Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна
Навчально-науковий інститут комп'ютерних наук та штучного інтелекту

Кафедра кібербезпеки інформаційних систем, мереж і технологій

До захисту допущено
Кафедрою КІСМіТ протокол №____ від «___» грудня 2025 р.

завідувач кафедри _____ Марина ЄСІНА
(підпис) *(ім'я, прізвище)*
«___» грудня 2025 р.

Кваліфікаційна робота
здобувача другого (магістерського) рівня вищої освіти

«Дослідження та аналіз механізмів маніпуляції в соціальній інженерії та методи їх виявлення»
(назва роботи)

Спеціальність (спеціалізація) 125 «Кібербезпека та захист інформації»

Освітня програма «Безпека інформаційних і комунікаційних систем»

Виконавець _____ Антон ГАЄВСЬКИЙ
(підпис) *(ім'я, прізвище)*

Науковий керівник _____ Марина ЄСІНА
(підпис) *(ім'я, прізвище)*

Харків – 2025

РЕФЕРАТ

Пояснювальна записка до магістерської роботи містить 63 сторінки, включає 7 рисунків, 6 таблиць, 1 додаток та 37 посилань на джерела. Робота присвячена комплексному дослідженню механізмів забезпечення приватності в інформаційних системах, що працюють із великими масивами даних, а також аналізу практичних методів підвищення стійкості до атак соціальної інженерії.

Метою роботи є дослідження та порівняльний аналіз соціально-інженерних методів впливу, виявлення їхніх психологічних механізмів, розробка формальних підходів до виявлення аномалій у поведінкових журналах та створення практичних методик підвищення стійкості організацій до фішингових атак.

Об’єкт дослідження – процес виявлення та попередження атак соціальної інженерії в корпоративних інформаційних системах.

Предмет дослідження – психологічні техніки фішингу, поведінкові індикатори атак у журналах, алгоритми розпізнавання аномалій та організаційні моделі підвищення стійкості до соціально-інженерних впливів.

Методи дослідження включають контент-аналіз фішингових повідомлень, статистичні методи виявлення відхилень, машинне навчання, моделювання сценаріїв поведінки користувачів, а також систематичний аналіз нормативно-правових документів і міжнародних стандартів кібербезпеки.

У роботі досліджено психологічні механізми фішингових атак, класифіковано основні техніки маніпуляції та наведено операціоналізацію цих технік у вигляді лінгвістичних маркерів. Отримані результати дали змогу сформувати системний набір методик підвищення стійкості організації.

Результати роботи можуть бути використані у діяльності підрозділів інформаційної безпеки, у проєктуванні систем раннього виявлення вторгнень, у корпоративних тренінгах з протидії фішингу, а також у дослідженнях, спрямованих на підвищення стійкості організацій до соціально-інженерних атак.

Ключові слова: СОЦІАЛЬНА ІНЖЕНЕРІЯ, ФІШИНГ, ПСИХОЛОГІЧНІ ТЕХНІКИ, ПОВЕДІНКОВІ ЛОГИ, ANOMALY DETECTION, IoA, ІНДИКАТОРИ АТАК, МАШИННЕ НАВЧАННЯ.

ABSTRACT

The explanatory note to the master's project comprises 63 pages and includes 7 figures, 6 tables, 1 appendix, and 37 references. The work is devoted to a comprehensive study of privacy-protection mechanisms in information systems that operate with large-scale datasets, as well as to the analysis of practical methods for strengthening resilience to social engineering attacks.

The purpose of the study is to investigate and compare social engineering techniques, identify their underlying psychological mechanisms, develop formal approaches for detecting anomalies in behavioral logs, and design practical methodologies that enhance organizational resistance to phishing attacks.

The object of the study is the process of detecting and preventing social engineering attacks in corporate information systems.

The subject of the study encompasses phishing manipulation techniques, behavioral attack indicators found in system logs, anomaly-detection algorithms, and organizational models aimed at increasing resilience to social engineering threats.

The methods of research include content analysis of phishing messages, statistical deviation-detection techniques, machine learning methods, user behavior scenario modelling, as well as systematic analysis of regulatory documents and international cybersecurity standards.

The work examines the psychological mechanisms of phishing attacks, classifies the main manipulation techniques, and presents their operationalization in the form of linguistic markers. The findings made it possible to develop a structured set of methodologies for improving an organization's resilience to social engineering.

The results of the research may be applied in the operation of security departments, in the design of early intrusion-detection systems, in corporate anti-phishing training programs, as well as in scientific studies focused on enhancing an organization's protection against social engineering attacks.

Keywords: SOCIAL ENGINEERING, PHISHING, PSYCHOLOGICAL TECHNIQUES, BEHAVIORAL LOGS, ANOMALY DETECTION, IoA, ATTACK INDICATORS, MACHINE LEARNING.

ЗМІСТ

ПЕРЕЛІК ПОЗНАЧЕНЬ І СКОРОЧЕНЬ

APT	– Advanced Persistent Threat, розширена постійна загроза
AUC	– Area Under the Curve, площа під кривою
BCE / BEC	– Business Email Compromise, компрометація бізнес-листування
BETH	– Behavioral Engagement Threat Hunting, поведінковий пошук загроз
CNN	– Convolutional Neural Network, згорткова нейронна мережа
CRM	– Customer Relationship Management, керування взаєминами з клієнтами
CSV	– