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

До захисту допущено
Завідувач кафедри

_____ **Дмитро ЛАНДЕ**

(підпис)

«_____»_____2025 р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Системи, технології та математичні методи кібербезпеки»
спеціальності 125 «Кібербезпека та захист інформації»

на тему: Оцінка витоків інформації підприємств на основі аналізу відкритих джерел

Виконав (-ла): здобувач вищої освіти IV курсу, групи ФБ-13
(шифр групи)

Клименко Дар'я Олегівна _____
(прізвище, ім'я, по батькові) (підпис)

Керівник к.т.н, доцент, Стьопочкіна Ірина Валеріївна

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

Рецензент К.ф.-м.н., доцент кафедри ММАД, Терещенко І.М. _____
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище, ім'я, по батькові) (підпис)

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

Київ – 2025 року

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

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 125«Кібербезпека та захист інформації»

Освітньо-професійна програма «Системи, технології та математичні методи кібербезпеки»

ЗАТВЕРДЖУЮ
Завідувач кафедри
Дмитро ЛАНДЕ
(підпис)
«_____»_____2025 р.

ЗАВДАННЯ
на дипломну роботу здобувачу вищої освіти

Клименко Дар'ї Олегівни

1. Тема роботи: Оцінка витоків інформації підприємств на основі аналізу відкритих джерел, керівник роботи: к.т.н., доцент, Стьопочкіна І.В, затверджені наказом по університету від «26» травня р. №1761- с.
2. Термін подання здобувачем вищої освіти роботи «13» червня 2025 р.
3. Вихідні дані до роботи: літературні та інформаційні джерела по методам пошуку у відкритих джерелах та соціальній інженерії.
4. Зміст роботи: огляд методик пошуку у відкритих джерелах, аналіз векторів інтересу соціального інженера, проектування методики пошуку чутливої інформації за допомогою дорків, розробка програмного забезпечення для підрахунку фінальної оцінки.
5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): презентація.
6. Дата видачі завдання: 30 вересня 2024р.

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів дипломної роботи	Примітка
1	Отримання завдання	30.09.2024	Виконано
2	Огляд наукової літератури	01.12.2024-01.02.2025	Виконано
3	Робота над першим розділом	01.02.2025-31.03.2025	Виконано
4	Робота над другим розділом: проектування власної методики	01.04.2025-14.04.2025	Виконано
5	Побудова дорків	14.04.2025-27.04.2025	Виконано
6	Проведення експерименту	28.04.2025-18.04.2025	Виконано
7	Робота над третім розділом: аналіз результатів	19.04.2025-10.06.2025	Виконано
8	Створення презентації	10.06.2025-12.06.2025	Виконано
9	Передзахист дипломної роботи	13.06.2025	Виконано
10	Доопрацювання	13.06.2025-19.06.2025	Виконано
11	Захист дипломної роботи	20.06.2025	

Здобувач вищої освіти _____ Дар’я КЛИМЕНКО

(підпис) **(Власне ім’я, ПРІЗВИЩЕ)**

Науковий керівник _____ Ірина СТЬОПОЧКІНА

(підпис) **(Власне ім’я, ПРІЗВИЩЕ)**

РЕФЕРАТ

Обсяг роботи 69 сторінки, 8 ілюстрацій, 6 таблиць, 20 джерел літератури, 1 додаток.

Об’єкт дослідження: витоки інформації організацій у відкритий доступ.

Предмет дослідження: методики виявлення чутливих даних, які витекли у відкритий доступ.

Мета дослідження: оцінка відкритості підприємства щодо витоків інформації по даних, знайдених у відкритих джерелах.

Методи дослідження: збір і аналіз інформації з відкритих джерел, огляд літератури на відповідну тему, практичний експеримент для перевірки методики.

Ключові слова: доркінг, витоки, оцінка, вразливість, відкриті джерела.

ABSTRACT

Volume of work 69 pages, 8 illustrations, 6 tables, 20 literary sources, and 1 appendix.

Object of research: information leaks of organizations into open access.

Subject of research: methods for detecting sensitive data leaked into open access.

Purpose of the study: to assess the exposure of an enterprise to information leaks based on data found in open sources.

Research methods: collection and analysis of open-source information, literature review on the topic, selection of tools and analysis of their documentation.

Keywords: dorking, leaks, assessment, vulnerability, open sources.

ЗМІСТ

ВСТУП

Атаки соціальної інженерії сьогодні грають найбільшу роль в атаках на установи критичної інфраструктури [1]. Попри велику кількість технічних рішень запобігання витокам, найчастіше інформація випадково чи цілеспрямовано компрометується саме завдяки людському фактору [2]. Це пояснює високу частоту використання методів соціальної інженерії в сучасних кібератаках [3]. Більше того, майже всі успішні атаки базуються на використанні зібраних даних про установу з відкритих джерел [4]. Ця робота є продовженням і логічним розвитком наведених робіт.

Системи збору даних про наявні атаки є одним з векторів задач, які реалізують завчасне виявлення атак [5]. Іншим видом превентивних заходів є системи виявлення витоків даних, що і є актуальною задачею цієї роботи. Використання інструментів OSINT надає можливість виявляти чутливу інформацію у відкритих джерелах ще до того, як її можуть використати зловмисники. Знайдена інформація стає підґрунтям для оцінки відкритості організації для дослідження її зловмисником.

Задачі дослідження:

1) Проаналізувати існуючі рішення.
2) Сформувати алгоритм дій для пошуку та побудувати відповідні дорки;
3) Провести експеримент на тестовій організації;
4) Провести паралель між отриманими даними та MITRE att&ck matrix;
5) Запропонувати критерії дієвості політики безпеки підприємства критичної інфраструктури по витоку даних до відкритого доступу;
6) Зробити висновки по успішності методики.

Метою роботи є оцінка відкритості підприємства щодо витоків інформації по даних, знайдених у відкритих джерелах.

Об'єктом дослідження є витоки інформації організацій у відкритий доступ.

Предметом дослідження є методики виявлення чутливих даних, які витекли у відкритий доступ.

Апробація результатів роботи: робота була апробована на XXIII Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених «теоретичні і прикладні проблеми фізики, математики та інформатики».

Роботу було ***опубліковано*** у збірнику матеріалів XXIII Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених «теоретичні і прикладні проблеми фізики, математики та інформатики».

1 Аналіз існуючих рішень

1.1 Загальний огляд OSINT

OSINT (Open Source Intelligence) - розвідка, яка дає змогу добути інформацію різного роду (військову, політичну, економічну і тд) з відкритих джерел. OSINT розвідники це спеціалісти, які збирають і аналізують ці дані для її подальшого використання.

Існує таке поняття як етикет OSINT. Це набір етичних норм, правил і принципів, яких варто дотримуватись, щоб розвідка була легальною, відкритою та безпечною.

1) Така розвідка завжди має відбуватись відповідно до закону, місцевого та міжнародного. Джерела мають бути виключно відкриті, доступ до яких не потребує зламу, або незаконних маніпуляцій. Використання даних не має порушувати закони про авторське право, персональні дані та інші.

2) Якщо знайдені дані можуть нашкодити людині чи компанії, потрібно переглянути мету розвідки і скоректувати її щоб уникнути негативних наслідків.

3) Те що дані є у відкритому доступі, не означає що етично було б їх поширювати далі, порушуючи принцип конфіденційності. Саме тому, знайшовши приватну інформацію потрібно утриматись від її публікації.

4) Шукати тільки ту інформацію, яка задовольняє чітко сформульовану мету пошуку, яка виправдовує збір даних.

5) При подальшій публікації дослідження, потрібно вказувати джерела, поважаючи авторське право.

6) Завжди підтримувати особисту безпеку і зберігати анонімність при пошуку даних. Не вказувати свої особисті дані при реєстраціях, використовувати VPN, чистити цифровий слід.

Розвідувальний цикл OSINT поділяється на 5 основних етапів [6]:

- планування;
- добування;
- обробка розвідувальних даних;

- аналіз інформації;
- презентація та адаптація результатів.

На початку роботи визначається мета, об’єкт вивчення та обираються джерела для подальшого первинного збору інформації. Після того, як добувається уся доступна інформація, починається її обробка: класифікація, відбір та верифікація. Це створює підґрунтя для аналізу, що включає в себе перевірку достовірності, виявлення закономірностей, визначення загроз та дослідження методів їх протидії, а також перевірка чи задовольняє знайдена інформація початковій меті. На фінальній стадії, отримані дані оформлюється у вигляді звітів, які включають в себе основні висновки та рекомендації, залежно від поставленої мети, грунтуючись на добутих даних.

Добування інформації має три види: [7]

- пасивне;
- напівпасивне;
- активне.

Найчастіше з них використовують саме пасивне. В цьому випадку ціль не буде знати про намагання зібрати про неї інформацію, оскільки цей вид не передбачає прямого контактування з предметом інтересу, є повністю анонімним і досягається розслідуванням лише відкритих джерел. Напівпасивний збір інформації вже використовує більш проактивні дії відносно об’єкту розвідки. Для досягнення мети відправляється обмежений трафік на систему користувача який уподібнюється до звичайного інтернет трафіку. Це дозволяє зібрати поверхневу інформацію і при цьому не піднімає тривогу в жертви. Однак, якщо, все таки, про ці махінації стало відомо, компанія не може пов’язати цей трафік з технікою атакуючого, бо все так само проводилось анонімно. Найбільш ризиковим є активний збір даних, оскільки він включає в себе активну взаємодію з системою користувача/компанії, яка найбільш імовірно лишить сліди в системі і повідомить всім про свою присутність. В процесі сканується система і збирається інформація про відкриті порти, вразливості, додатки веб-серверів

тощо. Атаки за допомогою соціальної інженерії теж належать до цього виду збору інформації.

1.2 Джерела інформації для OSINT

Джерела, які використовуються для OSINT-у є загальнодоступними для всіх користувачів інтернету, на відміну від секретних або обмежених доступом джерел. Основними з них є:

1) соціальні мережі;
2) ЗМІ (Засоби масової інформації): друковані видання, новинні агентства та електронні медіа;
3) урядові звіти та реєстри: економічні та фінансові звіти, звіти про законодавство, звіти про міжнародні відносини, аудити та розслідування;
4) наукові публікації.
5) геопросторові дані: геолокація, картографічні дані, супутникові знімки;
6) метадані знайдених файлів;
7) сіра література: технічні звіти, робочі документи, ділові документи;
8) публічні форуми;
9) витоки баз даних.

Використовуючи якомога більше різних джерел, за конкретним запитом можна знайти багато релевантної інформації. Оскільки кожне з цих джерел є унікальним, дані які воно може надати теж є унікальними. Аналізуючи їх, можна сформувати більш глибоке розуміння ситуації і зробити відповідні прогнози.

Соціальні мережі вже давно є невід’ємною частиною суспільства, джерелом особистої інформації. За допомогою них можна відслідковувати конкретну особу чи групу осіб, дізнаючись більше про спосіб життя, поточну діяльність, наміри чи місцезнаходження об’єкту пошуку. Зокрема, соцмережі можуть використовуватись для ідентифікації кіберзлочинців і для подальшого моніторингу їх активності. Також вони можуть бути корисні якщо потрібно знайти фото або відео матеріали про ту чи іншу подію з різних ракурсів,

ідентифікувати зв’язки між абсолютно не пов’язаними, на перший погляд, людьми, знайти особисті профілі керівників компанії, тощо. Соцмережі можуть надати інформацію про психологічний стан особи або якоїсь групи осіб, що, наприклад, може бути вразливістю до атак соціальної інженерії. З коментарів можна дізнатись реакцію людей на ту чи іншу заяву від офіційних представників або в цілому проаналізувати настрої суспільства.

Нерідко сторінки містять на перший погляд невинну інформацію, яку за бажанням можуть використати зловмисники. Такою інформацією може бути місце роботи, посада або корпоративна пошта. Так наприклад з профілю у LinkedIn можна зрозуміти коли людина вийшла на поточну роботу чи перебуває у пошуку нової. Ця інформація може бути використана для надсилання таргетованого спаму з фейковими вакансіями. Це все може призвести до великих втрат і витоку даних, якщо співробітники не навчені протидії соціальній інженерії і мали необережність відкрити такі листи, або ще гірше, ввести конфіденційні дані компанії.

Зараз, у час процвітання Тік Току, багато компаній стирають формальні межі між працівниками і користувачами, знімаючи короткі відео з життя офісу. Такі відео можуть бути використані зловмисниками для збирання інформації про розташування та планування офісу, кількість працівників або техніку, яка використовується співробітниками для роботи. Використовуючи цю інформацію, не буде складним встановити стеження за життям офісу, вивчити охоронні засоби та складність потрапляння в приміщення офісу, або чи потрібні для цього перепустки. До речі, вигляд самих перепусток теж є чутливою інформацією, знаючи яку зловмиснику легше буде видати себе за співробітника. Поширюючи такі відео, варто розуміти ризики і перевіряти публікації, щоб наприклад, випадково не поширити фото або відео монітору з чутливою інформацією: паролі, програми для моніторингу виконання задач, операційні системи, робочі чати, файли на робочому столі, директорії тощо. Все це можна використати для формування і реалізації атак на мережу або пристрої компанії.

Аналізуються також особисті профілі в Інстаграмі або Фейсбуці (іноді можна підключити аналіз російських Однокласників та ВКонтакте, оскільки раніше ці соцмережі були доволі популярні і зараз ще деяка частина людей користуються ними, попри блокування). Імена домашніх тварин, історії з дитинства, посилання на профілі членів сім'ї – цього вистачить для злому деяких паролів, які можна відновити за допомогою відкритих особистих питань. Інформація про звички, хобі, часто відвідувані місця дають можливість не тільки відслідкувати місце перебування особи, а й допомагають влитись в довіру, знайти чутливі точки дотику до людини. Історії, які дійсні і є доступні до перегляду 24 години, з подорожей можуть підказати зловмисникам що людина не перебуває в місті, її квартира імовірно пустує і на робочому місці її точно не буде.

