швидкого та ефективного обміну командами та подіями між вашим кодом і браузером без необхідності багаторазового встановлення зв'язку. Етапи "Handshake", "Connection Open/Closed" ілюструють життєвий цикл цього з'єднання.

Server (Сервер) - цей компонент відповідає фоновому процесу Playwright, який запускається, коли ви вперше виконуєте скрипт Playwright. Він реалізований на Node.js, що підтверджується логотипом. Роль цього сервера полягає в отриманні команд від клієнта через WebSocket-з'єднання, трансляції цих команд у специфічні для браузера інструкції та керуванні запущеними екземплярами браузерів.

CDP / CDP+ (Протоколи керування браузером) - сервер Playwright взаємодіє безпосередньо з запущеними екземплярами браузерів. Для цього використовуються протоколи автоматизації:

- CDP (Chrome DevTools Protocol) - використовується для керування браузерами на базі Chromium (Chrome, Edge). Цей протокол дозволяє глибоко взаємодіяти з внутрішніми процесами браузера.

- CDP+ - це узагальнене позначення може стосуватися власних протоколів автоматизації, які Playwright використовує для Firefox та WebKit. Хоча вони не є "CDP" в чистому вигляді, їхня функціональність аналогічна – вони дозволяють Playwright контролювати Render Process (Процес візуалізації), Browser Process (Головний процес браузера) та Network Process (Мережевий процес) кожного екземпляра браузера. Різні іконки браузерів праворуч підкреслюють здатність Playwright автоматизувати різні типи браузерів.

Таким чином, рисунок 2.2 точно представляє багатошарову архітектуру Playwright, де клієнтський код високого рівня спілкується з низькорівневими процесами браузера через ефективне постійне з'єднання та сервер-

оркестратор, що використовує нативні протоколи автоматизації браузерів. Це забезпечує швидкість, надійність та крос-браузерну сумісність Playwright.

2.2.2. Налаштування конфігурації системи та виконання пошукових запитів

Інструмент було конфігуровано для функціонування в середовищі браузерів на основі Chromium. Різноманітні параметри були тонко налаштовані для точнішої симуляції реальної взаємодії користувача. Зокрема, скрепер працює з вимкненим безголовим режимом (headless mode) та імітує паузи між виконанням команд. Ця конфігурація була реалізована для мінімізації ймовірності ідентифікації як автоматизованого агента (бота) та забезпечення репрезентативності отриманих результатів щодо типового користувацького досвіду.

Скрепер був запрограмований для виконання пошукових запитів за попередньо визначеними термінами, пов'язаними з популярним програмним забезпеченням та веб-сайтами, що містилися у словнику, який нараховував 50 термінів. Цей список був сформований на основі двох джерел даних (дата-фідів): списків Tranco та Kantar. Кожен термін у словнику асоціюється з переліком дозволених (whitelist) відомих легітимних доменів, які надалі використовуються для аналізу оголошень. Ця допоміжна інформація представлена у Додатку А.

Одним із ключових критеріїв, на якому Google базує показ оголошень, є географічне місцезнаходження користувача. З метою отримання повнішого уявлення та максимізації шансів на виявлення оголошень, що використовують імітацію бренду, було необхідним дослідження різних локацій. Це досягається за допомогою сервісу Mullvad VPN.

Mullvad VPN — це провайдер послуг віртуальної приватної мережі (VPN), відомий своїм сильним акцентом на приватність, анонімність та безпеку користувачів.

Замість традиційних імен користувача та паролів або прив'язки до електронної пошти, Mullvad генерує унікальний 16-значний номер облікового запису при реєстрації. Цей номер є єдиним ідентифікатором вашого облікового запису в системі Mullvad, що додатково підвищує анонімність

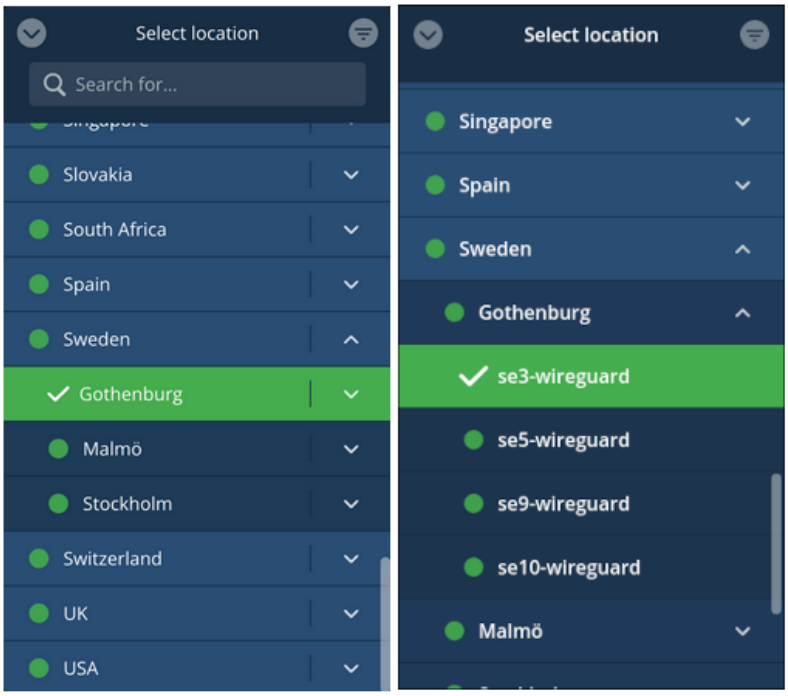

Рисунок 2.3 – Вибір сервера засобами Mullvad VPN

Кожен цикл збору даних виконується з 10 різних країн, а результати записуються незалежно. Список країн включає: США (us), Канада (ca), Австралія (au), Велика Британія (gb), Нідерланди (nl), Швеція (se), Румунія (ro), Бразилія (br), Південна Африка (za), Таїланд (th).

2.3. Процес збору даних

Процес збору даних включає збір оголошень з результатів пошукових систем на основі заздалегідь визначених пошукових термінів. Цей розділ детально описує процедури, використані для захоплення відповідних даних,

включаючи виявлення оголошень, збір інформації (наприклад, скриншоти, відеозаписи, HAR-файли) та методи, використані для забезпечення точності та релевантності зібраних даних. Мета — створити всеохоплюючий набір даних для наступного аналізу.

Інструмент прокручує сторінку результатів пошуку до обраного пункту та відображає всі оголошення. Потім він обробляє кожне оголошення окремо та оцінює його релевантність. Цей фільтр мінімізує кількість помічених оголошень, оцінюючи, наскільки кожне оголошення відповідає пошуковому терміну або його асоційованому відомому домену. Оскільки мета скрепера — виявляти спроби несанкціонованим використанням брендів, всі оголошення, які не схожі на пошуковий термін, вважаються нерелевантними. Наприклад, якщо скрепер аналізує оголошення для "Amazon", він буде перевіряти лише ті оголошення, які містять пошуковий термін або його асоційований відомий домен.

Коли знайдено цікаве оголошення, скрепер захоплює скриншоти ключової інформації, включаючи саме оголошення (рисунок 2.4) та розділ "Чому це оголошення?".

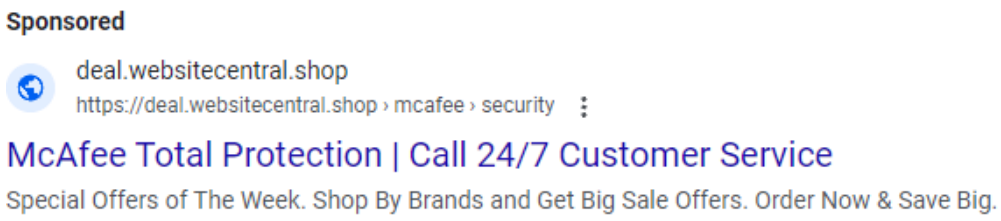

Рисунок 2.4 – Приклад оголошення

Цей розділ містить інформацію про рекламодавця (рисунок 2.5) та надає уявлення про те, за якими критеріями пошукова система відобразила оголошення (рисунок 2.6). Під час цього кроку також записується ідентифікатор рекламодавця, доступний через кнопку "Подивіться більше оголошень". Цей ідентифікатор можна використати пізніше для перегляду всіх оголошень, які зараз розміщує ця компанія або особа (рисунок 2.7).

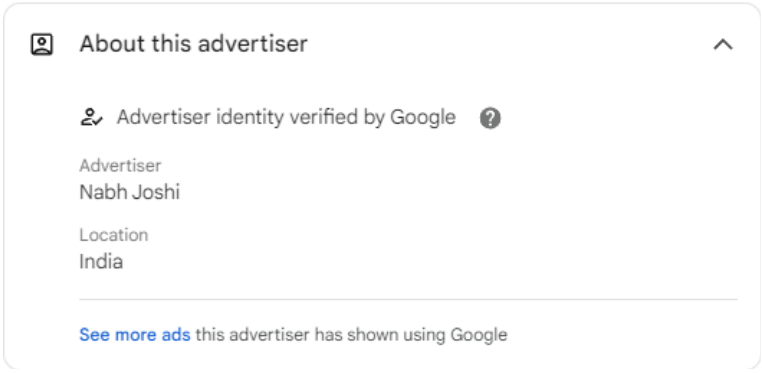

Рисунок 2.5 - Інформація про рекламодавця

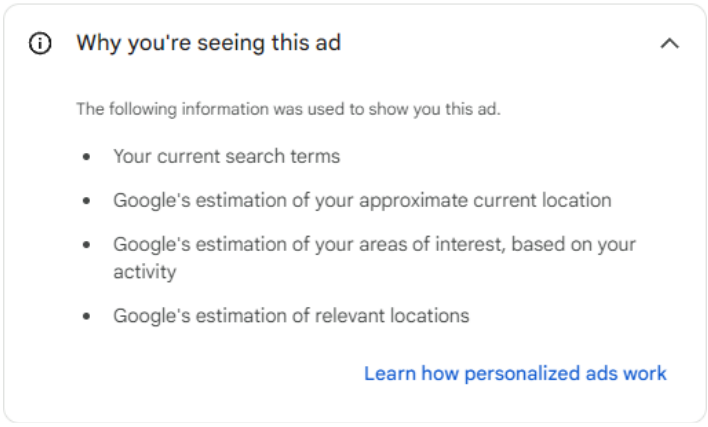

Рисунок 2.6 - Критерії пошукової системи

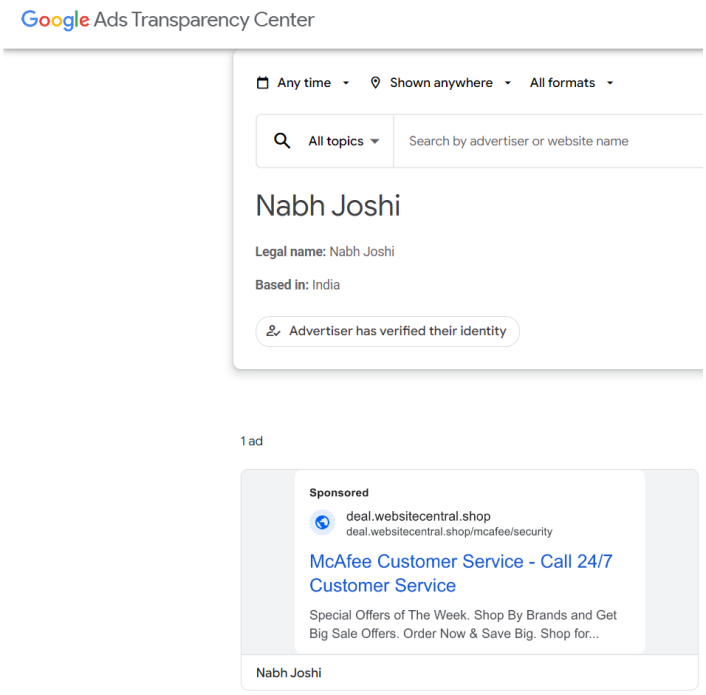

Рисунок 2.7 - Фрагмент сторінки "Показати більше оголошень"

Також у цей момент у процесі увімкнено слухач запитів. Цей слухач призначений для збереження ланцюжка перенаправлення для будь-яких навігаційних запитів браузера. Деякі оголошення можуть вирішити в реальному часі, що скрепер не входить до цільової аудиторії, і перенаправити його на легітимну лендінг-сторінку. Ланцюжок перенаправлення допомагає захопити будь-які підозрілі веб-сайти в процесі. Наступним кроком є натискання на оголошення. Потім скрепер робить скриншот лендінг-сторінки веб-сайту (рисунок 2.8). Він також зберігає лендінг-URL для подальшого огляду.

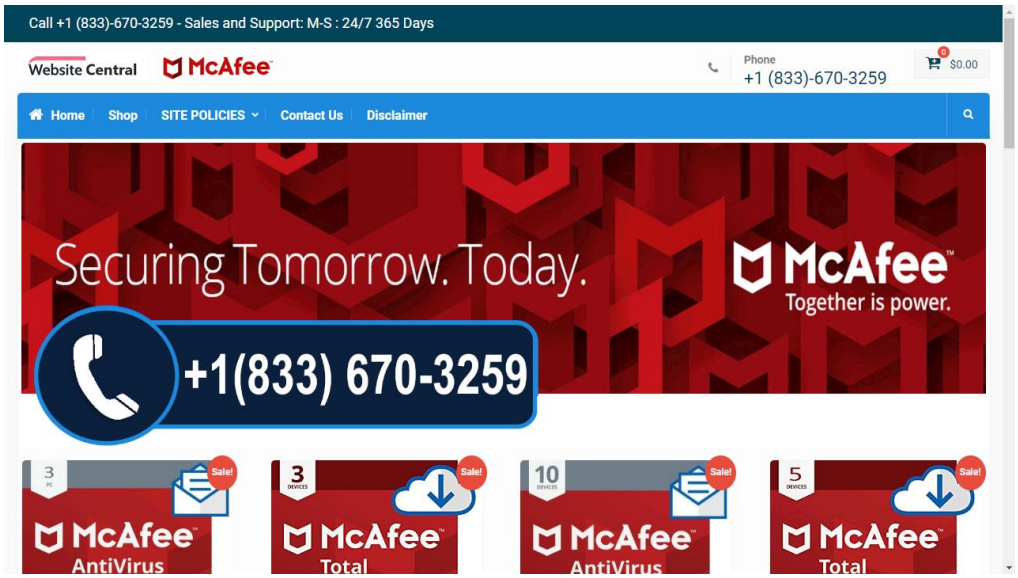

Рисунок 2.8 - Знімок екрана цільової сторінки, захоплений скрепером

Під час усього виконання пошукових запитів скрепер також налаштований на запис відео кожної сторінки [28]. Таким чином, після завершення запису лендінг-сторінок та головної сторінки зберігаються для подальшого посилання. Ці записи корисні для аналізу непередбачуваних граничних випадків та слугують доказом, якщо лендінг-сторінка стане недоступною в майбутньому.

Для полегшення більш детального аналізу HTTP-трафіку та перенаправлень скрепер також зберігає HTTP-архів веб-браузера. Якщо

пошуковий термін цікавий (тобто для нього ведуться рекламні кампанії), скрепер записує HAR-файл усієї взаємодії. Це дозволяє відтворити середовище, в якому були знайдені оголошення, а також провести подальший аналіз будь-яких підозрілих запитів та сторінок.

2.4. Аналіз та оцінка зібраних рекламних оголошень

Цей розділ присвячений опису технік, застосованих для аналізу зібраного масиву рекламних оголошень. Мета аналізу полягає у верифікації їхньої легітимності та ідентифікації випадків зловмисної рекламної діяльності, зокрема тих, що базуються на імітації фірмової ідентичності. Процес аналізу включає перевірку цільових URL-адрес, а також детальний розгляд лендінг-сторінок та ланцюжків перенаправлень. Зважаючи на специфіку зловмисної рекламної діяльності, що використовує імітацію брендів, може виникати необхідність у проведенні остаточної ручної верифікації для підтвердження характеру виявлених випадків.

Після завершення повного циклу сканування, тобто виконання всіх запланованих пошукових запитів, результати автоматично піддаються оцінці розробленим скрепером. Цей етап реалізовано з метою представлення лише релевантної інформації та мінімізації часових і ресурсних витрат на експертну перевірку доменів. Різноманітні зібрані дані порівнюються із попередньо визначеним "дозволеним списком" (whitelist), на основі чого генеруються відповідні попередження (alerts). На початковому етапі, цільова URL-адреса (landing URL) кожного рекламного оголошення порівнюється з відомим доменом або доменами, асоційованими з конкретним пошуковим терміном, за допомогою регулярних виразів. У випадку виявлення невідповідності генерується попередження. Надалі, ланцюжок перенаправлень (redirect chain) оголошення також піддається аналізу та порівнянню з "дозволеним списком" доменів та допустимих перенаправлень.

Якщо в послідовності запитів виявляється невідомий веб-сайт, генерується відповідне попередження. Система передбачає можливість легкої інтеграції додаткових автоматизованих перевірок для підвищення точності.

Система генерації попереджень функціонує у двох режимах: "суворому" (strict) та "послабленому" (relaxed). У суворому режимі цільова URL-адреса кожного оголошення зіставляється виключно з відомим доменом, що відповідає конкретному пошуковому терміну, за яким було знайдене оголошення. Аналогічно, кожне перенаправлення в ланцюжку оголошення порівнюється лише з відомим доменом конкретного терміну та обмеженим списком дозволених перенаправлень. У послабленому режимі як цільова URL-адреса, так і всі елементи ланцюжка перенаправлень порівнюються з усіма відомими доменами, що присутні у "дозволеному списку" скрепера. Використання послабленої версії системи попереджень призводить до меншої кількості помічених оголошень, але не впливає на кількість ідентифікованих зловмисних випадків. Обґрунтування такого підходу полягає у припущенні, що перехід на веб-сайт, який включено до загального "дозволеного списку" доменів, не є шкідливим, за винятком вкрай малоймовірної ситуації компрометації одного з легітимних відомих доменів.

Результати виконання тестового запуску системи відображаються у вигляді структурованого звіту. Цей звіт містить релевантну інформацію для кожного зібраного рекламного оголошення, згруповану за відповідними пошуковими термінами. Звіт також надає статистику щодо кількості оголошень, виявлених за кожним терміном, та перелік згенерованих попереджень. Для кожного попередження відображається відповідна контекстна інформація; наприклад, при спрацьовуванні попередження, пов'язаного з URL, у звіті також вказується цільова URL-адреса. Аналогічний принцип застосовується до інформації про ланцюжок перенаправлення. Для полегшення подальшого аналізу та розслідувань, особливо у випадках генерації попереджень, у звіті також відображається ідентифікатор

рекламодавця. Фрагмент такого звіту можна переглянути нижче для кращого уявлення про формат представлення результатів. Розбіжність між загальною кількістю знайдених оголошень та кількістю оголошень, відображених у звіті, пояснюється застосуванням фільтрації. Відсутність 4 оголошень у звіті свідчить про те, що вони були визнані нерелевантними до конкретного пошукового терміну, відповідно до критеріїв фільтрації.

```
avast - Знайдено оголошень: 9
    Оголошення 1: 0/2 попереджень
    Оголошення 3: 0/2 попереджень
    Оголошення 4: 1/2 попереджень
        Ланцюжок перенаправлення: {'mcafeeinc.demdex.net', 'www.mcafee.com', ...}
        Ідентифікатор прозорості оголошень: AR07041635852870483969? origin=ata
    Оголошення 6: 0/2 попереджень
    Оголошення 7: 2/2 попереджень
        Лендінг-URL: https://blitzhandel24.co.uk/avast/[...]
        Ланцюжок перенаправлення: {'monitor.clickcease.com', 'blitzhandel24.co.uk',
    Ідентифікатор прозорості оголошень: AR03282036780072697857? origin=ata
```

Рисунок 2.9 – Приклад попереднього звіту

2.4.1. Позначення (маркування) даних

У попередньому звіті кількість помічених доменів досить висока. Наприклад, найбільша частина помічених доменів складається з магазинів-аутлетів. Терміни, такі як "samsung" або "nike", призводять до значної кількості помічених оголошень через різні магазини-аутлети, які продають ці продукти. Приклад такого оголошення можна побачити на рисунку 2.10. Скрепер правильно помічає ці оголошення, оскільки вони не ведуть на відомий домен. Однак це не означає, що оголошення є шкідливим.

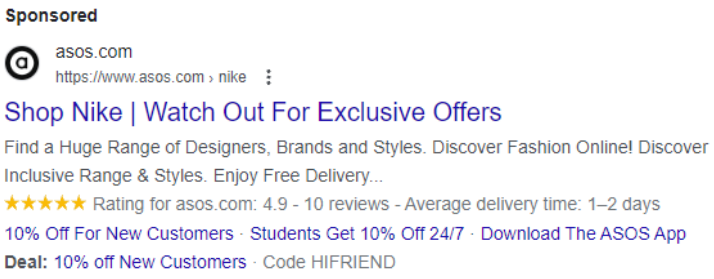

Рисунок 2.10 - Оголошення магазину для пошукового терміну "nike"

Другою за величиною причиною помічених доменів є пов'язане, але легітимне програмне забезпечення. Іноді пов'язане програмне забезпечення може цільово використовувати ключові слова, включені в список пошукових термінів інструменту. Це призводить до того, що фільтр оголошень вважає ці оголошення релевантними. Потім ці оголошення помічаються як підозрілі через різницю між лендінг-URL та відомим доменом. Приклад такого пов'язаного оголошення можна побачити на рисунку 2.11.

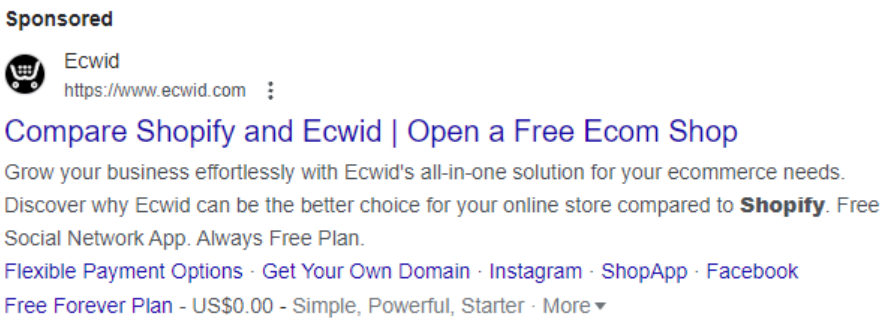

Рисунок 2.11 - Оголошення пов'язаного програмного забезпечення для ключового слова "shopify"

Окремий позначник даних використовується для подальшої категоризації помічених доменів та оголошень, присутніх у попередньому звіті. Цей позначник використовує базу даних веб-сайтів, де кожен веб-сайт є частиною категорії. Усі попередні звіти проходять через позначник, який аналізує результати кожного оголошення окремо. Якщо попередження лендінг-URL присутнє в попередньому звіті, тоді позначник порівнює цей URL зі своєю базою даних. Флаги призначаються оголошенню на основі категорії, до якої відноситься лендінг-URL. Якщо лендінг-URL недостатньо для винесення вердикту, тоді позначник також порівнює ідентифікатор рекламодавця зі своєю базою даних та позначає оголошення відповідно. Якщо автоматичний вердикт не може бути даний, наприклад, через те, що лендінг-URL і/або ідентифікатори рекламодавців не є категоризованими, тоді

користувачу повідомляється, що потрібна перевірка домену. Це повідомлення складається з повідомлення, яке вказує користувачу на оголошення (та його асоційований файл), яке потрібно перевірити.

База даних веб-сайтів, згадана вище, складається за допомогою позначника. Процес є циклічним, оскільки результати позначника використовуються для розширення бази даних, яка, у свою чергу, покращує продуктивність цього позначника. Під час позначення підраховується кількість кожного домену. Якщо це число більше або дорівнює значенню відсікання (обране для нашого інструменту як 3), надсилається сповіщення. Це сповіщення повідомляє користувача, що домен потребує категоризації. Користувач може вручну перевірити цей домен і розподілити його в одну з дев'яти категорій: магазини-аутлети, пов'язане програмне забезпечення, блоги/форуми, платформи курсів, кур'єрські служби, пошукові системи, непов'язані, платформи для завантаження додатків, що використовує імітацію бренду. Категорія "непов'язані" стосується оголошень, які були помічені через кілька інтерпретацій ключового слова. Наприклад, "UPS" може означати "United Parcel Service" (наша ціль) та "Uninterrupted Power Supply" (непов'язані). Подібним чином, "facebook" може повертати оголошення на контактні сторінки інших веб-сайтів (рисунок 2.12), а "booking" широко використовується платформами, схожими на booking.com, без наміру на заміну бренду (рисунок 2.13).

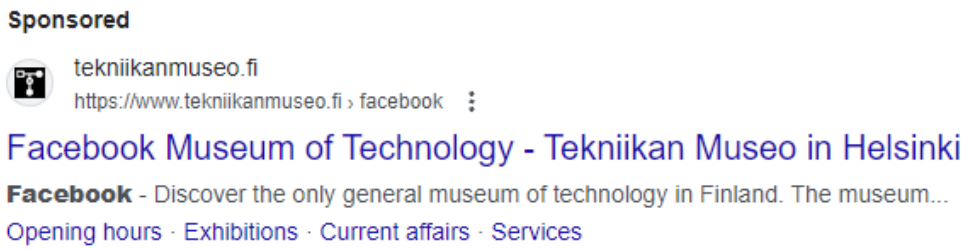

Рисунок 2.12 - Непов'язане оголошення, яке вказує на контактну сторінку Facebook

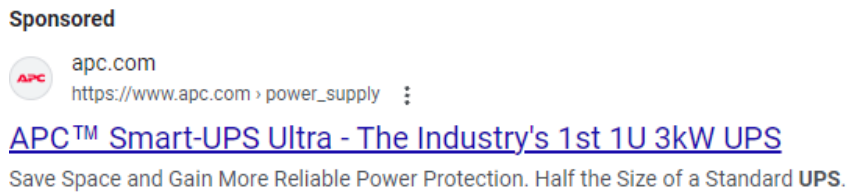

Рисунок 2.13 - Непов'язане оголошення для ключового слова "ups"

Ця система категоризує лендінг-URL кожного оголошення на основі раніше представленої бази даних. Відповідні флаги потім призначаються оголошенню, що дозволяє краще візуалізувати результати. Система флагів складається з наступного, що показано на рисунку 2.14.

F	**- нешкідливий**	**S**	**- шкідливий**
O	- магазин-аутлет/роздріб	**E**	- пошукова система
R	- пов'язане програмне забезпечення	**P**	- платформа для завантаження додатків
C	- платформи курсів	**M**	- брендджекінг
B	- блоги/форуми	**X**	- перевірено вручну
I	- кур'єрські служби		
U	- непов'язані		
L	- легітимний		

Рисунок 2.14 – Система флагів (міток)

Уривок проміжного звіту, повернутого позначником, можна побачити нижче. Оголошення, спочатку записані скрепером, тепер категоризовані.

Оголошення 4 було визнане легітимним, оскільки ідентифікатор рекламодавця асоційований з антивірусною компанією "McAfee".

Оголошення 7 було визнане нешкідливим, оскільки blitzhandel24 є відомим магазином для онлайн-продуктів.

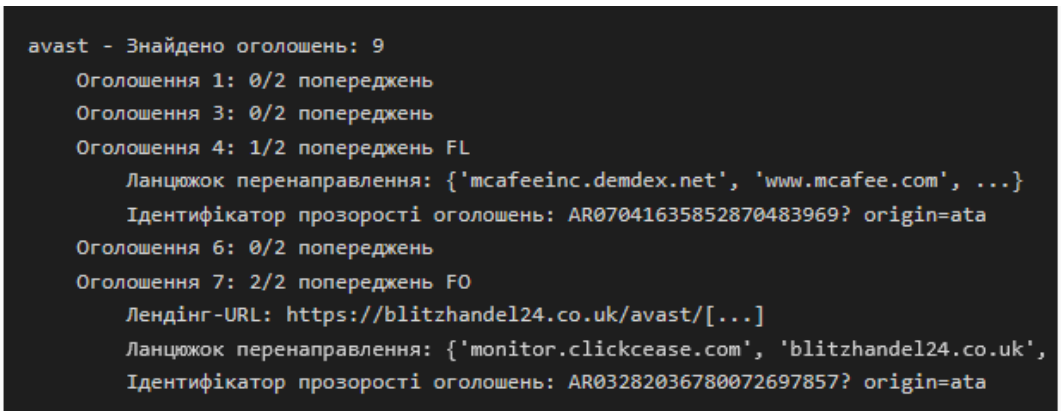

Рисунок 2.15 – Фрагмент проміжного звіту

2.4.2. Ефективність збору даних та автоматизована категоризація

Виконання процесу збору даних за допомогою розробленого скрепера вимагає значних часових витрат. Проведені вимірювання показали, що один повний цикл збору даних для окремої країни в середньому триває близько однієї години. Цей фактор суттєво обмежує можливості оперативного оновлення даних, необхідних для аналізу в режимі реального часу або для швидкої адаптації до змін у ландшафті зловмисної рекламної діяльності. Зокрема, для збору нової вибірки результатів, що відображає актуальний стан, потрібно до десяти годин роботи скрепера на кожну країну.

Зважаючи на це обмеження, компонент категоризації даних (data labeler) був розроблений як незалежний програмний модуль, функціонально відокремлений від модуля скрепінгу. Таке архітектурне рішення дозволяє здійснювати ефективну категоризацію нових, щойно зібраних наборів даних, а також проводити повторну категоризацію раніше зібраних результатів з мінімальними зусиллями та без необхідності перезапускати тривалий процес скрепінгу. Це є особливо критично важливим у сценаріях, коли статус раніше визнаних нешкідливими доменів змінюється на потенційно небезпечний, і потрібно оперативно переоцінити відповідні оголошення.

Як завершальний етап фази аналізу даних, реалізовано процес автоматизованої перевірки доменів. Цей процес має на меті верифікувати та підвищити точність результатів, отриманих за допомогою початкових

автоматизованих методів виявлення. Необхідність у перевірці доменів виникає для доменів, що часто зустрічаються у зібраних оголошеннях, або для оголошень, які не були однозначно категоризовані попередніми правилами. В обох випадках логіка перевірки ґрунтується на розрахунку інтегрального "рівня загрози" для досліджуваного домену.

Для оцінки рівня загрози, ефективний домен верхнього рівня (eTLD) та, якщо це можливо, безпосередньо наступний рівень домену (eTLD+1) піддаються аналізу за допомогою трьох різних зовнішніх платформ перевірки репутації та наявності шкідливої активності: VirusTotal, URLVoid та IPQS. "Рівень загрози" для домену інкрементується на 1 бал за кожну з цих платформ, яка класифікує домен як шкідливий або підозрілий. Додатково, якщо вік домену (що можна отримати, зокрема, через сервіси на кшталт VirusTotal) становить менше одного року, до "рівня загрози" додається 0.5 бали, оскільки молоді домени часто використовуються в короткострокових зловмисних кампаніях.

Фінальна перевірка здійснюється на ресурсі scammer.info, що є відкритою платформою для звітів про онлайн-шахрайство. Якщо досліджуваний домен або пов'язаний з оголошенням номер телефону фігурує у повідомленнях про шахрайство на цьому веб-сайті, "рівень загрози" також збільшується на 1 бал. Інтеграція інформації з чотирьох незалежних джерел суттєво підвищує ймовірність ідентифікації зловмисних оголошень та знижує ризик хибнонегативних спрацьовувань.

У даній роботі в якості "порогового значення" для рівня загрози, яке використовується для остаточної класифікації, обрано значення 1. Слід зазначити, що це порогове значення може бути змінене дослідниками для порівняння результатів та оптимізації балансу між точністю та повнотою виявлення. Результати, представлені в даній роботі, базуються на застосуванні вищезазначеного числового критерію. Якщо розрахований фінальний "рівень загрози" для домену рекламного оголошення перевищує

встановлене порогове значення, таке оголошення класифікується як зловмисне. В іншому випадку воно відноситься до категорії нешкідливих оголошень.

Висновки до другого розділу

У другому розділі було представлено розробку цілісної методології, спрямованої на запобігання поширенню шкідливого програмного забезпечення через онлайн-рекламу. Запропонований підхід базується на поєднанні сучасних інструментів автоматизації, зокрема фреймворку Playwright, та методів обробки великих обсягів даних.

У підрозділі 2.1 описано загальну концепцію методології, що охоплює виявлення підозрілих рекламних оголошень через автоматизований моніторинг пошукових систем та подальшу обробку зібраної інформації. В межах підрозділу 2.2 проведено проектування інструменту для реалізації цієї методології, зокрема детально розглянуто можливості Playwright як ефективного інструменту для взаємодії з вебінтерфейсами в умовах реального часу.

Підрозділ 2.2.2 присвячено налаштуванню конфігурації системи та здійсненню цільових пошукових запитів, що дозволяє автоматизовано виявляти потенційно шкідливу рекламу.

У підрозділі 2.3 описано процес збору даних, включно з критеріями вибору релевантних джерел і способами збереження інформації для подальшого аналізу.

У підрозділі 2.4 розглянуто етап аналізу та оцінки рекламних оголошень, з особливим акцентом на ручне та автоматизоване маркування, а також на показники ефективності збору даних і можливості їхньої класифікації за допомогою автоматизованих засобів.

У результаті, розроблена методологія демонструє високий потенціал у сфері кібербезпеки, зокрема у виявленні та протидії шкідливим елементам онлайн-реклами шляхом гнучкої автоматизації, структурованого збору даних і подальшої аналітики. Запропоноване рішення є масштабованим і може бути адаптоване до різних сценаріїв використання в реальному середовищі.

РОЗДІЛ 3. ОЦІНКА РЕЗУЛЬТАТІВ ІМПЛЕМЕНТАЦІЇ МЕТОДОЛОГІЇ ЗАПОБІГАННЯ ПОШИРЕННЯ ШКІДЛИВОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ НА РІВНІ МАРКЕТИНГОВИХ СТРАТЕГІЙ

3.1. Представлення результатів збору даних

Процес збору даних здійснювався безперервно протягом періоду розробки скрепера, що охоплював проміжок з лютого 2025 року по травень 2025 року. Результати, отримані в ході цих початкових (пілотних) скрепінгів, були використані для ітеративного вдосконалення інструменту та розширення бази відомих легітимних доменів. Для даного дослідження було проаналізовано результати, зібрані скрепером протягом червня 2025 року, що включає дані 64 окремих запусків.

Загалом у ході 3200 запитів (що відповідає 64 запускам по 50 запитів кожен) скрепером було зафіксовано сукупність 13 967 рекламних оголошень. Після застосування фільтраційного механізму, 10 067 з цих оголошень було визнано релевантними.

Таблиця 3.1 - Результати скрепера за країною

країна	запуски	загальні оголошення	релевантні	легітимні	нешкідливі	шкідливі	невизначені
us	8	1,676	1,325	739	451	111	24
gb	6	1,807	1,140	539	491	45	65
au	7	1,203	826	440	298	39	49
ca	7	1,852	1,269	680	428	124	37
nl	7	1,638	1,294	655	583	50	6
ro	7	1,685	1,139	494	606	31	8
se	7	1,644	1,270	603	609	42	16
za	5	765	558	288	227	17	26
br	5	1,129	819	383	391	29	16
th	5	568	427	237	132	25	33
Всього	**64**	**13,967**	**10,067**	**5,058**	**4,216**	**513**	**280**

У таблиці 3.1 представлено розподіл результатів категоризації для кожної країни. Стовпець "Легітимні" відображає кількість оголошень, визначених як легітимні. До цієї категорії віднесені оголошення, які не ініціювали жодних попереджень у процесі скрепінгу, тобто їхній перехід здійснювався на відомий домен без виявлення підозрілих перенаправлень. Стовпець "Нешкідливі" показує кількість оголошень, які були ідентифіковані як нешкідливі (безпечні) зовнішніми інструментами виявлення загроз, як описано в попередньому розділі. Кількість випадків, які наш інструмент класифікував як зловмисні, представлено у стовпці "Шкідливі". Останній стовпець, "Невизначені", містить оголошення з невизначеними результатами категоризації. Така ситуація виникала, коли скрепер стикався з непередбачуваними обставинами у ході виконання, наприклад, поява спливаючих вікон, що вимагали взаємодії користувача, для якої скрепер не був спроектований через низьку відтворюваність таких сценаріїв.

Таблиця 3.2 - Відношення результатів за країною

країна	**легітимні%**	**нешкідливі%**	**шкідливі%**	**невизначені%**
us	55,77	34,04	8,38	1,81
gb	47,28	43,07	3,95	5,70
au	53,27	36,08	4,72	5,93
ca	53,59	33,73	9,77	2,92
nl	50,62	45,05	3,86	0,46
ro	43,37	53,20	2,72	0,70
se	47,48	47,95	3,31	1,26
za	51,61	40,68	3,05	4,66
br	46,76	47,74	3,54	1,95
th	55,50	30,91	5,85	7,73
Всього	**50,24**	**41,88**	**5,10**	**2,78**

Відсоткові показники розподілу оголошень по категоріях для кожної країни більш наочно візуалізовано у таблиці 3.2. Ці відсотки були розраховані на основі загальної кількості релевантних оголошень, зібраних

скрепером. Приблизно половина (близько 50%) релевантних оголошень були визнані інструментом легітимними. Нешкідливі оголошення становили приблизно 42% від загальної кількості релевантних оголошень. З 10 067 записаних релевантних оголошень, 5% були позначені як шкідливі. Рівень невизначених результатів становив менше 3%.

3.2. Представлення розподілу шкідливих та нешкідливих результатів. Статистика відвідуваності доменів

У таблиці 3.3 представлена кількість ідентифікованих зловмисних доменів у розрізі країн, а також кількість рекламних оголошень, асоційованих із цими доменами. Загалом, у ході всіх запусків дослідження було визначено 49 унікальних зловмисних доменів. Аналіз їхнього характеру показав наступний розподіл: 22 домени демонстрували ознаки імітації фірмової ідентичності різного ступеня вираженості, 20 доменів функціонували як зловмисні пошукові системи, а 7 були платформами, що використовувалися для поширення шкідливого програмного забезпечення під виглядом легітимних додатків.

Таблиця 3.3 - Розподіл шкідливих результатів за країною

країна	шкідливі домени	шкідливі оголошення	% від релевантних оголошень
us	24	111	8,38
gb	9	45	3,95
au	5	39	4,72
ca	24	124	9,77
nl	6	50	3,86
ro	2	31	2,72
se	8	42	3,31
za	7	17	3,05
br	8	29	3,54
th	5	25	5,85
Всього	49	513	5,10

П'ятьма пошуковими термінами, які найчастіше асоціювалися зі зловмисними оголошеннями, були, у порядку спадання частоти, "mcafee", "avast", "microsoft", "anydesk" та "teamviewer".

У таблиці 3.4 представлена категоризація рекламних оголошень, які були позначені системою як нешкідливі (безпечні). Кожний запис у таблиці відповідає одній з категорій, визначених у другому розділі.

Таблиця 3.4 - Розподіл нешкідливих випадків

категорія	домени	оголошення	%
аутлет	212	1,855	44,00
пов'язане	83	987	23,41
курси	10	222	5,27
блоги	26	320	7,59
кур'єр	14	136	3,23
платформи для додатків	8	88	2,09
пошукові системи	21	194	4,60
непов'язані	41	414	9,82

Згідно з отриманими даними, 44% зібраних нешкідливих оголошень належать до категорії інтернет-магазинів (аутлетів), що пропонують широкий асортимент товарів. Оголошення, пов'язані з програмним забезпеченням, складають приблизно 24%. До цієї категорії віднесені оголошення легітимних програмних продуктів, які таргетуються за ключовими словами, включеними до списку пошукових термінів дослідження. Приблизно 10% оголошень (відображено у стовпці "Непов'язані") були класифіковані як релевантні виключно через наявність пошукового терміну.

Однак, через полісемію або схожість формулювань пошукових термінів (як у випадку "United Parcel Service" проти "Uninterrupted Power Supply"), ці оголошення виявилися концептуально непов'язаними з очікуваною тематикою. Решта 22.77% нешкідливих випадків включали різноманітні платформи, що пропонували освітні курси, блоги/форуми, послуги

кур'єрської доставки, платформи для завантаження програмного забезпечення або пошукові сервіси.

Таблиця 3.5 – Статистика відвідуваності доменів

кількість відвідин	домени	оголошення
1	293	293
≤ 2	425	557
≤ 3	511	815
≤ 5	641	1,383
≤ 10	769	2,351

На основі зібраного набору даних було скомпільовано перелік усіх відвіданих доменів із зазначенням частоти їх відвідуваності. Ці дані були використані для визначення порогового значення частоти відвідувань, при перевищенні якого домен підлягає додатковій верифікації. Загалом, скрепер здійснив доступ до 854 унікальних доменів, що асоціювалися з 4729 рекламними оголошеннями. 11 доменів з цього переліку були відвідані понад 50 разів, що визначило їх як найбільш часто зустрічаються у вибірці.

3.3. Поглиблений аналіз отриманих результатів на основі показників

У цьому і наступних підрозділах буде проведено поглиблений аналіз результатів Деякі виявлені приклади зловмисної рекламної діяльності будуть деталізовані разом з підтверджуючими даними. Також будуть обговорені обмеження розробленого скрепера та окреслені можливі шляхи його подальшого вдосконалення.

3.3.1. Географічний розподіл загроз

Найвищі показники поширеності зловмисних оголошень були зафіксовані на території Сполучених Штатів Америки та Канади. Цю

обставину можна пояснити кількома взаємопов'язаними факторами. По-перше, сукупне населення цих країн має значну кількість активних користувачів інтернету та високий рівень онлайн-економічної активності, що робить їх привабливими мішенями для суб'єктів, що здійснюють зловмисну рекламну діяльність. По-друге, слід відзначити більш високий рівень розвитку рекламних екосистем у США та Канаді [34]. Це дозволяє здійснювати більш точне таргетування рекламних оголошень на основі детального аналізу поведінки користувачів та їхніх демографічних характеристик. Така точність у націлюванні може підвищити ефективність зловмисних оголошень, оскільки вони можуть бути адаптовані для створення більш переконливого вигляду для конкретних сегментів користувачів або спрямовані на більш вразливі групи населення.

3.3.2. Цільова аудиторія та вектори атаки

Значна поширеність зловмисних оголошень, спрямованих на користувачів антивірусного програмного забезпечення (зокрема, McAfee, Avast), продуктів Microsoft та програм для віддаленого доступу (AnyDesk, Teamviewer), ймовірно, мотивована демографічними характеристиками користувацької бази цих програмних продуктів. Таке програмне забезпечення часто використовується старшим поколінням користувачів, які можуть бути більш сприйнятливими до фішингових атак та менш обізнаними щодо актуальних методів кібербезпеки. Ця обставина робить їх привабливою цільовою аудиторією для тих, хто прагне експлуатувати широку та потенційно вразливу групу користувачів. Як зазначалося раніше, суб'єкти, що використовують імітацію фірмової ідентичності, експлуатують відомі та довірені бренди для підвищення довіри до своїх фішингових спроб. Видаючи себе за ці бренди, вони можуть успішно вводити користувачів в оману та спонукати їх до завантаження зловмисного програмного забезпечення або

розголошення конфіденційної інформації, що робить згадані категорії програмного забезпечення основними мішенями для таких атак.

3.3.3. Аналіз віку доменів

Одним із вартих уваги спостережень, отриманих з аналізу результатів, є кореляція між віком домену та ймовірністю його належності до категорії зловмисних. Переважна більшість виявлених зловмисних доменів мали вік менше одного року, що відповідає відомим тактикам, які застосовуються кіберзлочинцями для уникнення виявлення шляхом частої реєстрації нових доменів. Однак, були зафіксовані й винятки, коли деякі зловмисні домени існували більше одного року, а окремі навіть понад 5 років. Ці знахідки узгоджуються з результатами дослідження [16]. Наявність старіших зловмисних доменів може свідчити про більш досвідчених кібератакувальників, які здатні тривалий час підтримувати контроль над доменом без його компрометації. Цікаво також відзначити, що веб-сайти, які існували більше 5 років, були класифіковані як легітимні такими автоматизованими сервісами, як VirusTotal, URLVoid та IPQS. Єдині повідомлення про їхню зловмисність надходили від користувачів через платформу scammer.info. Стійкість таких доменів підкреслює нагальну потребу у постійному моніторингу та адаптації методів виявлення для ефективної протидії як новим, так і тривалим загрозам.

3.3.4.Обґрунтування вибору дизайну системи

В другому розділі було згадано про необхідність досягнення "оптимального балансу" між обсягом автоматизованої перевірки доменів та обсягом перевірки окремих рекламних оголошень. Цей баланс визначає порогове значення частоти відвідувань домену для його включення до процесу верифікації. Результати, представлені в попередньому розділі, будуть використані для обґрунтування вибору порогового значення у 3

відвідування. Компонент категоризації даних був розроблений з метою мінімізації потреби у ручній експертній перевірці. Чим вища частота відвідувань певного домену, тим більша кількість рекламних оголошень, пов'язаних з цим доменом, може бути автоматично покрита результатами однієї верифікації домену. І навпаки, нижча частота домену означає менше охоплення оголошень однією перевіркою.

Розглядалися наступні варіанти стратегії перевірки:

1. Перевірка всіх (854) унікальних доменів ⇒ автоматична категоризація всіх асоційованих оголошень.

2. Перевірка доменів, відвіданих щонайменше 2 рази (561 доменів) ⇒ необхідність додаткової перевірки оголошень, пов'язаних з рештою 293 доменів (що відвідані 1 раз).

3. Перевірка доменів, відвіданих щонайменше 3 рази (429 доменів) ⇒ необхідність додаткової перевірки оголошень, пов'язаних з рештою 557 доменів (відвідані 1 або 2 рази).

4. Перевірка доменів, відвіданих щонайменше 5 разів (261 домен) ⇒ необхідність додаткової перевірки оголошень, пов'язаних з рештою 1143 доменів (відвідані 1, 2, 3 або 4 рази).

Варіанти 1 і 2 вимагали б теоретично однакової сукупної кількості перевірок (домени + оголошення, що потребують окремої перевірки) — 854 одиниці перевірки. Варіант 3 вимагав би 429 перевірок доменів та 557 перевірок оголошень, сумарно 986 одиниць перевірки. Варіант 4 вимагав би 261 перевірку доменів та 1143 перевірки оголошень, сумарно 1404 одиниці перевірки. Збільшення порогового значення частоти відвідувань призводило б до експоненційного зростання кількості окремих оголошень, що потребують перевірки.

Незважаючи на те, що варіант 3 (поріг 3 відвідування) вимагає дещо більшої сукупної кількості перевірок (986) порівняно з варіантами 1 і 2 (854), саме він був обраний. Це рішення обґрунтовано тим, що, з огляду на

структуру зібраної інформації, перевірка окремих рекламних оголошень є більш ефективною та швидшою, оскільки вся необхідна інформація (скриншоти, відео, HAR-файли) вже доступна в папці, асоційованій з оголошенням. На противагу цьому, верифікація домену безпосередньо з подальшим пошуком відповідних рекламних матеріалів вимагає додаткових зусиль. Таким чином, розподіл зусиль у варіанті 3 (менше перевірок доменів, більше – оголошень) призвів до нижчого ефективного рівня необхідних ручних зусиль, що стало вирішальним фактором при виборі порогового значення 3.

3.4. Деталізація виявлених зловмисних випадків

3.4.1. Аналіз ідентифікованих шкідливих доменів

Із загальної кількості 854 відвіданих унікальних доменів, 49 були класифіковані як зловмисні на підставі інтегрального "рівня загрози", розрахованого автоматизованими методами. Виявлені зловмисні домени були розподілені за трьома основними категоріями: випадки імітації фірмової ідентичності, зловмисні пошукові системи та платформи для завантаження додатків. Детальний аналіз кожної з цих категорій представлено у наступних підрозділах.

3.4.2. Випадки імітації фірмової ідентичності

З сукупності зловмисних доменів, ідентифікованих у наборі даних, 22 домени були класифіковані як такі, що використовують імітацію фірмової ідентичності. Переважна більшість доменів у цій категорії (17 з 22) відповідають шаблонам так званих "технічних шахрайств" (tech scams). Характерними ознаками таких веб-сайтів є, зокрема, безпідставні твердження про інфікування комп'ютера користувача шкідливим програмним забезпеченням, використання логотипів відомих брендів без належного

дозволу для підвищення рівня довіри та відображення контактних телефонних номерів під виглядом служб клієнтської підтримки. Рекламні оголошення, що ведуть на такі ресурси, переважно націлені на пошукові терміни, пов'язані з антивірусним програмним забезпеченням та продуктами Microsoft ("mcafee", "avast", "microsoft"). На цих веб-сайтах, як правило, пропонуються до придбання нібито ліцензії на зазначене програмне забезпечення або послуги з надання технічної підтримки, що насправді є спробою шахрайства.

Наступні приклади ілюструють типові випадки зловмисної рекламної діяльності, що базується на імітації фірмової ідентичності, виявлені у ході дослідження.

Приклад 1 - windowstechies.com

Цей веб-сайт пропонує інструмент, призначений для "вирішення поширених проблем ПК". Ресурс активно цільово використовує значну кількість ключових слів. Пошукові терміни асоціюються з окремими рекламними оголошеннями (рисунок 3.1), які ведуть на різні версії цільової сторінки веб-сайту (рисунок 3.2), що може свідчити про адаптацію контенту під конкретний запит. Домен windowstechies.com мав значний вік – 12 років на момент збору даних, і був позначений як легітимний трьома автоматизованими інструментами виявлення (VirusTotal, URLVoid, IPQS). Повідомлення про його зловмисний характер надходять виключно від користувачів на платформі scammer.info.

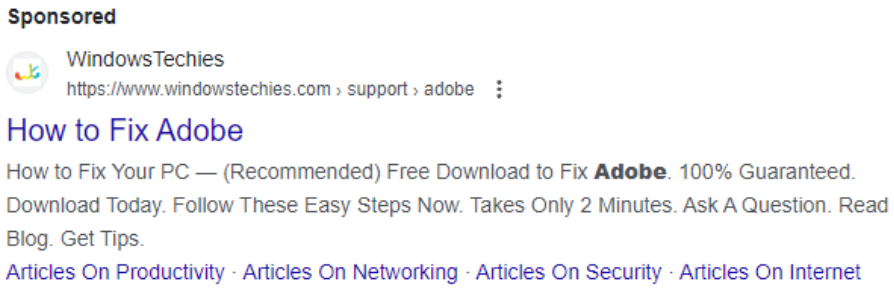

Рисунок 3.1 - Приклад рекламного оголошення windowstechies.com, спрямованого на ключове слово "adobe".

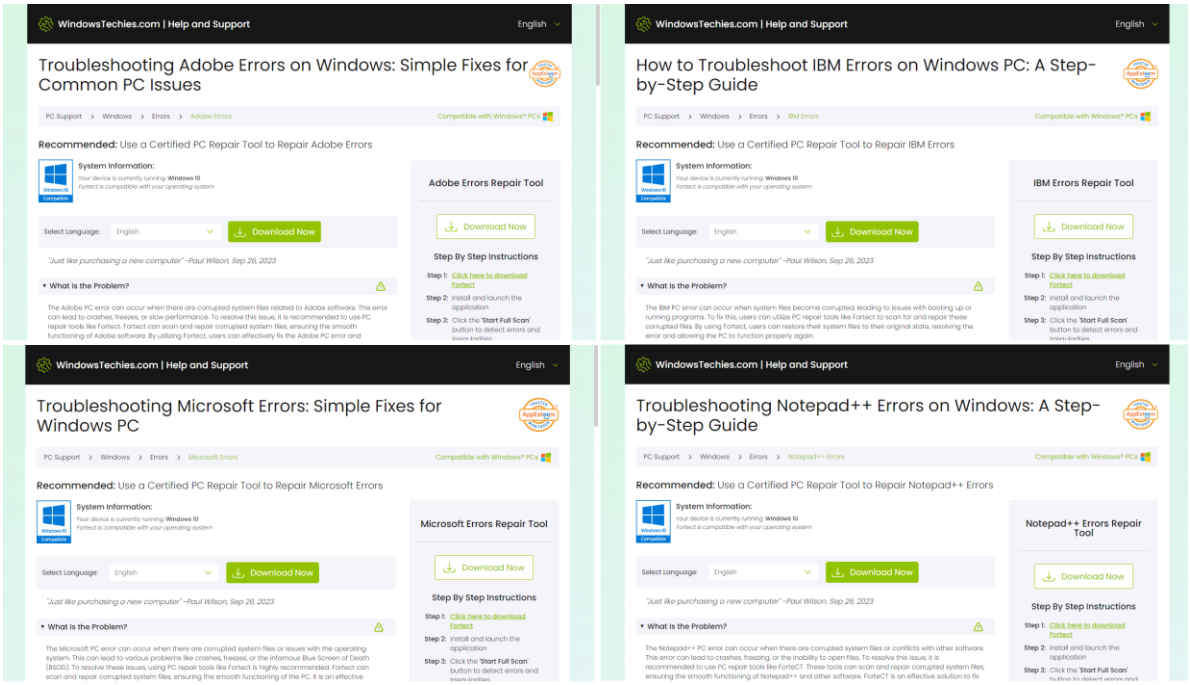

Рисунок 3.2 - Варіанти цільових сторінок веб-сайту windowstechies.com

Приклад 2 - deal.risecenter.shop

Цей веб-сайт рекламує придбання ліцензій на антивірусне програмне забезпечення McAfee. Він відповідає характерним шаблонам "технічних шахрайств", згаданих раніше. Домен deal.risecenter.shop позначений як зловмисний на платформах VirusTotal та IPQS, а також фігурує у повідомленнях про шахрайські дзвінки на scammer.info. Використання бренду McAfee без відомої легітимної афіліації є яскравим випадком імітації фірмової ідентичності. Дуже схожий випадок було зафіксовано для веб-сайту deal.websitecentral.shop, який має практично ідентичний макет сторінки. Відмінності полягають лише у відображеному контактному номері телефону та логотипі на веб-сайті.

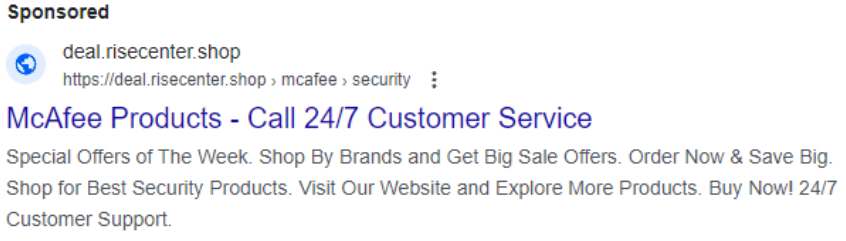

Рисунок 3.3 - Рекламне оголошення deal.risecenter.com для слова "mcafee"

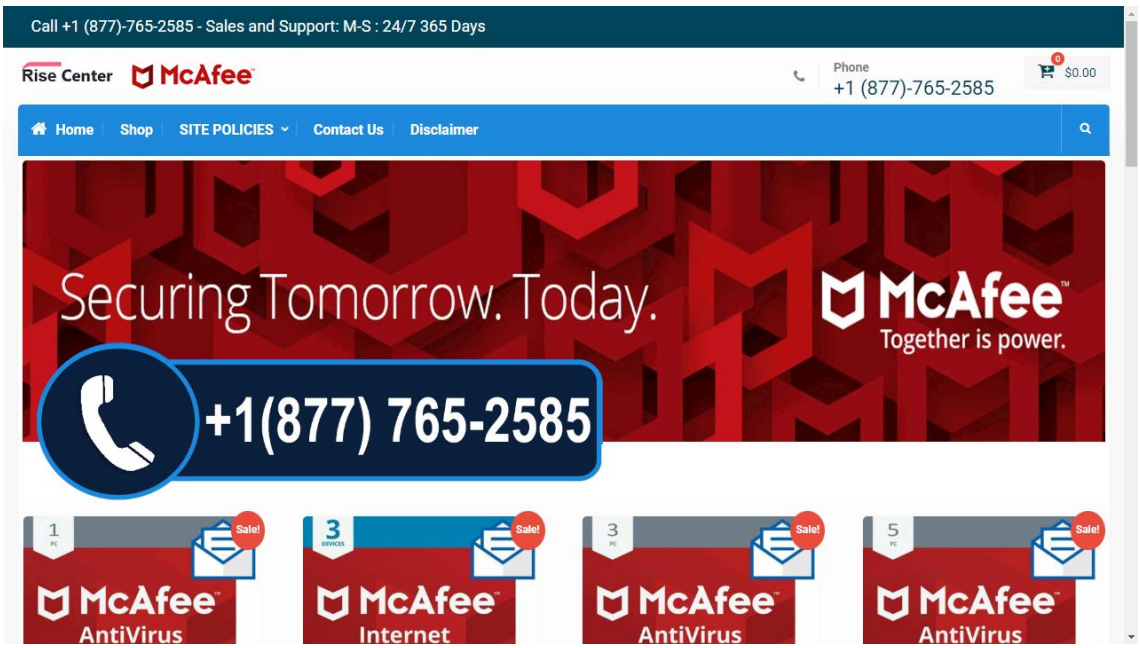

Рисунок 3.4 - Цільова сторінка веб-сайту deal.risecenter.com.