Дуже часто люди не приховують свої політичні або релігійні вподобання на особистих сторінках у Twitter. Особливо уважно проаналізуючи всю інформацію з соцмереж можна зробити висновок про фінансовий стан людини, що дає ще більше можливостей для шахрайства, причому добираючи відповідні маніпуляції. Наприклад, якщо людина має достатньо грошей і зацікавлена в інвестиціях, зловмисники шляхом маніпуляцій і фіктивних документів можуть легко вкрасти в неї ці гроші, пропонуючи вкласти їх в якийсь неіснуючий проект. Або, якщо в особистому профілі багато згадувань про політичні погляди людини, шахраї, використовуючи цю інформацію, можливо спробують завербувати або підкупити цю особу задля власних інтересів. При цьому розголошуючи інформацію про фінансові проблеми, можна іноді отримати пропозиції легкого та швидкого заробітку, що дуже часто насправді виявляється фішингом і чимось незаконним.

Журналістські розслідування часто стають предметом аналізу для Osint розвідників, оскільки надають не тільки актуальну і сучасну інформацію (якщо це свіжі роботи), а й історичну. Онлайн ЗМІ випереджають своєю зручністю та економлять час, але нерідко онлайн ресурсів не вистачає, і доводиться звертатись до традиційних ЗМІ, архівів тощо, до часів, коли онлайн ресурси не були такі

поширені, або їх навіть не існувало. Зі ЗМІ, як і з соцмереж, можна отримати інформацію про настрої суспільства щодо тої чи іншої теми, оскільки нерідко змі публікують опитування чи голосування, з подальшим обговоренням теми. Інтерв'ю та подкасти можуть містити конфіденційну інформацію, яку монтажери не вирізали, особливо якщо розмова ведеться з впливовою людиною яка має доступ до багатьох секретних даних.

А найголовніше що можуть дати ЗМІ — це новини, актуальні і історичні новини. Завдяки новинам можна встановлювати часові шкали для розуміння послідовності подій, що неабияк корисно, якщо потрібно провести паралелі і аналізувати подібні ситуації, адже, як відомо, історія часто повторюється. Особливо корисним тут можуть бути вузьконаправлені змі чи змі окремих громад, що дає змогу добути унікальні деталі чи специфічні аспекти події, особи чи компанії, яка є предметом аналізу. Для отримання більш менш достовірної інформації про подію, не завадить проаналізувати декілька різних змі, оскільки одна і та сама інформація часто може висвітлюватись з різним контекстом і суттю. Але навіть в глобальних ЗМІ іноді трапляються помилки, і який би вотум довіри до ЗМІ не був, варто завжди перевіряти інформацію знайдену на просторах інтернету і шукати підтвердження. Часто, навіть поширення дезінформації чи фейків теж стає приводом для аналізу і дає змогу вийти на першоджерело фейку, зробити відповідні висновки про те чи куплене змі, або наскільки добросовісно відбирає новини для публікації.

На відміну від ЗМІ, урядові звіти є офіційним державним документом і можуть використовуватись для підтвердження деякої інформації. Зрозуміло не варто нехтувати можливістю того, що ці звіти також можуть нести в собі маніпулятивний характер для коректування настроїв в суспільстві чи впливу на іншу державну структуру або навіть іншу країну. Але часто державні реєстри є джерелом достовірної інформації. Урядові звіти включають в себе: законодавчі акти та нормативні документи, економічні звіти, аудиторські звіти, безпекові звіти, дипломатичні звіти, статистичні звіти, тощо. Всі ці звіти є джерелом даних,

які показують, позицію та стан держави на світовій арені. Наприклад, дані про галузі, в які держава найбільше інвестує коштів, або наскільки фінансовий стан держави залежний від запозичень інших країн. Окрім урядових звітів існують ще різноманітні реєстри. В Україні основні реєстри такі:

- ЄДР (Єдиний державний реєстр юридичних осіб);
- судовий реєстр;
- реєстр боржників;
- реєстр декларацій НАЗК;
- Державний реєстр речових прав на нерухоме майно;
- реєстр судових експертиз, адвокатів, нотаріусів, арбітражних керуючих;
- список осіб, що переховуються від органів влади (МВС).

З цих джерел є можливість дізнатись перелік майна яким володіє людина, або в яких судових провадженнях вона приймає участь, чи перебуває в розшуку.

Урядові звіти охоплюють достатньо великий спектр тем, чим стають корисні для OSINT розвідників. Але в той самий час, інформація в них може бути представлена так, щоб вигідно висвітлити роботу уряду, тож на це теж треба звертати увагу.

Якщо ж при аналізі урядових звітів виникла проблема з розумінням специфічних термінів, можна звернутись до наукових публікацій. Це джерело корисно коли є потреба більше розібратись в темі з якою не надто знайомий, і точно бути певним в достовірності інформації яку вивчаєш, оскільки наукові публікації є найбільш авторитетним та об’єктивним джерелом знань загалом про весь світ. Такі публікації включають в себе дисертації, статті, монографії, наукові журнали, тощо. В контексті кібербезпеки, наприклад, дають можливість ознайомитись з новими методами атаки та алгоритмами захисту від них.

Одним з основних джерел даних також є геопросторові дані — дані, які базуються на використанні супутникових знімків, карт та все що дає розуміння про географічне розташування природного чи побудованого об’єкту. З цих даних

можна дістати інформацію про координати об’єкту інтересу, його переміщення, зміни стану з часом, час доби чи пори року. Це джерело особливо корисне для відстеження ситуацій чи подій які мають геопросторове підґрунтя. Наприклад, переміщення військової техніки та її концентрацію, відстеження і фіксація наслідків стихійних лих, визначення розташування об’єктів критичної інфраструктури, тощо.

На додачу до геопросторових даних, є корисним проаналізувати загалом метадані файлів. Це такі собі дані про дані, представляють собою опис історії, властивостей та походження файлів. Вони дадуть уявлення про те, хто створив файли, покажуть автора/ів документів. Дата створення та редагування файлів дає інформацію про те, коли у файл вносили зміни, що може бути корисно при встановлені достовірності документів. Іноді можуть містити вбудований шлях до файлу, що може дати уявлення про внутрішню файлову структуру системи.

В переліку основних джерел зазначена сіра література – білі книги, тези конференцій та презентацій, документи неурядових компаній, технічні документи, внутрішня документація організацій. Вона, також, надає дані, які могли бути неопубліковані або були розсекречені внаслідок витоку. Це можуть бути звіти про хакерські атаки на компанію, результати тестування безпеки, внутрішні політики та регламенти доступу, чернетки документів, які могли бути необережно опубліковані. Залежно від мети OSINT аналізу, сіра література може надати унікальні дані, які ще не увійшли до офіційних джерел.

Публічні форуми також вносять свою частку корисної інформації, завдяки активним обговоренням різних тем. На форумах можна почитати про нові вразливості, моніторити активність хакерів, їх обговорення нових чи старих атак. Співробітники можуть випадково або навмисно поширювати на форумах секретну інформацію про компанію.

Також можна знайти обговорення злитих баз даних. Іноді їх можна знайти у відкритому доступі, іноді для цього потрібно зайти в даркнет. Такі дані містять особисту інформацію компанії чи її користувачів, або ж іншу секретну

інформацію яка була скомпроментована внаслідок хакерської атаки, витоки через співробітників чи проблеми з безпекою системи.

Отже, для OSINT аналізу використовується багато джерел, особливо якщо їх комбінувати можна добути ще більше інформації для аналізу.

1.3 Методи та інструменти OSINT для пошуку чутливої інформації

Зазвичай, OSINT розвідка не обмежується одним методом, що логічно, оскільки для обширного збору інформації варто використовувати декілька різних методик залежно від цілі і джерела пошуку. З основних і найбільш поширених методик можна виділити:

- пошук у пошукових системах;
- аналіз соціальних мереж;
- моніторинг блогів, новин та медіа;
- моніторинг форумів;
- веб-скрейпінг та парсинг;
- аналіз доменів та інфраструктури;
- аналіз файлів, документів та метаданих;
- геолокаційний аналіз;
- аналіз та пошук зображень і медіафайлів;
- збір інформації з даркнет.

Комбінування цих методів дає змогу отримати різноманітні дані, що розширює подальші пошуки і є підгрунтям для якісного аналізу.

1.3.1 Пошук у пошукових системах

Будь-яка OSINT розвідка починається з ключового інструменту пошуку – пошукових систем. Звісно такий процес не обмежується простим вводом запиту в браузер, і кожна пошукова система має свої особливості, якими треба користуватись для цілеспрямованого і конкретного добування даних. Перше, на що потрібно звернути увагу, це ціль дослідження і її геолокацію. Якщо це США, то для роботи обирається Google та Bing, оскільки це основні пошукові системи

для цієї країни. Для росії це буде Yandex, а для Китаю – Baidu. Друге, обов'язково використовуються пошукові оператори (дорки), які значно звужують область пошуку. Це такі команди, які розширюють можливості звичайного пошуку, і дозволяють знайти специфічний тип контенту або витоки даних. Базово, правило пошукового оператору для більшості браузерів складається з типу оператора (наприклад site для пошуку контенту на конкретному домені), і самого параметру, ключової фрази, що саме потрібно шукати. Тип оператора і сам параметр розділяються знаком ":". Використання таких операторів може показати навіть ті сторінки, що були змінені на приватні, але вже встигли бути проіндексовані Google або іншим браузером. Ще деякі популярні оператори, та що вони можуть робити наведено в таблиці нижче:

Таблиця 1.1 - Популярні оператори для покращення пошуку

Оператор:	**Приклад:**	**Призначення:**
"<запит>"	"OSINT"	шукає точну фразу
OR	password OR pass	задає альтернативу пошуку
-	login -facebook	вилучає зі списку результатів
site:	site:linkedin.com	шукає тільки на вказаному сайті
filetype: ext:	filetype:pdf ext:xls	шукає конкретний тип файлу
inurl:	inurl:daria	шукає сторінки з певним словом в URL
intitle:	intitle:leak	певні слова повинні бути в заголовку
allintitle:	allintitle:leak oci	всі слова повинні бути в заголовку
intext:	intext:"confidential"	слова мають бути в тілі сторінки
cache:	cache:example.com	показує кешовану версію сторінки

Якщо попередній пошук не дав жодних корисних даних, можна скористатися сервісом Google Alerts. Він застосовується для автоматизованого

моніторингу нових індексацій Google, тобто при налаштуванні, буде надсилати сповіщення, якщо десь з’явиться інформація, яка відповідає введеному запиту. Але є один мінус – без Google пошти він не буде працювати. Для випадку коли важливо зберегти анонімність є аналогічний сервіс TalkWalker Alerts. На відміну від першого, цей не вимагає реєстрації, і моніторить трохи більше джерел, наприклад соцмережі та коментарі [8].

Якщо ж і в цьому випадку нічого немає, або веб-сторінки які потрібні є недоступними, оскільки були видалені чи змінені, застосовуються кешовані версії веб-сайтів або веб-архіви. Вони зберігають історичні копії веб-сторінок і надають можливість їх перегляду. Google, Bing, Yandex, Baidu зберігають кешовані версії сторінок. Завдяки ним можна ознайомитись з останньою версією веб-сторінки. Якщо ж потрібно проаналізувати зміни, перевірити інформацію на достовірність чи порівняти різні версії зроблені в різні дні/місяці/роки підійдуть історичні копії. Одним з найкращих архівів для цього є Wayback Machine. Архів зберігає текст, зображення, повний інтерфейс веб-сторінок, і на додачу регулярно створює копії, що і надає можливість перегляду веб-сторінки в різні роки.

На початку розслідування, коли ще не відомо що саме потрібно шукати, можна скористатись інструментом автоматизації, а саме IntelTechniques Search Tool. Зробивши запит, відповідно до мети розслідування, одночасно запуститься пошук у 30 різних джерелах, які включають в себе вже розглянуті браузери, а також Wayback Machine. Таким чином можна знайти персональні дані, новини, блоги, обговорення на форумах, архіви сайті, наукові публікації і навіть витоки даних, дані з Darknet.

1.3.2 Аналіз соціальних мереж

Як вже було сказано, соцмережі є джерелом чутливих даних про особисте життя людини. Аналіз соціальних мереж включає в себе вивчення публічних профілів, для того щоб зібрати досьє на конкретну особу або компанію. Додатково, аналізуються дописи, коментарі та відповіді на них, списки

підписників та відмітки, з метою пошуку зв’язків, а також метадані. Також вивчається активність користувачів, зміни опису профілів, часові мітки [7].

Для анонімного стеження за профілями, публічними та приватними, використовують фейковий акаунт. Аналіз публікацій може показати сталі шаблони поведінки, часові мітки дадуть розуміння про часи роботи чи часову зону. Щоб автоматизувати цей процес, існує багато різноманітних інструментів, деякі з них націлені на конкретну соц-мережу. Наприклад, Hadzy, показує статистику коментарів на YouTube, є можливість фільтрації за датою, кількістю лайків та відповідей, а також здатен шукати коментарі за конкретним нікнеймом. telegram-tracker – інструмент для парсингу Telegram каналів. Працює за принципом пошуку цільової інформації і виводить список каналів, в яких згадується та чи інша інформація. Social Searcher навпаки, націлений на більшість популярних соцмереж і моніторить їх в режимі реального часу. Пошук можна провести за іменем користувача, тренду, або вказавши згадки які мають бути в публікації. Для візуалізації залежностей використовують Maltego – інструмент націлений на побудову та аналіз зв’язків між суб’єктами та об’єктами. Окрім встановлення зв’язків між людьми, також враховує пошти, акаунти в соцмережах, адресами та номерами телефону, компаніями, веб-сайтами, метаданими та інтернет інфраструктурою, що включає в себе доменні імена, DNS записи та IP адреси. Social Links – інструмент що працює на базі Maltego, та додає до нього функціонал подібний до Graph Search. Тобто показує пролайкані пости, відмітки та теги в публікаціях.

1.3.3 Моніторинг блогів, новин та медіа

Сучасні інструменти полегшили роботу аналізу різних новин та медіа. Наприклад, Google пропонує такі архіви новин: Google News Archive і Google Newspaper Archive. Різниця між ними в тому, що перший містить архіви діджитал медіа, а другий пропонує скани деяких друковані видання. Оскільки Google не охоплює всі публікації, є інший ресурс, який займається архівуванням

друкованих видань. Newspaper Archive тримає найбільшу базу даних друкованих газет, починаючи з 1700 років і дотепер [8]. Ще одним з найбільших ресурсів новин є All You Can Read. Це велика база даних ЗМІ різних країн, при користуванні дає зручні можливості фільтрації за країною походження журналу/газети, і за темами.

Якщо говорити не лише про архіви, а й про актуальні публікації новин або блогів, то тут корисним будуть RSS-канали. RSS (Really Simple Syndication) – це такий формат, завдяки якому легко моніторити актуальні новини і читати їх в одному місці. Однак не всі веб ресурси підтримують RSS. Щоб його додати на свою сторінку, потрібно створити окремий файл з розширенням .xml і підключити до свого сайту. Після цього користувачі, які бажають отримувати свіжі новини, можуть підписатись на цей файл в своїй RSS-стрічці і читати всі різноманітні публікації в одному місці, як тільки вони публікуються [9].

Для взаємодії з RSS існує багато інструментів, наприклад inoreader. Це платформа, яка збирає для користувачів новини на які вони підписані, й окрім того допомагає їх структурувати за тегами, фільтрувати контент і налаштовувати сповіщення. Feedly – подібний інструмент, крім звичайного новинного збирача, пропонує також моніторинг блогів з безпеки, бази даних вразливостей, державні сайти, тощо. Також має тариф з штучним інтелектом, який аналізує стрічку і здатний підбирати новини, які можуть бути корисні для OSINT-розвідувальників. Схожим інструментом є Silobreaker, який здатний моніторити та надавати глибоке і своєчасне звітування про нові кіберзагрози. Крім того, він оцінює їхню важливість для компанії і аналізує схожість між загрозами і MITRE att&ck, а також пропонує візуалізацію отриманих даних [10].

1.3.4 Моніторинг форумів

Онлайн форуми незамінне джерело для моніторингу активності хакерів. Їх аналіз здійснюється для того, щоб бути в курсі більшості нових загроз і розрахувати відповідні ризики.

Не всі онлайн форуми індексуються пошуковими системами. Саме тому, потрібну з них інформацію часто неможливо знайти через Google. Але деякі форуми створюють свій пошуковий механізм, наприклад Reddit і його Reddit search, який дозволяє прямо на сторінці форуму шукати потрібну інформацію. Також за допомогою Reddit Search можна спробувати пошукати дані сторінок форуму, які були видалені, або ж звернутись до сайту Pushshift, який є архівом з понад 300гб даних того ж таки Reddit починаючи з 2005 року [8]. Не зле буде перевірити 4chan, Hacker News, Craigslist, або інші форуми на більш спеціалізовану тему, яка залежить від мети дослідження. Також не варто знецінювати вже не активні форуми.

1.3.5 Веб-скрепінг та парсинг

Веб-скрепінг – це процес автоматичного збирання даних з веб-ресурсів, в тому числі соцмереж і форумів. На відміну від ручного збору даних, цей метод використовує автоматизовані інструменти [11].

Процес веб-скреппінгу складається з таких етапів:

1) За допомогою веб-скрепінг утиліти або саморобного коду надсилаються HTTP-запити до відповідного веб-ресурсу.

2) Очікується відповідь, якщо вона позитивна і сторінка існує, програма почне аналіз і збір відповідних даних.

3) Після цього дані впорядковуються і зберігаються в читабельних форматах, таких як XLS, CSV, SQL або XML.

Цей метод корисний, якщо, наприклад, потрібно зібрати всі доступні пошти з сайту з великою кількістю сторінок, щоб не перебирати їх усі вручну.

Парсинг – це наступний етап, обробка вже зібраних даних. Зазвичай інструменти для веб-скрепінгу вже використовують парсинг для обробки даних знайдених веб-скрепінгом.

theHarvester – одна з можливих утиліт для цього завдання. Цей інструмент використовується для збирання субдоменів, поштових адрес, хостів, відкритих

портів або імен співробітників з веб-сторінок у Goggle, Bing, LinkedIn, Twitter, Yahoo, тощо. Звісно є й інші інструменти, деякі з них використовуються в більш специфічних випадках. Наприклад, Email Extractor – інструмент націлений саме на збір поштових адрес. Metagoofil або OOMetaExtractor збирають метадані з веб-сторінок. Censys та Certificate Search шукають сертифікати пов'язані з заданим доменом. Spy On Web, Alexa або Moon Search покажуть різноманітний аналіз та статистику вебсторінок [7] Instant Data Scraper ще один з інструментів для парсингу та веб-скрепінгу.

1.3.6 Аналіз доменів та інфраструктури сайтів

Аналіз доменів та інфраструктури сайтів є важливим кроком в OSINT розвідці для пошуку технічної інформації про цільову веб сторінку. Також, цей метод використовується для пошуку зв'язків між доменами, виявлення пов'язаних сервісів, інших доменів, технічні характерстики хостингу, сертифікати безпеки, IP-адреси та інші артефакти. Веб-сайти містяться на конкретних доменах. Наприклад, для сайту www.example.com, example – це домен. Аналіз домену дає можливість отримати відомості про:

- реєстратора домену;
- дату створення та дії домену;
- DNS-записи (A, MX, NS, TXT);
- сертифікати SSL та їхню історію;
- IP-адреси, до яких прив'язаний домен;
- субдомени.

Такий підхід дозволяє візуалізувати інфраструктуру сайтів, дає підґрунтя для аналізу вразливостей або зв'язків. В ході цього аналізу, розглядається така інформація:

- поточна реєстрація домену;
- конфігурації IP-адрес та DNS-записів;

- історія реєстрацій домену;

дані про сервер та вміст сайту;

- розташування та наявність субдоменів;
- інформація з файлу robots.txt;
- реплікація (копіювання) контенту на інших сайтах.

При реєстрації домену сайт вимагає введення інформації про реєстранта, адміністративного та технічного контактів, пов'язаних із доменом. Це може бути як одна, так і три різні особи. Контактні дані включають повне ім'я, назву компанії, фізичну адресу, номер телефону та адресу електронної пошти. Ця інформація надається реєстратору доменного імені сервісом, через який домен було придбано. Далі ці дані передаються до ICANN (Internet Corporation for Assigned Names and Numbers). Після цього інформація стає загальнодоступною і може бути отримана через сотні онлайн-ресурсів [8].

Майже кожен вебсайт має в своєму кореневому каталозі файл robots.txt, який вказує на те, які його сторінки не варто індексувати пошуковими системами. Його легко відкрити і перевірити за посиланням http://www.example.com/robots.txt, в якому потрібно вказати цільовий веб-сайт. Цей файл може містити цікаві для дослідника сторінки, які адміністратор сайту з якихось причин не хоче, щоб були знайдені пошуковими системами.

В процесі аналізу домену неможливо обійтись без аналізу IP-адрес. IP (Internet Protocol) адреса — це числове позначення пристрою в мережі. При введенні доменного імені браузер насправді підключається до IP-адреси. До одного домену може бути прив'язана лише одна IP адреса, в той час як до однієї IP-адреси може бути прив'язано декілька доменів [8]. Окрім аналізу веб-сайтів, дає можливість визначити географічне розташування користувача або компанії.

Для базового ознайомлення з технічними характеристиками домену сайту використовують сервіс Whois, але є й інші інструменти. Сервіс ViewDNS пропонує декілька різних інструментів: Whois, Reverse IP, Port Scanner, IP History, TraceRoute. Для аналізу доменів існують інструменти зворотного пошуку.

Whoisology один з таких інструментів, за вхідними даними, типу поштова адреса, ім’я, адреса, компанія, знаходить домен або домени які пов’язані з цими даними. Аналогом попереднього інструменту є DNS Trails, який фокусується на зв’язки між доменами. DomainHistory.net показує історію Whois-записів [8]. Visual Site Mapper візуалізує інфраструктуру сайту, показує які сторінки є центральними. VirusTotal надає можливість перевірити чи зловмисний сайт. Сервіси типу bitly, tiny.cc, goo.gl, bit.d скорочують великі URL, окрім того фіксують інформацію, хто, коли і скільки разів переходив за посиланням. Certificate Details покаже домени прив'язані до одного SSL-сертифікату [12]. Для роботи з IP адресами зручними будуть інструменти IPLocation або вже розглянуті з сервісу ViewDNS. Неможливо не згадати про Shodan, це пошуковик, який шукає не сайти, а пристрої, підключені до Інтернету. Він сканує інтернет і зберігає банери пристроїв — метадані, які пристрій надсилає у відповідь при з’єднанні. Корисно, якщо потрібно перевірити які пристрої підключені до мережі, виявити сервіси, які працюють на цільовому IP, а також виявити вразливі або незахищені пристрої.

1.3.7 Аналіз файлів, документів та метаданих

Одним з основних пунктів дослідження OSINT розвідки є перевірка метаданих знайдених файлів, документів. Для витягування метаданих з файлів, спеціалісти використовують спеціальні готові інструменти. EXIF Viever, EXIF Viever Pro, EXIFdata.com, ExifTool by Phil Harvey, – це деякі приклади з безкоштовних інструментів, які виконують пошук метаданих. fotoforensics.com витягує метадані з зображень. FOCA - додаток що може проаналізувати метадані зібрані з веб сторінок [12]. Також метадані шукає CyberChef recipe, але основна його мета це все таки декодування та дешифрування даних. Часто метадані містять мініатюру зображень, і навіть якщо фотографія була обрізана, ця мініатюра може містити повне початкове зображення. Інструменти типу Jeffrey’s Exif Viewer допоможуть перевірити зображення на наявність цієї мініатюри.

Аналіз метаданих здійснюється з метою визначити власника файлу, виявлення геолокації, пошуку слідів змін зображення чи відео, де і коли створено було файл. Наприклад, за фотографією на сайті компанії, можна визначити точне розташування офісу, якщо метадані міститимуть цю інформацію. Загалом зібрати інформацію яка на перший погляд не завжди є доступна.

1.3.8 Геолокаційний аналіз

Як вже було помічено, більшість методик між собою пересікаються, отож щоб знайти та проаналізувати точні координати використовуються інструменти, для пошуку метаданих. Але є і окремі інструменти, націлені конкретно на пошук геоданих і розроблені для роботи з ними, оскільки крім аналізу метаданих, використовується метод візуального спостереження та аналізу. Цікавим є інструмент 3D SUN-PATH, що показує розташування сонця і його променів щодо географічного об’єкту. Цей інструмент може бути використаний для аналізу зображень та тіней, і пошуку геолокації, де саме зображення було зроблено. Або наприклад MW Geofind, який дає змогу за геолокацією здійснити пошук відео на YouTube. Також, може бути корисним використання сервісів EarthCam, InseCam, Opentopia, Windy, Pictimo, які містять базу онлайн вебкамер [12]. Звісно також використовуються базові інструменти для перегляду мап, такі як Google Maps, Earth View, Bing Maps. Для візуалізації знайдених локацій зручно використовувати GPS Visualizer, який підтримує багато форматів.

Аналізуються геодані зазвичай з метою визначити місце зйомки враховуючи деталі зображення (вивіски, дорожні знаки, напрямок тіней від сонця, і особливо природа) для пошуку локації об’єкту або його переміщення.

1.3.9 Аналіз та пошук зображень та медіафайлів

В часи коли камеру має кожен телефон, інтернет переповнений фото та відео різних видів. Основними цілями аналізу зображень та медіафайлів є пошук першоджерела, підтвердження або спростування інформації, встановлення

геолокації, тощо. В основному вся суть в тому щоб добути ще більше інформації. Для базового пошуку можна використовувати Google Images та Bing Images, оскільки більшість зображень які є в інтернеті імовірно мають бути індексовані і будуть знайдені цими пошуковими системами. flickr та gettyimages сервіси які надають фотографії з різних заходів та подій. Якщо ж потрібно провести пошук за зображення, замість словесних пошукових запитів підійде реверс-пошук зображень. Це може бути корисним, якщо потрібно ідентифікувати сайти, які містять ідентичне або подібне зображення до наявного. Одним з найбазовіших та найпотужніших таких інструментів є Google Reverse Image Search. Менш потужний, але який не варто недооцінювати, є TinEye. В окремих випадках, корисним буде пошук в російському Yandex Images та Baidu Images. Але завжди варто пам'ятати про безпеку і не завантажувати для пошуку фотографії, які не мають витекти в інтернет. Деякі сервіси пропонують декілька варіантів завантаження фотографій, наприклад Image Raider. Окрім звичайного завантаження файлу зображення, можна ввести URL і сервіс завантажить з нього зображення. Але головна особливість Image Raider в тому, що він дозволяє завантажити до 20 зображень одночасно, що неабияк розширює кількість та точність знайдених результатів. Інший інструмент Plag Hunter було створено для пошуку зображень які порушують закон про авторське право, але воно також може бути корисним для моніторингу. Сервіс одразу показує де використовується це зображення, а в подальшому сканує інтернет, і як тільки зображення десь з’явиться, прийде оповіщення на вказану пошту. Що ж до відео, базовим інструментом для пошуку можна вважати YouTube. Лише за хвилину завантажуються відео тривалістю 48 годин, згідно з даними з офіційного сайту YouTube. Однак не дивлячись на такий великий трафік, Google Videos все ж містить ще більшу базу відео ніж YouTube, тож при відсутності результатів на YouTube, можна використати Google Videos. Internet Archive зберігає відео як аматорські та комерційні, які були колись опубліковані в інтернеті. А ще в ньому існує сервіс TV News Archive який зберігає теленовини, і окрім того надає

субтитри до кожного випуску, тож існує можливість шукати теленовини за ключовими словами. В пошуку відеофайлів також можна використовувати реверс-пошук. Наприклад, YouTube надає 4 кадри для кожного відео яке було завантажено, тож маючи URL цих кадрів, можна здійснити реверс-пошук, так само як з зображеннями. В Instagram так само можна здійснити пошук, використавши основний кадр для відео, який знаходиться у вихідному коді сторінки. IntelTechniques має інструмент Reverse Video Search Tool, який автоматизує цю схему. Достатньо просто ввести URL цільового відео і сервіс видасть результати зворотного пошуку в різних пошукових системах.

Після використання технік збору даних, можна перейти до самого аналізу зображень та відео. Тут можна використати вже розглянуті інструменти для добування метаданих та геоданих. Якщо потрібно вивести транскрипцію аудіо або відеофайлу допоможе Sonix. imageenlarger та letsenhance допоможуть відновити якість зображення і збільшити його чіткість. Fotoforensics показує зміни, якщо такі були, створює темну копію зображення, на якому світлі ділянки позначають що зображення було змінене. Forensically схожий інструмент на попередній, здатний проганяти зображення через світлові спектри і відстежувати зміни. Abby Fine Reader зчитає будь-яку мову з картинок. Тож, інструментів для роботи з зображеннями, а також їх багатьох аналогів, доволі багато і для більшості задач їх вистачає.