Приклад 3 - aolsolution.info

Цей веб-сайт також класифікується як "технічне шахрайство", пропонуючи цього разу послуги з "клієнтської підтримки". Ресурс позначений як зловмисний на VirusTotal та також згадується користувачами на платформі scammer.info. Веб-сайт вдає, що надає підтримку для антивірусного програмного забезпечення Avast, використовуючи імітацію бренду.

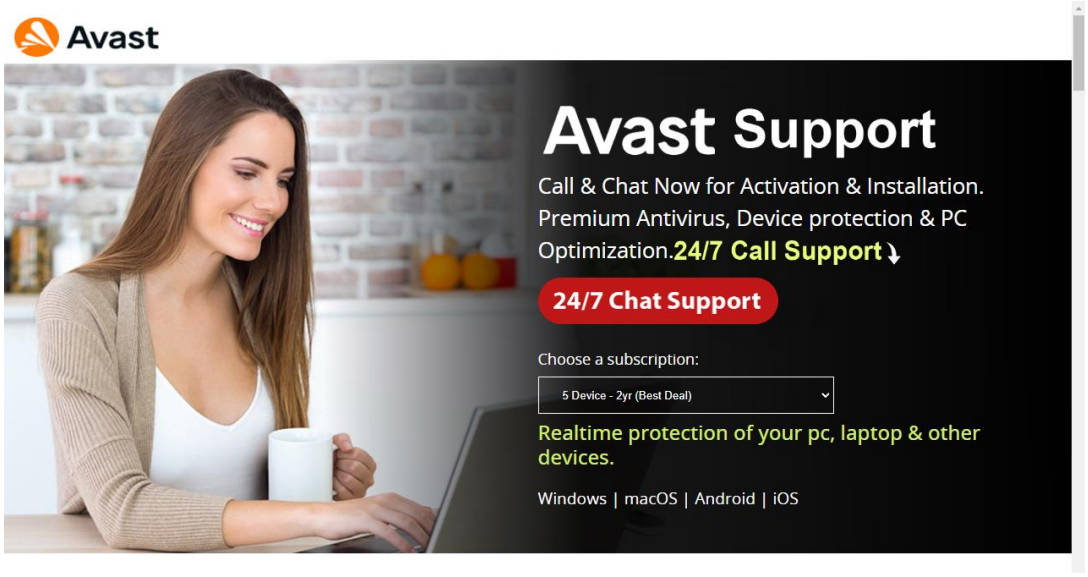

Рисунок 3.5 - Цільова сторінка веб-сайту aolsolution.info

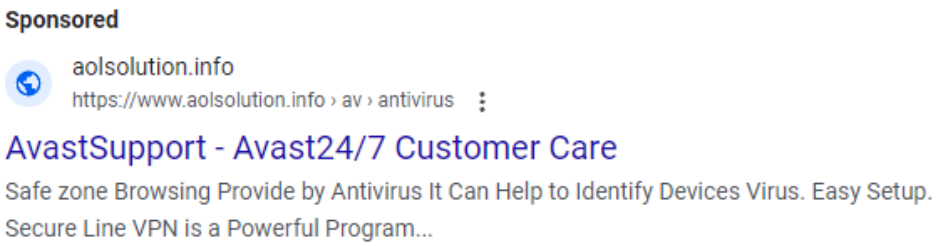

Рисунок 3.6 - Рекламне оголошення aolsolution.info для ключового слова "avast"

Приклад 4 - office-staples.org

Цей веб-сайт пропонує до придбання ліцензії на "Microsoft Windows". Проте, він не повністю відповідає раніше описаним шаблонам "технічних скамів".

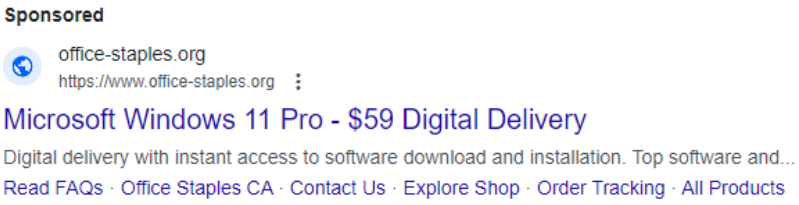

Рисунок 3.7 - Рекламне оголошення office-staples.org для ключового слова "microsoft"

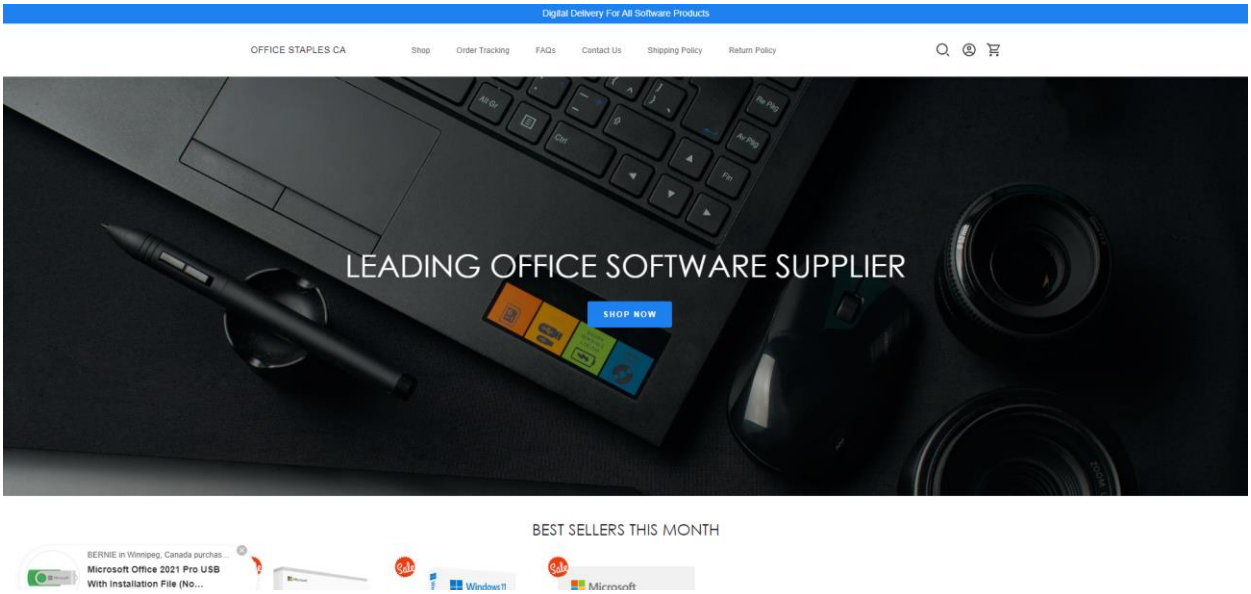

Рисунок 3.8 - Цільова сторінка веб-сайту office-staples.org

Цільова сторінка не відображає контактного телефону та не заявляє про безпосередню афіліацію з корпорацією Microsoft. Незважаючи на це, домен був позначений як зловмисний платформами VirusTotal та IPQS.

Приклад 5 - auth.onuberconnect.com

Цей веб-сайт функціонує як перенаправлення з інших проаналізованих рекламних посилань, зокрема eats-uder.online та couponcave.online. Обидва згадані веб-сайти також доступні для ручного відвідування, але при цьому вони ведуть на іншу цільову сторінку, відмінну від тієї, що була зафіксована скрепером (Рисунок 3.10). При спробі ручного доступу до URL-адреси, на яку потрапив скрепер, відбулося перенаправлення на головну сторінку "Yahoo!", що може слугувати ознакою застосування тактики маскування реклами (cloaking). Веб-сайт auth.onuberconnect.com позначений як зловмисний лише платформою IPQS і розроблений для збору облікових даних користувачів, вдаючи пропозицію купону (як видно з рисунку 3.10, який ілюструє спробу імітації фірмової ідентичності Uber). Через неможливість прямого ручного доступу до захопленої URL-адреси, провести додаткове дослідження намірів розробників сторінки щодо подальшої обробки зібраних даних не вдалося.

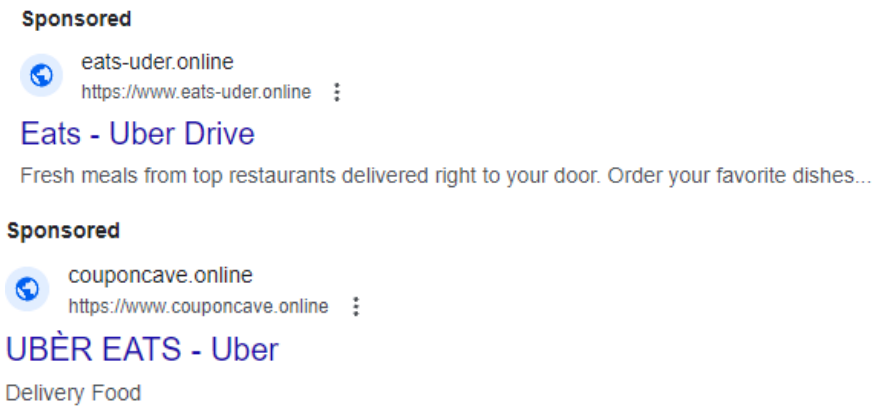

Рисунок 3.9 - Два різних рекламних оголошення, що перенаправляють на auth.onuberconnect.com для ключового слова "uber".

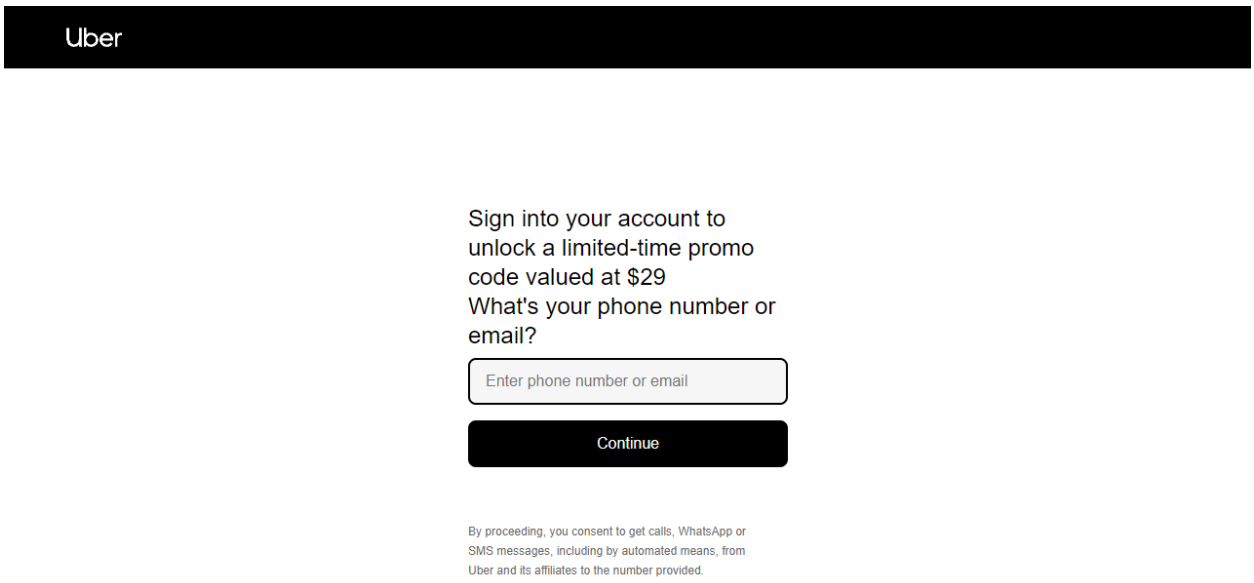

Рисунок 3.10 - Цільова сторінка веб-сайту auth.onuberconnect.com, захоплена скрепером

3.4.3. Шкідливі пошукові системи

У процесі скрепінгу рекламних оголошень було ідентифіковано загалом 41 веб-сайт, що функціонував як пошукова система. Приблизно половина з них була класифікована як зловмисні згідно з оцінками платформ VirusTotal або URLVoid.