1.3.10 Збір інформації з даркнет [7]

Інтернет умовно поділяється на три рівні доступності інформації: surface web (поверхневий інтернет), deep web (глибинний інтернет) і dark web (темна мережа, даркнет). Різниця між цими рівнями у доступності та індексації інформації.

- Surface web – це частина Інтернету, яка індексується пошуковими системами (Google, Bing тощо). Сюди входять публічні веб-сторінки, доступні кожному користувачеві за допомогою стандартних запитів.

- Deep web – не індексується пошуковими системами Інтернету. Цей рівень включає в себе закриті, непублічні або приватні ресурси: особисті кабінети, внутрішні бази даних, корпоративні системи тощо. Для доступу до deep web достатньо використати звичайний браузер і знати що шукати, мати пряме посилання або відповідні облікові дані. Також цей рівень є легальним.
- Dark web – частково є deep web, але потребує спеціального програмного забезпечення для його доступу, типу Tor (The Onion Router), I2P, Freenet. Більшість сайтів в даркнеті мають специфічні домени, типу .onion. Окрім сайтів, даркнет містить в собі багато чатів, поштові сервіси, P2P-зв'язки, файлові хаби тощо. Такий рівень забезпечує високу анонімність, конфіденційність і велику складність відстеження, що робить його найпопулярнішою платформою для ведення нелегальної діяльності.

Більшість інформації легко можна знайти у surface web, використовуючи пошукові системи і методи, розглянути в попередніх пунктах. Однак якщо такий пошук не дав результату, Osint розвідники починають спускатись рівнями інтернету, спочатку аналізуючи deep web, а потім доходять і до даркнету. Цілями такого пошуку є виявлення витоків даних, моніторинг ринків даркнету (з продажу зброї, наркотиків, фальшивих документів або викраденої конфіденційної інформації), а також виявлення планування атак різного виду, в тому числі кібератак, інструкції для цих атак. Програмне забезпечення яке використовується для доступу до даркнету, може використовуватись для збереження анонімності при пошуку в surface і deep web. Також, даркнет, а конкретно мережа Tor, має свої пошукові системи: Ahmia, Candle, DuckDuckGo, Torch, Searx, а також соцмережі: Atlayo, BlackBook, Daniel’s Chat і деякі поштові сервіси: Onion Mail, TorBox, Mail2Tor. Другою, найбільш популярною мережею даркнету є I2P (Invisible Internet Project), домени розміщені тут, мають розширення .i2p. Ці дві мережі мають суттєві відмінності:

- Tor був створений для анонімного перегляду звичайного інтернету, в той час як I2P розроблена як самостійна мережа, і обмін даними відбувається виключно всередині неї.
- В них різні методи маршрутизації: Tor використовує circuit switching (встановлюється постійний канал на час сеансу зв’язку), а I2P – packet switching (дані передаються у вигляді окремих пакетів, які можуть передаватись різними маршрутами).
- Tor використовує один двосторонній тунель, а I2P має окремі тунелі для вхідного і вихідного трафіку.
- Tor використовує централізовану структуру директорій, I2P — повністю децентралізований, кожен вузол зберігає дані лише про відомі йому вузли.
- Tor сумісний з будь-яким додатком, що підтримує SOCKS. I2P вимагає власну API-інтеграцію, що ускладнює використання, але підвищує рівень анонімності.

Третім інструментом для доступу в даркнет є Freenet – децентралізована peer-to-peer мережа для анонімної публікації та збереження інформації. На відміну від Tor та I2P, Freenet фокусується саме на довготривалому зберіганні контенту, а не на анонімному перегляді інтернету в реальному часі. Інформація в Freenet доступна не за URL-адресами, а через криптографічні ключі. Кожен користувач є одночасно і клієнтом, і сервером, одночасно надаючи дані для скачування і скачуючи їх.

Враховуючи можливості які даркнет надає своїм збереженням анонімності, це робить легким пошук не тільки для OSINT розвідників, а й, на жаль, для хакерів і зловмисників.

1.4 Приклади використання OSINT для реалізації атак на об’єкти критичної інфраструктури

Сьогодні критична інфраструктура як ніколи залежна від інформаційних систем, що створює підґрунтя для атак пов’язані з OSINT активністю. Через відкриті джерела зловмисники можуть отримати доступ до документів з конфігураціями систем, контактів відповідальних осіб, детального опису технологічних процесів тощо. Така інформація часто потрапляє у відкритий доступ внаслідок недбалості співробітників, помилок у налаштуванні прав доступу або публікації документації без належної перевірки. Ці дані можуть бути використані на наступному етапі атаки, зокрема у рамках соціальної інженерії. Соціальна інженерія — це будь-які дії, спрямовані на те, щоб змусити людину зробити дію, яка може не відповідати її власним інтересам.

В штатах, наприклад, найбільш поширеними атаками на критичну інфраструктуру є фішингові атаки. Такі атаки створені на базі масової розсилки електронних листів, sms-повідомлень або дзвінків, які містять запит на надання конфіденційної інформації. Таким чином зловмисники отримують логіни, паролі, поширюють шкідливе програмне забезпечення, з метою добути більше чутливих даних і дістати несанкціонований доступ до системи. У звичайного фішинга є більш маніпулятивний різновид – spear phishing (цілеспрямований фішинг), який на відміну від базової розсилки націлений на конкретну особу чи компанію. Підґрунтям для атаки цілеспрямованого фішингу можна вважати OSINT розвідку, оскільки зазвичай саме так зловмисники збирають особисту чи корпоративну інформацію. Також, при поширенні програм-вимагачів (ransomware) часто використовують цілеспрямований фішинг, тому можна вважати, що такі атаки також починаються з OSINT-у і базуються на витоку даних.

Одним з прикладів є атака на компанію Colonial Pipeline у 2021 році. Colonial Pipeline — це трубопровідна система протяжністю понад 5 500 миль, яка

з’єднує Техас із Нью-Йорком. 7 травня 2021 року компанія, імовірно внаслідок витоку облікових даних VPN, стала жертвою програми-вимагача. Шкідливе програмне забезпечення уразило їх систему ведення обліку і рахунків, і компанія, щоб не допустити поширення програми-вимагача, була змушена призупинити роботу. Ця атака призвела до дефіциту пального в 17 штатах, крім цього компанія сплатила хакерам великий грошовий викуп [13].

Іншим прикладом є нафтопромислова компанія Halliburton, яка стала жертвою кібератаки і підтвердила це 23 серпня 2024 року. Хоча деталей кібератаки не було опубліковано, було вказано, що зловмисники не обов’язково використовували новітні технології для зламу, а скористались помилками в заходах безпеки. Група хакерів RansomHub, якій приписують цю атаку, часто в своїх атаках використовує дані добуті з даркнету, які витекли за межі корпоративної мережі. Венкі Раджу, технічний директор ColorTokens, вказав, що завдяки інструментам типу Shodan чи smap, навіть хакерам-аматорам легко виявити незахищені пристрої в мережі. Отже, попри малу кількість інформації щодо самої атаки, враховуючи факти вище можна припустити що атака була реалізована на базі OSINT розвідки і використанні скомпрометованих даних [14].

З 14 по 25 квітня, хакери незаконно отримали доступ до хмарного середовища американського постачальника телекомунікаційних послуг AT&T, в результаті чого дані мільйонів користувачів були викрадені. Інформація щодо дзвінків і текстових повідомлень які відбулися приблизно з 1 травня по 31 жовтня 2022 року, а також 2 січня 2023 року була скомпрометована. Також дані містять ідентифікаційні номери станцій вишок, до яких підключався телефон під час дзвінка, що потенційно дає можливість визначити місцеположення телефону. Важливим є факт, що зловмисники для цієї атаки використовували попередні витоки даних. Більше того, дані добуті в процесі самої атаки також можуть використовуватись для подальших атак, наприклад для побудови графів дзвінків і визначення зв’язків між людьми [15].

Більш свіжим є приклад типового фішингу, реалізованим проти українського оборонного сектору, який став основою для поширення шкідливого програмного забезпечення DCRat через месенджер Signal. Атака передбачала надсилання нібито архівів звітів про проведення нарад, з проханнями ознайомитись з ними. Архів містив PDF-приманку та шифратор DarkTortilla, який розшифровує та запускає шкідливе програмне забезпечення DCRat, яке в свою чергу встановлює віддалене керування заражених пристроїв. Щоб шанс реалізації атаки був вище, хакери використовували вже зламані акаунти, до яким користувачі довіряли [16].

Висновки до Розділу 1

OSINT є дуже серйозним інструментом подвійного призначення: збору чутливих даних не тільки для кібер фахівців, а й хакерів. Його ефективність безпосередньо залежить від глибини пошуку, і вміння правильно інтерпретувати знайдені дані. Основні джерела OSINT, попри відкритість і доступність, нерідко містять чутливу і приватну інформацію, яка може бути використана для підготовки атаки. Методики OSINT дозволяють зловмисникам створювати деталізовані профілі цілей і знаходити вразливості системи.

Аналіз реальних інцидентів показує, що більшість кібератак, таких як фішинг або поширення шкідливого програмного забезпечення, реалізуються на базі витоків і використанні чутливих даних. Це робить OSINT розвідку ключовим елементом, що сприяє успішній реалізації подібних атак.

2 Розробка методики

2.1 Атаки соціальної інженерії на сектори критичної інфраструктури

Було проведено аналіз кібератак на об’єкти критичної інфраструктури в Україні здійснених з початку 2022 року і до кінця 2023 року, а також виділено атаки реалізовані за допомогою соціальної інженерії на об’єкти критичної інфраструктури, а саме:

- фішинг,
- цілеспрямований фішинг,
- доставка шкідливого програмного забезпечення за допомогою фішингу;
- імітація офіційного листування з використанням скомпрометованих акаунтів;
- використання підробок офіційних платформ, для поширення шкідливого програмного забезпечення [17].

Нижче наведена статистика атак відфільтрована відповідно за секторами критичної інфраструктури:

Таблиця 2.1 - Статистика атак за період 2022-2023 роки

Сектор: “Енергетика”	
Дата:	**Короткий опис атаки:**
01-02-2022	Цілеспрямована фішингова атака на співробітника української енергетичної компанії з метою доставки шкідливих програм SaintBot і OutSteel, призначених для завантаження додаткового ПЗ та викрадення документів.
04-09-2023	Кібератака на об’єкт критичної енергетичної інфраструктури України.

Продовження таблиці 2.1

Сектор: “Інформаційних технологій та комунікацій”	
Дата:	**Короткий опис атаки:**
19-07-2022	Група Turla створила фальшиву програму "CyberAzov" для здійснення DDoS-атак, і використовувала підроблений домен для його завантаження, що імітував домен полку “Азов”.
19-06-2023	Користувачам надсилали фальшиві електронні листи, що маскувалися під офіційну поштову службу України. Листи містили PDF-файл зі шкідливим посиланням, яке вело на підроблений сайт, що імітував справжній сервіс. Внаслідок цього зловмисники змогли отримати логіни та паролі користувачів.
08-07-2023	Фішингова атака, спрямована на отримання аутентифікаційних даних користувачів українських публічних поштових служб.
27-09-2023	Група UAC-0165 здійснила серію кібератак проти 11 українських телекомунікаційних провайдерів. Використовуючи раніше скомпрометовані системи, зловмисники сканували мережі на наявність відкритих портів і отримували віддалений доступ. Після проникнення в систему зловмисники розгорнули деструктивні скрипти, що призвело до переривань в наданні послуг споживачам.
Сектор: “Фінансові установи”	
Дата:	**Короткий опис атаки:**
06-10-2023	Використовувались скомпрометовані електронні листи для поширення шкідливого програмного забезпечення SmokeLoader.

Продовження таблиці 2.1

Сектор: “Державне управління”	
Дата:	**Короткий опис атаки:**
19-01-2022	Зловмисники використовували платформу для пошуку роботи як канал для доставки шкідливого програмного забезпечення. Вони завантажили шкідливу програму у вигляді резюме та подали заявку до західної урядової організації.
01-02-2022	Документ Word, прикріплений до фішингового листа, ймовірно був націлений на Міністерство закордонних справ України. У файлі містився VBS-скрипт, який виконував завантаження й встановлення шкідливого програмного забезпечення для закріплення на пристрої жертви.
07-03-2022	Кампанія фішингу, націлена на урядові установи України, з використанням шкідливого програмного забезпечення "MicroBackdoor".
17-03-2022	Розсилка електронних листів, що містять шкідливі файли, націлена на урядові та військові установи України. У результаті атаки комп'ютер цільової особи заражався шкідливим програмним забезпеченням SPECTR.
18-03-2022	Фішингові кампанії, спрямовані на українські організації, що поширювали бекдор LoadEdge.
23-03-2022	Фішингові атаки, зі спробою видати себе за Міністерство оборони України.
27-03-2022	Фішингова кампанія, спрямована на українські організації, з використанням шкідливого ПЗ GRIMPLANT і GRAPHSTEEL.
02-04-2022	Група українських державних службовців отримала фішингові сповіщення через Telegram, з закликом перевірити безпеку їхніх акаунтів, оскільки було помічено несанкціонований вхід до їх акаунтів з боку Росії. Після кліку на шкідливе посилання акаунти були скомпрометовані.

Продовження таблиці 2.1

26-04-2022	Фішингова кампанія, яка використовувала скомпрометований акаунт співробітника державного органу України для розповсюдження шкідливих програм "GraphSteel" і "GrimPlant".
28-04-2022	Кампанія цілеспрямованого фішингу проти державних органів України використовувала електронні листи з темою «Указ прес-служби Європейського Союзу № 576/22 про безперервні заходи безпеки» для доставки ISO-образу. В результаті пристрої були інфіковані шкідливим програмним забезпеченням "Meterpreter".
09-05-2022	Одному з державних органів України надійшов таргетований фішинговий лист на одну з їхніх загальних електронних адрес. Лист містив посилання на шкідливий файл, мета якого — збір інформації, яка може бути використана для компрометації цілі та отримання подальшого доступу до їхньої мережі.
02-06-2022	Шкідливий файл, який розповсюджував "Cobalt Strike Beacon". Файл поширювався серед українських державних організацій.
01-07-2022	Надсилаючи шкідливі файли у фішингових електронних листах та повідомленнях, зловмисник завантажував шкідливе програмне забезпечення LONEPAGE, що компрометувало пристрій цілі.
11-07-2022	Розповсюдження електронних листів з темою "Спільний офіційний звіт щодо гуманітарної ситуації. Україна" та вкладеним XLS документом "Гуманітарна катастрофа в Україні 24 лютого 2022 року" призвело до компрометації українських державних органів і впливу на системи, що були заражені програмою Cobalt Strike Beacon.
20-07-2022	CERT-UA виявив шкідливий файл, що розповсюджував шкідливе ПЗ для крадіжки даних "AgentTesla". Зважаючи на назву початкового файлу "Report_050722_4.ppt" та зміст презентації в PPT, CERT-UA вважає, що атака була спрямована на українські державні організації.

Продовження таблиці 2.1

21-10-2022	Державні організації України стали мішенню для кібератак за допомогою ШПЗ RomCom. Вектор атаки — фішинг-лист, який нібито надійшов від Прес-служби Генерального штабу Збройних Сил України.
06-02-2023	Масові фішингові листи, що розповсюджували шкідливе ПЗ Remcos і націлені на державні органи України.
24-02-2023	Фішингова кампанія, спрямована на корпоративну електронну адресу української урядової організації, що включала в себе листи з попередженням про закінчення дії паролю. Лист містив шкідливий вкладений файл, який спонукав користувачів ввести свої облікові дані, після чого ті були скомпрометовані.
01-04-2023	Цільові користувачі отримали шкідливий лист з темою "Windows Update" від адреси @outlook.com, що містила ім'я та ініціали реальної особи. Лист містив інструкції для запуску PowerShell-команди, яка завантажувала скрипт PowerShell, що дозволяло зловмиснику викрасти дані, використовуючи команди "tasklist" та "systeminfo". Крім того, скрипт міг відправляти отриману інформацію через HTTP-запит до API сервісу Mocky.
18-04-2023	Фішинг-електронні листи, які нібито походили від Посольства Таджикистану в Україні. Листи містили шкідливі файли, такі як LOGPIE keylogger, CHERRYSPY Backdoor та STILLARCH malware які могли скомпрометувати пристрій жертви.
12-05-2023	Цілеспрямований фішинг, який використовував відомі вразливості в Roundcube, що призводило до негайної компрометації пристрою цілі, як тільки лист був відкритий. Таким чином, зловмисник перенаправляв майбутні вхідні листи цілі на контрольовану адресу, що дало можливість викрасти дані з вхідних листів. Більше сорока українських організацій стали мішенями для цих атак.

Кінець таблиці 2.1

07-06-2023	Зловмисник створив фішингові вебсайти, які використовувалися для поширення троянської версії Devolution Remote Desktop Manager. Після завантаження, зловмисник отримував доступ до пристрою цілі та її даних.
07-07-2023	За допомогою фішингу зловмисник поширює шкідливе ПЗ PicassoLoader. Після його розгортання, це ПЗ завантажує та запускає утиліту віддаленого доступу njRAT, що дає зловмиснику доступ до пристрою цілі та можливість поширюватися по мережі пристрою.
15-12-2023	Цілям надсилалися електронні листи з шкідливими посиланнями, що при відкритті призводили до завантаження шкідливих програм, таких як MASEPIE, OCEANMAP і STEELHOOK.

Відповідно до створеної статистики, можна зробити висновки що найбільше від атак соціальної інженерії страждає сектор державного управління.

2.2 OSINT як інструмент протидії атакам соціальної інженерії

Реальні приклади, розглянуті в першому розділі, демонструють, як зловмисники використовують відкриті джерела для збору інформації і як ці дані стають основою для атак. Однак ці ж методи можна використовувати і для забезпечення безпеки організації. Використання інструментів OSINT надає можливість виявляти чутливу інформацію у відкритих джерелах ще до того, як її зможуть використати зловмисники. Знайдена інформація стає підґрунтям для оцінки відкритості організації. Для більш якісного аналізу, варто розуміти яка конкретно інформація цікавить соціального інженера.

Основне, що цікавить такого зловмисника, це ті дані, які дадуть йому можливість легко увійти в довіру, підробити щось офіційне, або зрозуміти слабкі місця в організації чи конкретної людини [18].

Перше, те як компанія використовує інтернет. Аналіз цифрової присутності компанії дозволяє встановити, які саме онлайн-сервіси, веб-сайти чи API

використовуються для комунікації з клієнтами або партнерами. Додатково, через дорки можна знайти доменну інформацію (назви доменів, субдоменів, корпоративні електронні адреси), що полегшує моделювання фішингових атак або створення фейкових сайтів.

Друге, чи є в компанії соц-мережі, та як вона використовує їх, оскільки вони можуть містити як прямі витоки (наприклад, фото з робочих місць, скріни систем, опис інструментів), так і непрямі — згадки про внутрішні процеси, святкування, партнерів або нові впровадження. З соціальних мереж зловмисники також можуть отримати контактну інформацію співробітників (номери телефонів, особисті та службові email), що спрощує реалізацію фішингових атак.

Також, соц інженеру цікаво чи є в корпорації політики щодо розміщення інформації її працівниками в Інтернеті. Окрім ПБ соц інженера цікавлять й інші файли внутрішнього використання, такі як службові документи, інструкції, політики безпеки. Сюди ж відноситься і використання особистих пристроїв (BYOD). Наявність або відсутність контролю за BYOD-пристроями працівників безпосередньо впливає на площу потенційної атаки.

Чутливою є інформація про постачальників, яких використовує організація і їх кількість. Інформація про партнерів і клієнтів (контракти, умови співпраці, комерційні пропозиції) може бути знайдена через витоки документів або публічні згадки, і використана для таргетованих атак. Як організація приймає платежі та проводить виплати теж може бути використано у схемах фінансового шахрайства, зокрема у фішингових кампаніях. При цьому, витоки фінансової інформації (рахунки, накладні, платіжні дані) значно полегшують розробку реалістичних сценаріїв атак.

Соціального інженера також цікавить чи має організація кол-центри і їх географічне розташування. Інформація про наявність або аутсорсинг кол-центрів дає можливість створити достовірний вектор атаки, орієнтований на голосову інженерію, особливо у випадках, коли працівники менш обізнані з кіберзагрозами. Географія діяльності організації дозволяє визначити часові

пояси, мову спілкування, правові зони та регіональні особливості атак. Також це дає змогу більш точно підробляти дзвінки або листи. У випадку, якщо компанія децентралізована з великою кількістю офісів, вона може мати не уніфіковану політики безпеки для кожного філіалу.

Іншим вектором для пошуку є організаційна структура компанії. Публічна наявність ієрархії компанії або списку її співробітників, полегшує моделювання атак типу whale spear phishing. Деталізовані дані про організаційну структуру (імена керівників, відповідальних осіб, ієрархія підрозділів) можуть бути знайдені через витоки документів або відкриті профілі працівників у соцмережах.

Крім того, можна виявити інформацію про внутрішні системи організації (VPN, внутрішні портали, CRM, ERP-системи) та файли внутрішнього використання (службові документи, інструкції, політики безпеки). Інколи через витоки у відкритому доступі можуть опинитися навіть облікові дані (логіни, паролі, токени доступу) або технічні метадані (версії ПЗ, конфігурації серверів, IP-адреси), що значно спрощує підготовку технічної атаки.

2.3 Побудова дорків

Пошук даних буде здійснюватись за допомогою дорків [19] у пошуковій системі Google, що є найпопулярнішою і найбільш індексованою пошуковою системою і дає змогу охопити якнайбільше даних. Використання пошукових операторів при цьому значно спрощує задачу і звужує область пошуку, а також дозволяє знайти те, що звичайні запити видати не можуть.

Таблиця 2.2 - Дані, по яким буде проводитись пошук

Вектор інтересу:	**Чутливі дані:**
Соціальні мережі	Мова та стиль спілкування
	Фото офісів
	Імена співробітників

Продовження таблиці 2.2

	Керівники відділів
	Імена розробників
Політика безпеки та внутрішні документи, зокрема політика щодо BYOD	Політика безпеки (пункт про BYOD)
	Шаблони договорів
	Внутрішні звіти
	Назви внутрішніх проектів
Список постачальників	Назви постачальників
	Контакти постачальників
Інструкції для платежів та виплат	Реквізити для оплат
	Процедури платежів
	Критерії підтвердження
	Формати рахунків
Наявність кол-центрів	Геолокація кол-центрів
Розташування офісів та філій	Адреси офісів
	Кількість офісів
Організаційна структура компанії	Взаємозв’язки між відділами і їх кількість
Відкриті репозиторії	Конфігураційні файли
	Журнали змін
Контактна інформація співробітників	Робочі електронні адреси
	Номера телефонів
Облікові дані	Логіни
	Паролі
	Токени доступу

Кінець таблиці 2.2

Доменна інформація	Піддомени
	Сервіси прив’язані до доменів
Інформація про внутрішні системи	Сервери пошти
	CRM/ERP-доступ
	Конфігурації серверів
	IP-адреси корпоративних серверів
Плани заходів або розклади	Графік роботи компанії
	Дати корпоративів
	Дати конференції
Технічні метадані	Тип використовуваного ПЗ
	Версії софту
	Технічні ролі пристроїв

Список чутливих даних був сформований на основі їхньої практичної цінності для реалізації атак соціальної інженерії. Це дозволяє обмежити обсяг дослідження тими елементами, що мають найвищу ймовірність бути використаними для компрометації організації.

Побудовані дорки виглядають так:

1. Соціальні мережі:

site:linkedin.com/in "$organization"

site:linkedin.com/company "$organization"

site:facebook.com "$organization"

(site:twitter.com OR site:x.com) "$organization"

site:instagram.com "$organization"

Більш деталізовані запити для пошуку конкретної посади співробітника:

site:linkedin.com/in "$organization" ("Team lead" OR "CTO" OR "Developer" OR "HR" OR "Manager")

site:linkedin.com/in "$organization"

Для пошуку фото офісів:

"офіс $organization" (site:linkedin.com OR site:behance.net OR site:pinterest.com OR site:instagram.com OR site:facebook.com)

2. Політика безпеки та внутрішні документи, зокрема політика щодо BYOD:

"$organization" ("політика" OR "політика безпеки" OR "BYOD" OR "bring your own device" OR "internal security policy" OR "внутрішні правила") (filetype:pdf OR filetype:doc OR filetype:docx)

"$organization" ("політика безпеки" OR "BYOD" OR "bring your own device" OR "internal security policy" OR "внутрішні правила")

Для шаблонів договорів:

"$organization" ("договір" OR"шаблон договору" OR "template contract" OR "agreement template" OR "договір з постачальником" OR"договір на надання послуг" OR "contract template") (filetype:doc OR filetype:docx OR filetype:pdf)

Для внутрішніх звітів:

"$organization" ("звіт" OR "report" OR "security report" OR "audit report" OR "звіт з безпеки") (filetype:doc OR filetype:docx OR filetype:pdf)

Назви внутрішніх проектів:

"$organization" ("проект" OR "project" OR "initiative" OR "roadmap" OR "road map") (filetype:doc OR filetype:docx OR filetype:pdf)

3. **Список постачальників:**

"$organization" ("постачальник" OR "постачальники" OR "supplier" OR "vendor"

OR "suppliers" OR "approved vendors" OR "контакти постачальників" OR "контакт постачальника" OR "supplier contact list") (filetype:pdf OR filetype:doc OR filetype:xlsx OR filetype:xls)

4. **Інструкції для платежів та виплат:**

site:$organization.com ("оплата" OR "реквізити" OR "інструкція з оплати" OR "payment instructions" OR "bank details" OR "invoice")

"$organization" ("оплата" OR "реквізити" OR "інструкція з оплати" OR "payment instructions" OR "bank details" OR "invoice") (filetype:pdf OR filetype:doc OR filetype:docx OR filetype:xls OR filetype:xlsx)

Процедури платежів:

"$organization" ("payment process" OR "порядок оплати" OR "внутрішня інструкція з платежів") (filetype:pdf OR filetype:doc OR filetype:docx OR filetype:xls OR filetype:xlsx)

Критерії підтвердження та формати рахунків::

"$organization" ("payment approval" OR "затвердження платежу" OR "ліміт оплати" OR "confirmation of payment" OR "transaction confirmation" OR "підтвердження платежу" OR "підтвердження транзакції" OR "документ на оплату" OR "документ про оплату" OR "платіжне доручення" OR "розрахунковий документ" OR "рахунок-фактура") (filetype:pdf OR filetype:doc OR filetype:docx OR filetype:xls OR filetype:xlsx)

5. Наявність кол-центрів:

"$organization" ("call center" OR "кол-центр" OR "контакт-центр") ("address" OR "location" OR "адреса" OR "локація")

6. **Розташування офісів та філій:**

("$organization" OR site:$organization.com) ("адреси офісів" OR "розташування філій" OR "офіс" OR "офіси" OR "locations" OR "приміщення" OR "відділення")

7. **Організаційна структура компанії**

"$organization" ("структура компанії" OR "наша команда" OR "внутрішня структура" OR "структура відділів" OR "відділи компанії" OR "організаційна структура" OR "company structure" OR "organizational chart") (filetype:pdf OR filetype:doc OR filetype:ppt OR filetype:docx OR filetype:xls OR filetype:xlsx)

"$organization" (site:$organization.com OR site:linkedin.com)

("структура компанії" OR "наша команда" OR "внутрішня структура" OR "структура відділів" OR "відділи компанії" OR "організаційна структура" OR "company structure" OR "organizational chart")

8. **Відкриті репозиторії**

Конфігураційні файли:

"$organization" (site:github.com OR site:gitlab.com) (filetype:env OR filetype:json OR filetype:yml OR filetype:py OR filetype:properties OR filetype:xml OR filetype:ini OR filetype:conf)

Журнали змін:

"$organization" (site:github.com OR site:gitlab.com) (filename:CHANGELOG.md OR filename:changelog.txt)

"$organization" (site:github.com OR site:gitlab.com) ("initial commit" OR "fix bug" OR "bug fix" OR "update config" OR "commit history")

9. **Контактна інформація співробітників:**

Робочі електронні адреси:

"$organization" (site:linkedin.com/in OR site:$organization.com OR site:pastebin.com) (filetype:txt OR filetype:pdf OR filetype:xlsx OR filetype:xls OR filetype:doc OR password filetype:docx OR filetype:json OR filetype:env OR filetype:yaml OR filetype:xml)

"$organization" "@organization.com" (filetype:txt OR filetype:pdf OR filetype:xlsx OR filetype:xls OR filetype:doc OR password filetype:docx OR filetype:json OR filetype:env OR filetype:yaml OR filetype:xml)

Номера телефонів:

"$organization" ("контактний номер" OR "телефон" OR "контактна особа"OR "тел:" OR "phone:" OR "+38")

"$organization" (site:$organization.com OR site:linkedin.com/in) ("контактний номер" OR "телефон" OR "контактна особа"OR "тел:" OR "phone:" OR "+38")

10. **Облікові дані**

Загальний запит:

"$organization" ("password" OR "login" OR "username" OR "пароль" OR "логін" OR "токен" OR "credentials" OR "authentication") (filetype:txt OR filetype:pdf OR filetype:xlsx OR filetype:xls OR filetype:doc OR filetype:docx OR filetype:json OR filetype:env OR filetype:yaml OR filetype:xml)

"$organization" (site:github.com OR site:gitlab.com OR site:bitbucket.org OR site:pastebin.com) (password OR token OR api_key OR secret_key OR authorization OR "DB_PASSWORD" OR "API_KEY" OR "SECRET_KEY")

11. Доменна інформація:

Піддомени:

site:.$organization.com -www.$organization.com*

Сервіси прив’язані до доменів:

site:.$organization.com (inurl:login OR inurl:admin OR inurl:login)*

12. Інформація про внутрішні системи:

Сервери пошти:

site:$organization.com ("mail server" OR intitle:"mail" OR inurl:mail OR "smtp" OR "imap" OR "pop3")

CRM/ERP-доступ:

site:$organization.com (inurl:crm OR inurl:erp OR inurl:"dashboard" OR inurl:admin OR inurl:portal OR inurl:user)

Конфігурації серверів

"$organization" "server" (filename:.env OR filename:config.json OR filename:config.yml OR filename:docker-compose.yml OR filename:settings.py OR filename:application.properties)

"$organization" "server" (filetype:env OR filetype:json OR filetype:yml OR filetype:py OR filetype:properties OR filetype:xml OR filetype:ini OR filetype:conf)

13. Плани заходів або розклади:

"$organization" ("план заходів" OR "розклад заходів" OR "event schedule") (filetype:xlsx OR filetype:pdf)

Графік роботи компанії:

"$organization" ("графік роботи" OR "робочий графік" OR "режим роботи")

Дати корпоративів:

"$organization" ("корпоратив" OR "party")

Дати конференції:

"$organization" ("конференція" OR "виступ" OR "подія" OR "conference" OR "summit" OR "конференція 2025")

14. Технічні метадані:

Після проведення пошуку файлів, потрібно буде за допомогою сторонньої програми проаналізувати метадані в знайдених файлах:

"site:$organization.com" (filetype:txt OR filetype:pdf OR filetype:xlsx OR filetype:xls OR filetype:doc OR password filetype:docx OR filetype:json OR filetype:env OR filetype:yaml OR filetype:xml)

Окремо варто вказати оператор *after: YYYY-MM-DD* який фільтрує результати проіндексовані після конкретної дати. Це може бути корисним при перевірці актуальності інформації.

2.4 Критерії для оцінки серйозності витоку

Для оцінки серйозності витоку інформації запропоновано систему зваженої бальної оцінки, що враховує критичність та актуальність знайдених даних. Запропонована модель допомагає оцінити вразливість даних організації доступних у відкритих джерелах.

Кожен тип знайдених даних має свою вагу серйозності, залежно від того, як саме соц інженер може використати їх. Вага кожного елементу визначається відповідно до серйозності атак, які можуть бути реалізовані з їх допомогою.

Важливість кожного критерію визначається шляхом попарних порівнянь і створенням матриці Сааті:

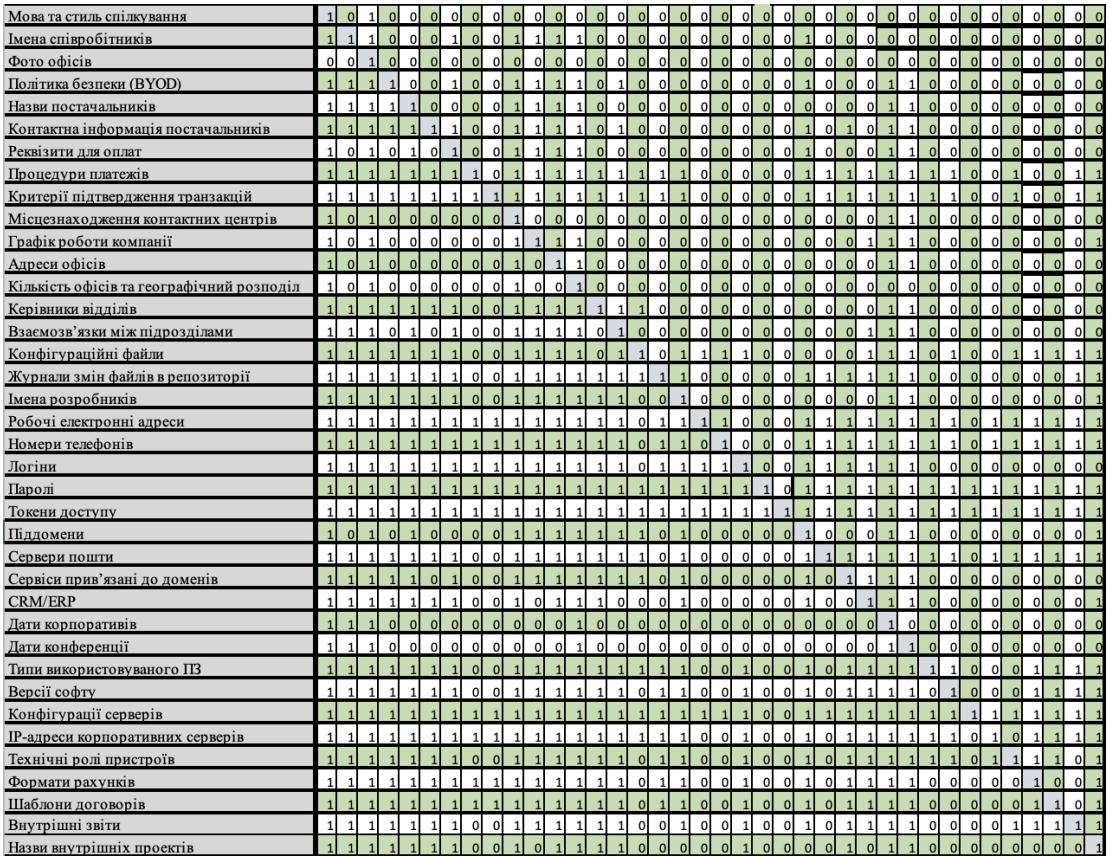

Рисунок 2.1 - Обрахунки матриці Сааті

Таблиця 2.3 - Фінальні обрахунки ваги

Критерій:	**Вага (ω):**
Мова та стиль спілкування	0,0027
Фото офісів	0,0013
Імена співробітників	0,0121
Керівники відділів	0,0216
Імена розробників	0,0229
Політика безпеки (BYOD)	0,0175
Шаблони договорів	0,0351
Внутрішні звіти	0,0337

Продовження таблиці 2.3

Назви внутрішніх проектів	0,0229
Назви постачальників	0,0148
Контактна інформація постачальників	0,0216
Реквізити для оплат	0,0148
Процедури платежів	0,0378
Критерії підтвердження транзакцій	0,0391
Формати рахунків	0,0337
Місцезнаходження контактних центрів	0,0067
Адреси офісів	0,0094
Кількість офісів та географічний розподіл	0,0054
Взаємозв’язки між підрозділами	0,0175
Конфігураційні файли	0,0351
Журнали змін файлів в репозиторії	0,0324
Робочі електронні адреси	0,0445
Номери телефонів	0,0432
Логіни	0,0351
Паролі	0,0499
Токени доступу	0,0513
Піддомени	0,0202
Сервіси прив’язані до доменів	0,0256
Сервери пошти	0,0391
CRM/ERP-доступ	0,0229
Конфігурації серверів	0,0486
IP-адреси корпоративних серверів	0,0418

Кінець таблиці 2.3

Графік роботи компанії	0,0135
Дати корпоративів	0,0067
Дати конференції	0,0081
Типи використовуваного ПЗ	0,0378
Версії софту	0,0351
Технічні ролі пристроїв	0,0378

Ваги нормалізовані так, щоб:

$$\sum_{i=1}^{n} \omega_i = 1 \qquad (2.1)$$

де:

n – кількість усіх типів даних;

ω_i – вага і-того елементу.

Після проведення пошуку чутливої інформації за допомогою відповідних дорків, на основі знайденої інформації експерти виставляють свою оцінку. Для цього використовуються дві ключові характеристики: наявність та актуальність. Наявність інформації — визначається як наявність або відсутність конкретних даних, що можуть бути використані зловмисниками для проведення атак. Актуальність інформації — визначає, наскільки отримані дані є релевантними на поточний момент для проведення атаки або здійснення інших дій. Щоб об'єктивно оцінити ці характеристики, кожна з них позначається оцінкою 0 або 1, де 1 означає наявність або актуальність інформації, а 0 — її відсутність або неактуальність. Завдяки цьому стає легше зрозуміти, наскільки знайдена інформація дійсно може становити загрозу. Також, важливо зазначити, що оцінка актуальності не виставляється окремо від наявності: якщо інформація не виявлена, то її актуальність автоматично вважається нульовою, оскільки оцінювати релевантність відсутніх даних неможливо.

Для більш точного визначення важливості кожного з критеріїв вводяться вагові коефіцієнти. Вага наявності визначається як коефіцієнт, що відображає важливість наявності даних для конкретної задачі, і приймає значення від 0 до 1. Вага актуальності показує ступінь впливу актуальності інформації на загальну оцінку її цінності в контексті потенційного використання в атаках, і теж приймає значення від 0 до 1. З практичної точки зору, навіть якщо певна інформація присутня у відкритому доступі, її цінність для зловмисника значно знижується, якщо вона втратила актуальність. Саме тому в критерію актуальності вага буде вище, ніж в наявності, однак це не означає що застарілі дані не надають жодної інформації зловмисникам.

Оцінка експертів розраховується за формулою:

$$q = \omega_a * q_{ai} + \omega_r * q_{ri} \quad (2.2)$$

де:

ω_a, ω_r – ваги наявності та актуальності відповідно;

q_a – експертна оцінка наявності типу даних;

q_r – експертна оцінка актуальності типу даних;

Остаточна оцінка відкритості організації розраховується за формулою:

$$Q = \sum_{i=1}^{n} \omega_i * (\omega_a * q_{ai} + \omega_r * q_{ri}) * 100\% \quad (2.3)$$

де:

n – кількість усіх типів даних;

ω_i – вага і-того елементу;

ω_a, ω_r – ваги наявності та актуальності відповідно;

q_{ai} – факт наявності і-го типу даних;

q_{ri} – експертна оцінка актуальності типу даних;

Знайдений результат розшифровується відповідно до запропонованих категорій:

Таблиця 2.4 - Запропоновані категорії для оцінки рівня відкритості

Відсоток:	Рівень відкритості:
81-100%	Високий (В)
61-80%	Вище середнього (ВС)
41-60%	Середній (С)
21-40%	Нижче середнього (НС)
0-20%	Низький (Н)

2.5 Методика у вигляді алгоритму на псевдокоді

Вхід: Список типів даних data_types, кожен з яких має назву та вагу ω_i

Вихід: Рівень відкритості організації та відсоток відкритості Q

wa ← вага наявності даних

wr ← вага актуальності даних

```
for each data_type in data_types do:
    found = search_data(data_type.name)
    if found:
        qa = 1
        qr = assess_relevance(data_type.name)
    else:
        qa = 0
        qr = 0
    data_type.score = (a * qa) + (r * qr)
Q = 0
for each data_type in data_types:
    Q += data_type.weight * data_type.score
```

```
Q = Q * 100%
#Інтерпретація результату
if Q >= 81:
   openness_level = "Високий (В)"
else if Q >= 61:
   openness_level = "Вище середнього (ВС)"
else if Q >= 41:
   openness_level = "Середній (С)"
else if Q >= 21:
   openness_level = "Нижче середнього (НС)"
else:
   openness_level = "Низький (Н)"

#Виведення фінального результату
print("Рівень відкритості організації:", openness_level)
print("Відсоток відкритості:", Q)
```

Висновки до Розділу 2

Запропоновано методику виявлення та оцінювання критичності витоків даних щодо об'єктів критичної інфраструктури. Детально проаналізовано дані, які є предметом пошуку соц інженера у відкритих джерелах і розроблено список який є основою для реалізації відповідних дорків. Крім цього проаналізовано чутливі дані, які представляють цінність з точки зору атак соціальної інженерії, і оцінено їх шляхом попарних порівнянь експертів. Для оцінювання було враховано актуальність, наявність та потенційна цінність знайденої інформації для зловмисника. На базі цього було розраховано вагу кожного типу даних і запропоновано формулу для фінальної оцінки відкритості організації.

3 Практичний експеримент

3.1 Практичне застосування методики на тестовій організації

Для даного дослідження і тестування методики під час проведення експерименту підійдуть компанії з високим рівнем онлайн присутності. Для цього досліду було обрано українську компанію, яка працює в секторі транспортних перевезень. Вхідними даними будуть:

- назва компанії, українською та англійською мовами;
- домен;
- приклад робочої пошти.

Тобто перед проведенням пошуку в дорках потрібно замінити $organization на назву компанії, і окремо провести пошук для української і англійської назви. @organization потрібно замінити на пошту компанії. Те саме стосується домену $organization.com.