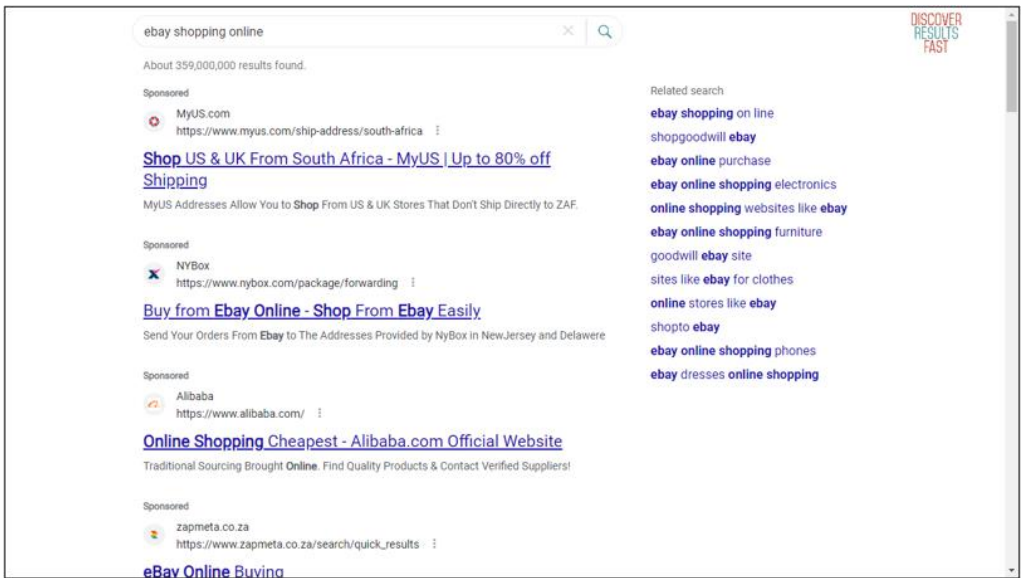

Рисунок 3.11 - Цільова сторінка discoverresultsfast.com, що є прикладом зловмисної пошукової системи

Такі зловмисні пошукові системи становлять загрозу тим, що вони можуть маніпулювати результатами пошуку, штучно просуваючи фішингові ресурси, контент, пов'язаний з рекламним шахрайством, та інший оманливий вміст.

На рисунку 3.11 проілюстровано цільову сторінку типової зловмисної пошукової системи. Відповідне рекламне оголошення було націлене на ключовий термін "ebay". На представленому знімку екрана видно, що легітимна сторінка веб-сайту "ebay" не відображається серед перших чотирьох результатів пошуку, що є індикатором маніпуляції.

3.4.4. Шкідливі платформи для завантаження додатків

Серед виявлених у ході скрепінгу зловмисних оголошень також були ідентифіковані платформи, що рекламували завантаження програмного забезпечення. З 15 доменів, що асоціювалися з такими платформами, 7 були класифіковані як зловмисні за результатами перевірки на VirusTotal, URLVoid або IPQS. Згідно з наявною інформацією, ці платформи часто асоціюються з розповсюдженням небажаного програмного забезпечення (adware).

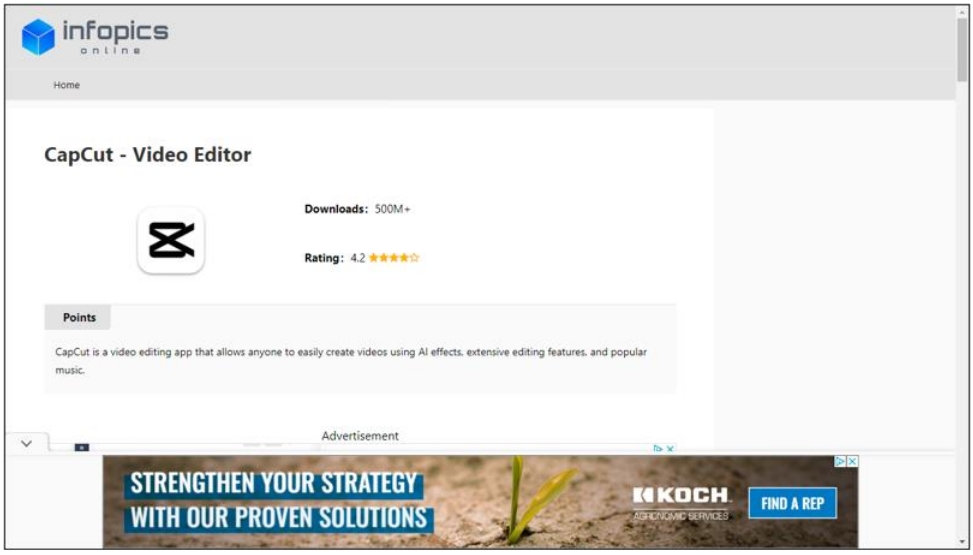

Рисунок 3.12 - Цільова сторінка infopics.online, що є прикладом зловмисної платформи для завантаження додатків

На рисунку 3.12 представлено приклад цільової сторінки такої платформи. Відповідне рекламне оголошення було націлене на ключовий термін "capcut", однак скрепером також були зафіксовані оголошення, що ведуть на цю платформу, за іншими пошуковими термінами, такими як "facebook".

3.5. Ефективність використаних методів виявлення

Як було показано у попередніх підрозділах, для оцінки потенційної зловмисності доменів використовувався комплекс із чотирьох різних платформ. Узагальнення ефективності використаних інструментів виявлення представлено нижче у таблиці 3.6.

Таблиця 3.6 - Узагальнені результати ефективності використаних інструментів виявлення

Платформа	**Кількість позначених**	**Імітація бренду**	**Пошукові системи**	**Платформи для додатків**
VirusTotal	37	11	20	6
URLVoid	23	4	15	4
IPQS	10	8	0	2
scammer.info	14	14	0	0

Платформа VirusTotal продемонструвала найвищий рівень загального виявлення зловмисних випадків, включаючи значну ефективність у ідентифікації зловмисних пошукових систем. Це, ймовірно, пояснюється агрегацією результатів з багатьох антивірусних рушіїв та інструментів сканування URL, що робить VirusTotal надійним ресурсом для широкого спектра загроз.

Платформа URLVoid показала хороші результати у виявленні зловмисних випадків, пов'язаних із пошуковими системами. Однак, її ефективність була нижчою у виявленні випадків імітації фірмової ідентичності та зловмисних платформ для завантаження додатків. Це свідчить про те, що сильні сторони URLVoid полягають насамперед в аналізі URL-адрес та використанні численних баз даних "чорних списків".

IPQS виявився ефективним у виявленні випадків імітації фірмової ідентичності, але не позначив жодного випадку, пов'язаного зі зловмисними пошуковими системами. Це свідчить про те, що IPQS більш спеціалізований на превенції шахрайства та оцінці репутації IP-адрес, зокрема в контексті захисту брендів. Ймовірно, ця платформа не має таких самих широких можливостей сканування URL, як попередні два інструменти.

Платформа scammer.info повністю сфокусована на зловмисній діяльності, пов'язаній з брендами та шахрайством, що відображає її краудсорсинговий підхід до збору інформації про шахрайські дії. Відсутність виявлення у інших категоріях підтверджує вузьку, але надзвичайно ефективну спеціалізацію на загрозах, що експлуатують довіру до брендів. Цей ресурс також виявився ефективним у позначенні доменів віком понад один рік, які могли залишитися непоміченими автоматизованими інструментами.

Оцінка ефективності показала, що кожен з використаних інструментів виявлення має свої сильні та слабкі сторони. Однак, значна непослідовність у результатах між інструментами свідчить про те, що жоден з них окремо не є надзвичайно ефективним у комплексному виявленні зловмисної рекламної діяльності, що базується на імітації фірмової ідентичності. Це підкреслює необхідність інтеграції даних з множинних джерел для підвищення загальної ефективності виявлення.

Одним з ключових обмежень даного дослідження є те, що його сфера була сфокусована виключно на рекламних оголошеннях, що

демонструвалися в рамках однієї рекламної мережі, а саме Google. Хоча Google посідає домінуюче становище в екосистемі онлайн-реклами, концентрація виключно на цій мережі може призвести до того, що отримані результати не будуть повністю репрезентативними для всього ландшафту онлайн-реклами. Такий вузький погляд ігнорує потенційні випадки зловмисної рекламної діяльності, присутні на інших значних рекламних платформах, таких як Facebook або Bing. Ці платформи мають відмінну від Google архітектуру відображення сторінок та оголошень, що може вимагати суттєвої адаптації або призвести до неефективної роботи розробленого скрепера.

Крім того, список пошукових термінів, використаний у даному дослідженні, був обмежений набором з 50 популярних програмних продуктів та веб-сайтів. Хоча ці терміни були обрані на основі їхньої високої релевантності та значного обсягу трафіку, вони не охоплюють увесь можливий спектр цілей для зловмисної рекламної діяльності. Відповідно, значна кількість зловмисних кампаній, націлених на менш поширені або новостворені пошукові терміни, могла залишитися невиявленою. Таким чином, реальна поширеність зловмисної рекламної діяльності, що базується на імітації фірмової ідентичності, у загальному рекламному просторі може бути вищою за показники, отримані в даному дослідженні.

Іншим значущим обмеженням, яке вже згадувалося раніше, є застосування техніки маскування реклами (cloaking) зловмисниками. Хоча у ході дослідження було зафіксовано випадок, коли веб-скреперу вдалося успішно подолати механізми маскування та потрапити на зловмисну цільову сторінку (приклад 5), це не гарантує, що скрепер буде ефективним проти всіх варіацій цієї техніки. Цілком імовірно, що значна кількість інших зловмисних оголошень, які використовують складніші методи маскування, уникнули виявлення, що призвело до недооцінки справжньої поширеності зловмисної рекламної діяльності, що базується на імітації фірмової ідентичності.

Незважаючи на відносно низьку частку випадків, що вимагають ручної верифікації (приблизно 5%), процес оцінки результатів може бути додатково оптимізований. Наприклад, перевірка доменів на зовнішніх платформах, таких як VirusTotal, URLVoid та IPQS, може бути повністю автоматизована шляхом інтеграції відповідних API. Такий підхід дозволить зробити процес верифікації більш інтегрованим та швидким.

Аналогічним чином, процедура автоматизованої категоризації доменів, описана у другому розділі також підлягає вдосконаленню. Це може бути досягнуто, зокрема, шляхом впровадження алгоритмів машинного навчання для автоматичного присвоєння категорій доменам. Подібним чином, існуючі зовнішні бази даних або класифікатори можуть бути використані для зменшення обсягу доменів, що потребують індивідуальної верифікації.

Щодо надійності джерела інформації scammer.info, слід зазначити, що хоча всебічне дослідження достовірності даних цієї платформи у рамках даної роботи не проводилося, вона була використана як джерело інформації у низці інших досліджень, пов'язаних з аналізом зловмисної активності. Зважаючи на її регулярну появу в дослідженнях, присвячених "технічним шахрайствам", ми вважаємо її релевантною та надійною базою даних для доповнення можливостей нашого інструменту.

Висновки до третього розділу

У третьому розділі було здійснено всебічну оцінку результатів впровадженої методології запобігання поширенню шкідливого програмного забезпечення на рівні маркетингових стратегій. Представлені результати збору та обробки даних засвідчили ефективність побудованої системи в реальному середовищі цифрової реклами.

У підрозділі 3.1 продемонстровано результати первинного збору рекламних матеріалів, зокрема — охоплення, обсяги даних та динаміку їх

накопичення. У 3.2 проаналізовано розподіл шкідливих і нешкідливих елементів, включно зі статистикою відвідуваності відповідних доменів, що дозволило визначити пріоритетні зони ризику для користувачів.

Підрозділ 3.3 зосереджений на глибшому аналізі зібраної інформації: визначено географічну локалізацію загроз, ідентифіковано цільову аудиторію атак та основні вектори розповсюдження. Проведено також аналіз віку доменів, який дозволив встановити, що значна частина шкідливих ресурсів є недавно створеними. Обґрунтування вибору архітектурного дизайну системи підтвердило його здатність адаптуватися до змін у середовищі загроз.

Підрозділ 3.4 деталізує конкретні випадки шкідливої активності, включаючи аналіз доменів, що маскуються під легітимні бренди, пошукові системи, які сприяють поширенню шкідливих посилань, а також виявлені небезпечні платформи для завантаження програмного забезпечення.

У завершальному підрозділі 3.5 оцінено ефективність використаних методів, зокрема поєднання автоматизованого сканування, семантичного аналізу та маркування даних. Отримані результати свідчать про високу точність і релевантність виявлених загроз, що підтверджує доцільність застосування запропонованої методології в межах систем цифрового маркетингу та кібербезпеки.

Таким чином, розділ продемонстрував не лише практичну дієвість розробленого підходу, але й відкрив можливості для його подальшого масштабування, зокрема через інтеграцію з системами реального часу та інструментами захисту корпоративної IT-інфраструктури

ВИСНОВКИ

В дипломній роботі проведена розробка методології запобігання поширення шкідливого програмного забезпечення на рівні маркетингових стратегій. Для цього було досліджено процес розробки та практичного застосування веб-скрепера на базі фреймворку Playwright з метою виявлення зловмисної рекламної діяльності, що базується на імітації фірмової ідентичності. У результаті аналізу набору даних, що включав 10 067 рекламних оголошень, було успішно ідентифіковано 49 зловмисних доменів, асоційованих з 513 зловмисними оголошеннями. З цієї кількості, 22 домени були класифіковані як випадки імітації фірмової ідентичності. Отримані дані емпірично підтверджують, що несанкціоноване використання брендів є реальною та актуальною загрозою в сучасній екосистемі онлайн-реклами, яка вимагає належної уваги та протидії.

Крім того, результати дослідження виявили певні прогалини в існуючих процесах модерації та верифікації рекламних оголошень на платформі Google, зокрема тих, що демонструються на сторінках результатів пошуку. Незважаючи на значні ресурси, якими володіє Google, виявлення зловмисних оголошень свідчить про потенційну недостатність ефективності механізмів перевірки контенту оголошень та легітимності рекламодавців. Ця ситуація створює суттєву загрозу для користувачів, які покладаються на уявну безпеку та надійність рекламних сервісів Google.

У рамках дипломної роботи також було проведено оцінку ефективності чотирьох інструментів виявлення загроз: VirusTotal, URLVoid, IPQS та scammer.info. Незважаючи на те, що кожен з цих інструментів продемонстрував певні сильні сторони при аналізі конкретних типів зловмисної активності, жоден з них не виявився універсальним рішенням, здатним ефективно ідентифікувати всі випадки зловмисної рекламної діяльності, що базується на імітації фірмової ідентичності. Варіабельність

ефективності між різними категоріями загроз підкреслює складність проблеми та нагальну потребу у вдосконаленні підходів до виявлення цього типу зловмисної рекламної діяльності.

Підсумовуючи, дана робота акцентує увагу на необхідності розробки більш досконалих методів виявлення та впровадження більш суворих процесів перевірки для ефективної боротьби з проблемою зловмисної рекламної діяльності, що базується на імітації фірмової ідентичності, яка все ще залишається актуальною. Створення кращих, більш спеціалізованих інструментів, що інтегрують сильні сторони існуючих рішень, є необхідним кроком для забезпечення безпеки онлайн-рекламних платформ та мінімізації негативних наслідків, пов'язаних з несанкціонованим використанням брендів у зловмисній рекламі і поширенні шкідливого програмного забезпечення.

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. L. Zhou, Z. Kehuan, X. Yinglian, Y. Fang, and W. XiaoFeng, “Knowing your enemy: understanding and detecting malicious web advertising,” in Proceedings of the 2012 ACM Conference on Computer and Communications Security, CCS ’12, p. 674–686, Association for Computing Machinery, 2012.
2. What is real-time bidding (RTB)? | AppsFlyer mobile glossary - https://www.appsflyer.com/glossary/real-time-bidding/
3. M. Milam, “The rise of brandjacking against major brands,” Computer Fraud & Security, vol. 2008, no. 10, pp. 10–13, 2008.
4. Unifying threat context with VirusTotal connectors ~ VirusTotal Blog - https://blog.virustotal.com/2023/10/unifying-threat-context-with-virustotal.html
5. A. Savˇcin, “Avast researchers detect a September surge in malvertising,” 2023. https://blog.avast.com/avast-threat-report-q3-2023-malvertising.
6. B. Krebs, “Using Google Search to Find Software Can Be Risky,” 2024. https://krebsonsecurity.com/2024/01/using-google-search-tofind-software-can-be-risky/.
7. M. Tober, “Zero-Clicks Study,” 2022. https://www.semrush.com/blog/zero-clicks-study/.
8. E. Nowroozi, Abhishek, M. Mohammadi, and M. Conti, “An Adversarial Attack Analysis on Malicious Advertisement URL Detection Framework,” IEEE Transactions on Network and Service Management, vol. 20, no. 2, pp. 1332–1344, 2023.
9. Peng Peng, Limin Yang, Linhai Song, and Gang Wang. 2019. Opening the Blackbox of VirusTotal: Analyzing Online Phishing Scan Engines. In Proceedings of the Internet Measurement Conference, 2019. ACM, 478–485.

10. Pi-hole. 2024. Network-wide ad blocking via your own Linux hardware. Retrieved January 1, 2024 from https://github.com/pi-hole
11. Victor Le Pochat, Tom van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyn-ski, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In 26th Annual Network and Distributed System Security Symposium, 2019. The Internet Society.
12. Quad9. 2024. An open DNS recursive service for free security and high privacy. Retrieved January 1, 2024 from https://www.quad9.net/
13. Aleieldin Salem, Sebastian Banescu, and Alexander Pretschner. 2021. Maat: Automatically Analyzing VirusTotal for Accurate Labeling and Effective Malware Detection. ACM Trans. Priv. Secur. 24, 4 (2021), 25:1–25:35.
14. VirusTotal. 2024. API v3 Overview. Retrieved January 1, 2024 from https: //docs.virustotal.com/reference/overview
15. Jun Wang, Weinan Zhang, and Shuai Yuan. 2016. Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting. CoRR abs/1610.03013 (2016).
16. Apostolis Zarras, Alexandros Kapravelos, Gianluca Stringhini, Thorsten Holz, Christopher Kruegel, and Giovanni Vigna. 2014. The Dark Alleys of Madison Avenue: Understanding Malicious Advertisements. In Proceedings of the 2014 Internet Measurement Conference. ACM, 373–380.
17. IAB, “IAB/PwC Internet Advertising Revenue Report 2024,” 2024. Available at: https://www.iab.com/insights/internet-advertisingrevenue-report-2024/.
18. Enabling Reconstruction of Web Attacks via Efficient Tracking of Live In-Browser JavaScript Executions. In 25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018.

19. Zhou Li, Kehuan Zhang, Yinglian Xie, Fang Yu, and XiaoFeng Wang. 2012. Knowing your enemy: understanding and detecting malicious web advertising. In Proceedings of the 2012 ACM conference on Computer and communications security. ACM, 674–686.
20. Bin Liu, Suman Nath, Ramesh Govindan, and Jie Liu. 2014. {DECAF}: Detecting and Characterizing Ad Fraud in Mobile Apps. In 11th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 14). 57–70.
21. Tim A. Majchrzak, Andreas Biørn-Hansen, and Tor-Morten Grønli. 2018. Progres-sive Web Apps: the Definite Approach to Cross-Platform Development?. In HICSS.
22. Ivano Malavolta, Giuseppe Procaccianti, Paul Noorland, and Petar Vukmirovic. 2017. Assessing the Impact of Service Workers on the Energy Efficiency of Progressive Web Apps. In Proceedings of the International Conference on Mobile Software Engineering and Systems, MOBILESoft '17, Buenos Aires, Argentina, May, 2017. to appear.
23. Rima Masri and Monther Aldwairi. 2017. Automated malicious advertisement detection using VirusTotal, URLVoid, and TrendMicro. In 2017 8th International Conference on Information and Communication Systems (ICICS). IEEE, 336–341.
24. Joseph Medley. 2019. Web Push Notifications: Timely, Relevant, and Precise. https://developers.google.com/web/fundamentals/push-notifications.
25. J. Segura, "Malvertising via brand impersonation is back again," 2023. https://www.malwarebytes.com/blog/threatintelligence/2023/05/malvertising-its-a-jungle-out-there.
26. J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: learning to detect malicious web sites from suspicious urls," in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009, pp. 1245–1254.

27. Z. Li, K. Zhang, Y. Xie, F. Yu, and X. Wang, “Knowing your enemy: understanding and detecting malicious web advertising,” in Proceedings of the 19th ACM Conference on Computer and Communications Security, 2012, pp. 674–686
28. Aditya K Sood and Richard J Enbody. 2011. Malvertising–exploiting web advertising. Computer Fraud & Security 2011, 4 (2011), 11–16.
29. Oleksii Starov, Yuchen Zhou, Xiao Zhang, Najmeh Miramirkhani, and Nick Nikiforakis. 2018. Betrayed by your dashboard: Discovering malicious campaigns via web analytics. In Proceedings of the 2018 World Wide Web Conference. International World Wide Web Conferences Steering Committee, 227–236.
30. M. Antonakakis, R. Perdisci, W. Lee, N. Vasiloglou II, and D. Dagon, “Detecting malware domains at the upper dns hierarchy.” in Proceedings of the 20th USENIX Security Symposium, 2011, pp. 1–16.
31. L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi, “Exposure: Finding malicious domains using passive dns analysis.” in Proceedings of the 18th Annual Network and Distributed System Security Symposium, 2011.
32. K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” in Proceedings of the 13th European Conference on Computer Vision, 2014, pp. 346–361.
33. M. D. Zeiler, “Adadelta: an adaptive learning rate method,” arXiv preprint arXiv:1212.5701, 2012.
34. M. Akiyama, M. Iwamura, and Y. Kawakoya, “Design and implementation of high interaction client honeypot for drive-by-download attacks,” IEICE transactions on communications, vol. 93, no. 5, pp. 1131–1139, 2010.
35. R. Ihaka and R. Gentleman, “R: A language for data analysis and graphics,” Journal of Computational and Graphical Statistics, vol. 5, no. 3, pp. 299–314, 1996.

36. LBE Tech. 2019. Parallel Space - Multiple accounts and Two face. http://parallel-app.com/.
37. Phani Vadrevu and Roberto Perdisci. 2019. What You See is NOT What You Get: Discovering and Tracking Social Engineering Attack Campaigns. In Proceedings of the Internet Measurement Conference. ACM, 308–321.
38. Antoine Vastel, Peter Snyder, and Benjamin Livshits. 2018. Who Filters the Filters: Understanding the Growth, Usefulness and Efficiency of Crowdsourced Ad Blocking. CoRR abs/1810.09160 (2018).

БІБЛІОГРАФІЧНА ДОВІДКА

Тема дипломної роботи: “Розробка методології запобігання поширення шкідливого програмного забезпечення на рівні маркетингових стратегій”

Обсяг пояснювальної записки: 76 аркушів.

Дата закінчення роботи: 10 червня 2025 р.

Підпис студента _____