Після визначення вхідних даних, провели пошук за допомогою дорків. Результати пошуку оформлено в таблиці нижче, де також вказано факт наявності конкретного типу даних і його актуальності:

Таблиця 3.1 - Результати пошуку під час експерименту

	Наявність	Актуальність
Мова та стиль спілкування	1	1
Фото офісів	1	1
Імена співробітників	1	1
Керівники відділів	1	1
Імена розробників	1	1
Політика безпеки (BYOD)	1	0
Шаблони договорів	1	1
Внутрішні звіти	1	1
Назви внутрішніх проектів	1	0
Назви постачальників	1	1

Кінець таблиці 3.1

Контактна інформація постачальників	1	1
Реквізити для оплат	1	0
Процедури платежів	1	1
Критерії підтвердження транзакцій	1	1
Формати рахунків	1	0
Місцезнаходження контактних центрів	0	0
Адреси офісів	1	1
Кількість офісів	1	0
Взаємозв’язки між підрозділами	0	0
Конфігураційні файли	0	0
Журнали змін файлів в репозиторії	0	0
Робочі електронні адреси	1	0
Номери телефонів	1	0
Логіни	0	0
Паролі	0	0
Токени доступу	0	0
Піддомени	1	1
Сервіси прив’язані до доменів	1	1
Сервери пошти	0	0
CRM/ERP	1	0
Конфігурації серверів	0	0
IP-адреси корпоративних серверів	0	0
Графік роботи компанії	1	1
Дати корпоративів	0	0
Дати конференції	1	0
Типи використовуваного ПЗ	0	0
Версії софту	0	0
Технічні ролі пристроїв	0	0

Знайдена інформація була оцінена на релевантність відповідно до того, наскільки її можна використати при побудові атаки соціальної інженерії. Враховувалось також те, наскільки дані були свіжими, і крім цього враховувалась кількість даних. Якщо дані були занадто поверхневими і не відкривали конфіденційні внутрішні процеси, актуальність цих даних оцінювалась як 0.

Для більш зручного розрахунку фінального індексу відкритості організації було реалізовано скрипт на версії Python 3.11.3, який наведено в додатку А. Код підтягує дані вагів, факту наявності і оцінку актуальності з відповідних json файлів і за допомогою функції calculate_score розраховує фінальну оцінки відкритості організації. Розроблена за допомогою інструменту Visual Studio Code, програма зчитує наперед задані конфігураційні файли і видає результат в консоль.

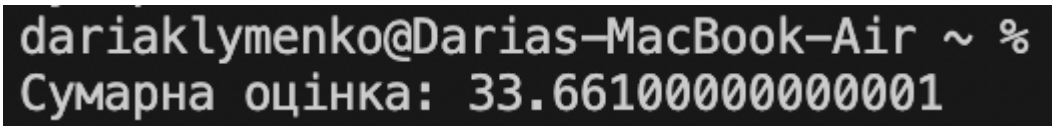

Рисунок 3.1 - Результат експерименту в консолі

За формулою (2.3), було пораховано індекс відкритості організації, що складає 33.661%, і за запропонованою шкалою означає рівень відкритості нижче середнього.

3.2 Аналіз знайденої інформації

Зібрана за допомогою OSINT інформація дозволяє не лише оцінити рівень відкритості організації, а й зіставити потенційні вектори атаки з реальними сценаріями, класифікованими за MITRE ATT&CK [20]. Загально, пошук чутливої інформації за допомогою дорків підпадає під техніку T1593.002: Search Open Websites/Domains: Search Engines. Додатково можна зазначити техніку пошуку інформації T1593.001: Search Open Websites/Domains: Social Media. Обидві

техніки спрямовані на збір чутливих даних, які підвищують шанси успішної реалізації наступних атак. Приклади таких атак будуть розглянуті нижче.

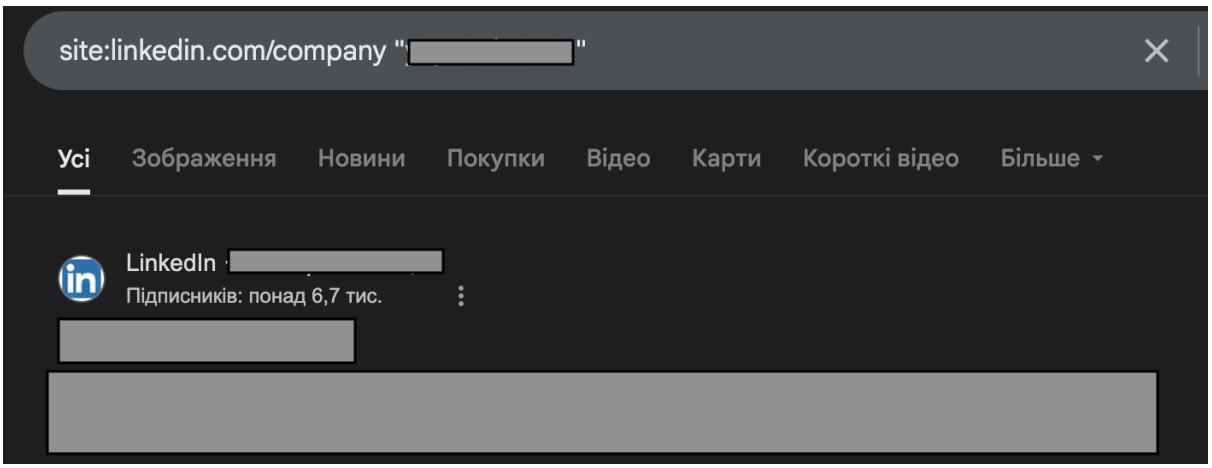

Рисунок 3.2 - Результат пошуку компанії в LinkedIn

Основним джерелом даних типу імен співробітників, розробників або керівників відділу стали соц мережі, а якщо конкретніше то LinkedIn. Самі імена вже є цінним джерелом для реалізації атаки соціального інженера. Вони розширюють пошук на особисті сторінки співробітників в соціальних мережах, і при більш детальному дослідженні, можна створити досьє на конкретну людину, дослідити її рівень інтернет гігієни. Знаючи слабку ланку в співробітниках легше давити саме туди, тоді вистачить і поверхневих даних для успішної атаки. За технікою T1589.003 класифікованій MITRE ATT&CK як Gather Victim Identity Information: Employee Names, така інформація може бути корисна для отримання електронних адрес, особливо коли є шаблон корпоративної пошти:

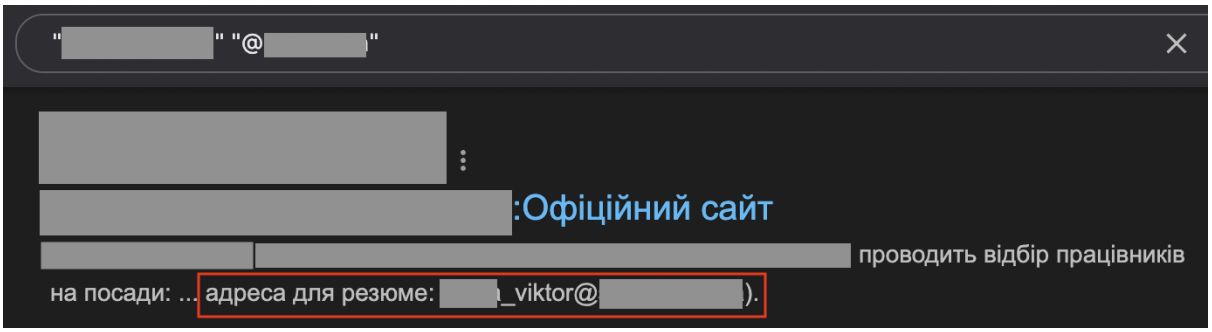

Рисунок 3.3 - Результат пошуку шаблону електронної пошти

В такому випадку компанія мало що може зробити для того щоб попередити збирання цих даних. Основною рекомендацією є мінімізація обсягу чутливих даних доступних в мережі.

При використанні адреси в парі з ім’ям співробітника можливе реалізування атаки цілеспрямованого фішингу: T1566.001 - Phishing: Spearphishing Attachment. В такому випадку рекомендаціями є обмеження доступу до сервісів пошти компанії тим співробітникам, яким це не потрібно для виконання робочих обов’язків, а також використання антивірусів і моніторинг журналів аудиту.

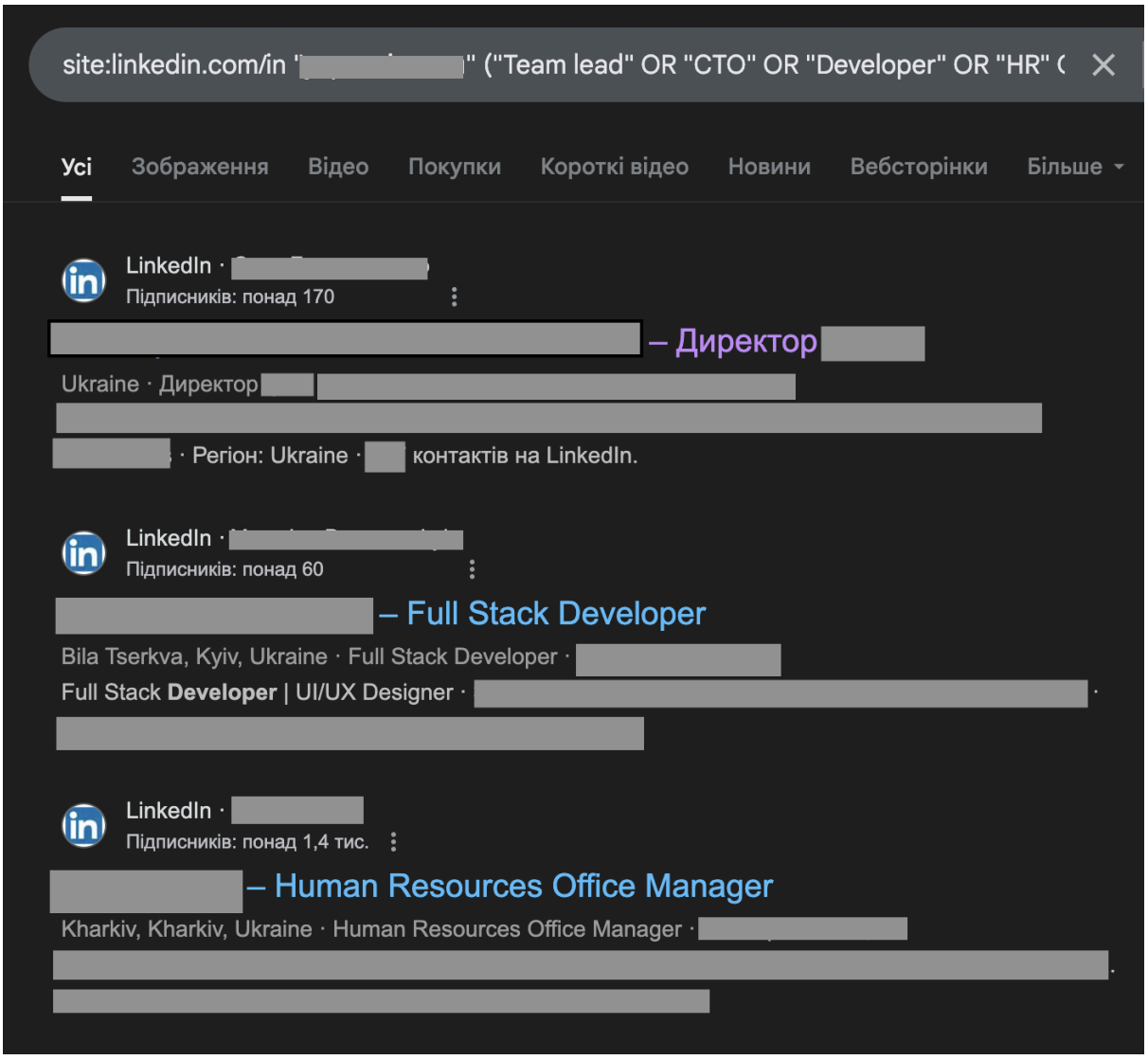

Рисунок 3.4 - Результат пошуку керівників або розробників в LinkedIn

За допомогою дорків не вийшло знайти файл зі структурою компанії, але її легко можна побудувати, знаючи хто яку посаду займає. Особливо цікавими є керівники відділів, які зазвичай мають більше доступу до конфіденційних даних ніж звичайні працівники. Техніка яка класифікує такі атаки має назву T1591.004 Gather Victim Org Information: Identify Roles і передбачає збір інформації про співробітників з привілеями для подальшого використання цієї інформації в атаках.

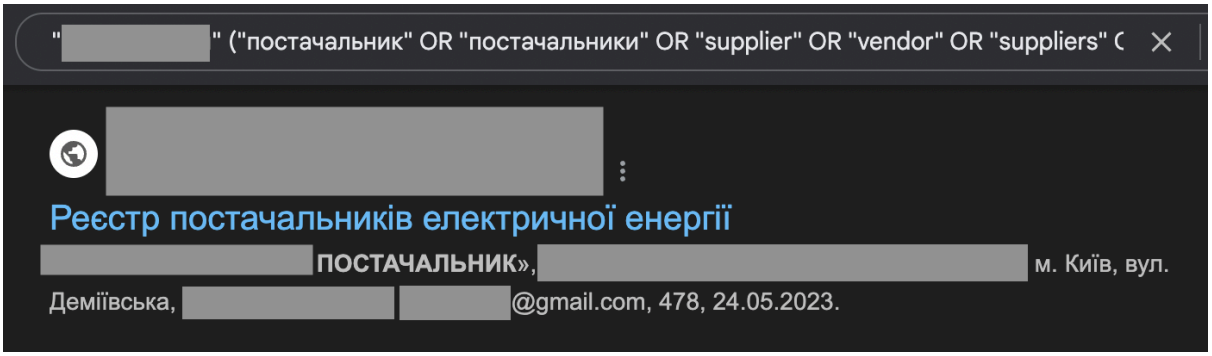

Рисунок 3.5 - Результат пошуку постачальників та їх контактів

Пошук постачальників має свою класифіковану техніку MITRE ATT&CK, яка називається T1591.002 Gather Victim Org Information: Business Relationships. Зазвичай, якщо вдається знайти список постачальників, це відкриває дорогу до атаки на ланцюжок постачальників T1195: Supply Chain Compromise. Проте в цьому експерименті ситуація зовсім інша, і можливості які відкриваються для соціального інженера при цьому теж інші.

Знаючи контактні дані постачальників (а також співробітників), можна використати атаку претекстингу, T1656: Impersonation, щоб представитись довірчою особою для отримання доступу до конфіденційної інформації. Для того щоб попередити такі атаки, варто навчати співробітників відповідним патернам поведінки, наприклад дзвінок підтвердження особи, щоб упевнитися в справжності представника постачальника.

Іншим вектором атаки може бути безпосередньо постачальник. Якщо після аналізу списку постачальників з’ясовується, що якийсь з них має поганий захист, легше спрямувати зусилля на його злам, і тоді застосувати техніку T1199: Trusted

Relationship. Основа цієї техніки в наявності довірчих стосунків між жертвою і майбутньою потенційною жертвою, що полегшує задачу створення правдоподібної легенди для атак соціальної інженерії, бо в цьому більше немає необхідності. Щоб уникнути наслідків цієї атаки, потрібно ввести двофакторну аутентифікацію для всіх делегованих корпоративних акаунтів і обов'язково сегментувати мережу, задля зменшення можливих збитків і уникнення компрометації чутливих даних.

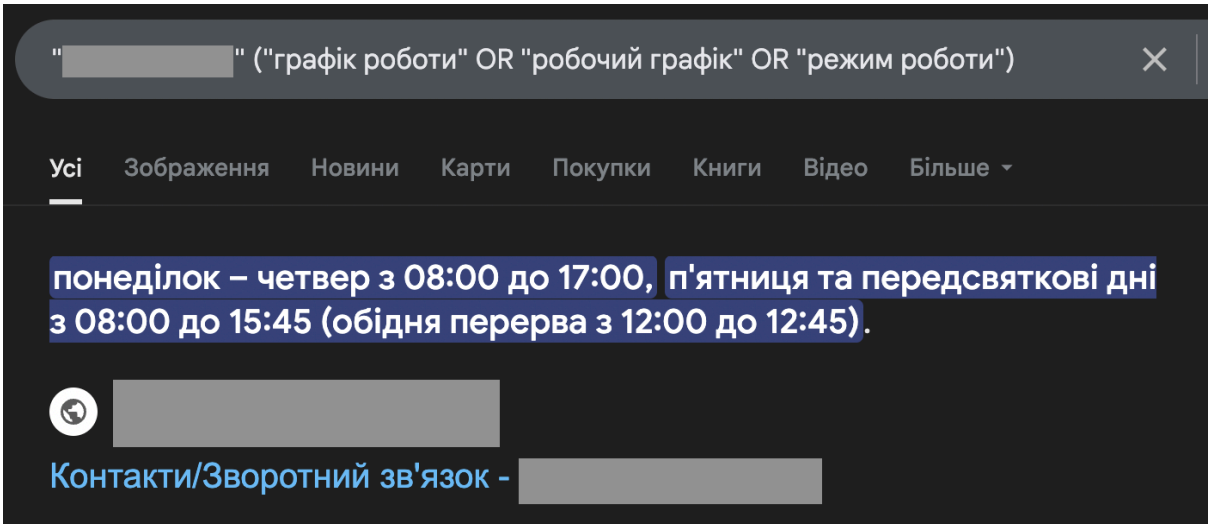

Рисунок 3.6 - Результат пошуку графіку роботи

Підготовкою до майбутньої атаки можуть бути спроби визначити геолокацію головного офісу, філій чи ключових ресурсів компанії, що має класифікацію T1591.001 Gather Victim Org Information: Determine Physical Locations. З цієї ж техніки збору даних для майбутніх атак є підтехніка T1591.003: Gather Victim Org Information: Identify Business Tempo, яка націлена на графік роботи компанії. Всі ці дані підвищують імовірність успішної атаки фішингу і додають правдоподібності легенді, наприклад якщо створити запит для термінової роботи.

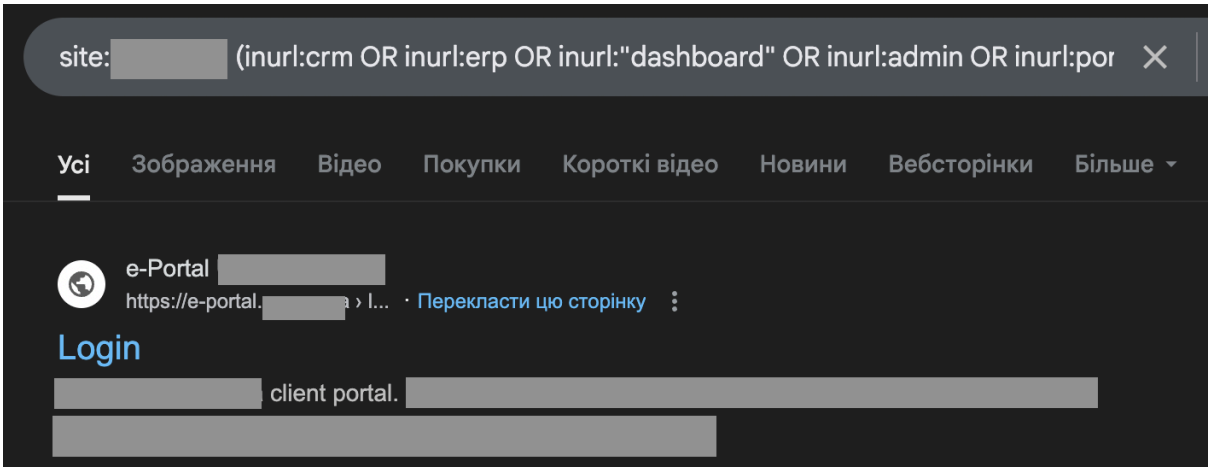

Рисунок 3.7 - Результат пошуку CRM/ERP

Зловмисник на базі порталу для клієнтів може створити фішинговий сайт ідентичний цьому, застосовуючи техніку T1583.001: Acquire Infrastructure: Domains, коли задля реалізування фейкового сайту купують нові домени. Після цього використати техніку T1566.002: Phishing: Spearphishing Link, для поширення нового посилання на сайт, який не є справжнім порталом, а насправді буде красти облікові дані та токени доступу в клієнтів.

Дані типу шаблону договору або внутрішні звіти підвищують успішну реалізацію атак розглянутих вище, наприклад додають правдоподібності при атаці T1656 Impersonation.

Якщо даних не було достатньо для серйозної атаки, вони можуть бути використані за технікою T1598: Phishing for Information, що означає фішинг задля збирання більшої кількості інформації про жертву.

Наслідками успішних реалізацій технік збору або атак описаних вище можуть бути інциденти, коли чітко продуманий фішинг змушує користувача відкрити небезпечне посилання чи завантажити шкідливе програмне забезпечення, відповідно до класифікації MITRE ATT&CK T1204: User Execution.

Однак в цьому експерименті компанія показала себе достатньо добре, оскільки чутливих даних типу паролів чи токенів доступу взагалі не знайшлось. Як і логінів, що унеможливлює атаку брутфорсу паролів на облікові записи,

оскільки не зрозуміло які з них є актуальними. Отже техніки типу T1087: Account Discovery та T1586: Compromise Accounts не можливо буде здійснити базуючись тільки на інформації з відкритих джерел. А без цього і техніка T1534: Internal Spearphishing не має сили. Також не вдалося зібрати достатньо технічних даних, T1082: System Information Discovery, оскільки доступних файлів в мережі були одиниці і їх аналіз нічого конкретного не показав.

Висновки до Розділу 3

Проведений експеримент підтвердив ефективність обраної методики та її здатність вирішувати поставлені завдання. Результати свідчать про хороший рівень інформаційної безпеки в компанії, оскільки більшість чутливих даних не було знайдено у відкритих джерелах. Аналіз отриманих даних підсвітив напрями, куди саме можуть цілити соціальні інженери і техніки, які для цього можуть застосовувати. Якщо дотримуватись рекомендацій і впровадити багаторівневу перевірку безпеки у вразливих векторах, можна уникнути ризиків від атак соціальної інженерії.

ВИСНОВКИ

В ході роботи було досягнуто мету, оцінити відкритість підприємства щодо витоків інформації по даних, знайдених у відкритих джерелах.

Огляд літератури показав, що проблема соціальної інженерії є дуже актуальною в наш час, оскільки не дивлячись на велику кількість технічних засобів запобігання витокам, завдяки людському фактору інформація продовжує витікати у відкриті джерела. Під час підготовки атаки соціальні інженери попередньо проводять розвідку у відкритих джерелах, щоб зібрати якомога більше чутливих даних і знати слабкі місця організації. Щоб попередити такі атаки було запропоновано проаналізувати інформацію про компанію, наявну в мережі, щоб завчасно знати про ризики її застосування.

Для цього було розглянуто методи пошуку даних у відкритих джерелах та детально проаналізовано вектори інтересу соціального інженера. Обрано метод пошуку за допомогою Google-dorks у пошуковій системі Google, що поєднує в собі найбільшу кількість релевантних результатів для аналізу. Розроблено список дорків, а також скрипт для оцінки знайдених даних, який полягає в обрахунку фактів наявності та актуальності даних в контексті атак соціальної інженерії, що в результаті показує найбільш реалістичну ситуацію з відкритістю організації.

Під час експерименту на тестовій організації було проаналізовано знайдені дані і порівняно їх з техніками MITRE att&ck, що дозволяє провести паралель між реальними інцидентами і оцінити відповідні ризики.

Сформований алгоритм дій є універсальним для багатьох компаній з високою інтернет присутністю і може допомогти уникнути ризиків пов’язаних з атаками соціальної інженерії. Методика дозволяє оцінити ступінь відкритості організації і виявити можливі вектори атак, що потребують впровадження відповідних заходів безпеки.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Shevchenko H., Stopochkina I., Babenko I., Peculiarities of phishing threats and preventive measures in the conditions of war in Ukraine // Theoretical and Applied Cybersecurity Vol. 4 No. 1 (2022). P.108-117. https://doi.org/10.20535/tacs.2664-29132022.1

2. Iryna Styopochkina; Mykola Ilin; Oleksandra Ponomarenko. Social Engineering in Modern Messengers: Applications for Offensive Security. //Information Technology: Computer Science, Software Engineering and Cyber Security, DOI: 10.32782/it/2023-2-10 (https://doi.org/10.32782/it/2023-2-10%22%20%5Ct%20%22_blank)

3. І. Стьопочкіна, І., Ільїн, К. (2024). Профілювання користувачів для підвищення стійкості персоналу об’єктів критичної інфраструктури до кібератак, які використовують людський фактор. Information Technology: Computer Science, Software Engineering and Cyber Security, 3, 159–168, doi: https://doi.org/10.32782/IT/2024-3-17

4. Oleksii Novikov, Mariia Shreider, Iryna Stopochkina, Mykola Ilin. Cyber Attacks Simulation for Modern Energy Facilities (https://ceur-ws.org/Vol-3887/paper4.pdf) // CEUR Workshop Proceedings. Selected Papers of the XXIII International Scientific and Practical Conference "Information Technologies and Security" (ITS 2023). URL: https://ceur-ws.org/Vol-3887/. Pp.35-49.

5. O. Chalyi, I. Stopochkina, Information retrieval and deanonymization in the tasks of early detection of potential attacks on critical infrastructure/ Кібербезпека: освіта, наука, техніка, Vol.2, Issue 26, pp. 305-322., 2024. https://doi.org/10.28925/2663-4023.2024.26.694

6. Ланде Д. В. OSINT у кібербезпеці: навч. посіб. – Київ: ТОВ "Інжиніринг", 2024. – 522 с.

7. Hassan N., Hijazi H., Alasmary W. Open Source Intelligence Methods and Tools: A Practical Guide to Online Intelligence. – 2nd ed. – Apress, 2018. – 321 c.

8. Bazzell M. Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information. – 7th ed. – IntelTechniques, 2019. – 575 c.

9. Що таке RSS? Як працює та навіщо потрібен. - ApiX-Drive. - URL: https://apix-drive.com/ua/blog/useful/shcho-take-rss

10. Molfar Global (Telegram-канал). - URL: https://t.me/molfar_global/869

11. Web Scraping: що це таке і як працює. - RapidSeedbox. - URL: https://www.rapidseedbox.com/uk/blog/web-scraping

12. Корисні інструменти для OSINT. - Molfar. - URL: https://molfar.com/useful-apps

13. Office of the New York State Comptroller. Cyberattacks on New York's Critical Infrastructure. - URL: https://www.osc.ny.gov/files/reports/pdf/cyberattacks-on-new-yorks-critical-infrastructure.pdf

14. Halliburton confirms cyberattack on its systems - URL: https://www.scworld.com/news/halliburton-confirms-cyberattack-on-its-systems

15. The Hacker News. AT&T Confirms Data Breach Affecting 109 Million Customers. – URL: https://thehackernews.com/2024/07/at-confirms-data-breach-affecting.html

16. The Hacker News. CERT-UA Warns DarkCrystal RAT Targets Ukraine. – URL: https://thehackernews.com/2025/03/cert-ua-warns-dark-crystal-rat-targets.html

17. CyberPeace Institute. Cyberconflicts: Threats and Attack Details. - URL: https://cyberconflicts.cyberpeaceinstitute.org/threats/attack-details

18. Hadnagy C. Social Engineering: The Science of Human Hacking. – 2nd ed. – Wiley, 2018. – 320 с.

19. Теорія та методи соціальної інженерії в кібербезпеці: Лабораторний практикум /[Електронний ресурс] : навч. посіб. для студ. спец. 125 «Кібербезпека», уклад. Ільїн К.І., Стьопочкіна І.В., 2023. https://ela.kpi.ua/handle/123456789/67176

20. MITRE ATT&CK®. Adversarial Tactics, Techniques, and Common Knowledge. – URL: https://attack.mitre.org/

ДОДАТОК А

```
import json

with open('items.json', 'r') as f:
    items = json.load(f)

with open('relevance_input.json', 'r') as f:
    relevance_input = json.load(f)

with open('presence_input.json', 'r') as f:
    presence_input = json.load(f)

w_presence = 0.1
w_relevance = 0.9

def calculate_score(weights, relevance_ratings, presence_ratings, w_relevance, w_presence):
    total_score = 0
    for key in weights:
        presence = presence_ratings.get(key, 0)
        relevance = relevance_ratings.get(key, 0)
        weight = weights[key]
            total_score += weight * (w_relevance * relevance + w_presence * presence)
    return total_score

score = calculate_score(items, relevance_input, presence_input, w_relevance, w_presence)*100
print(f"Сумарна оцінка: {score}")
```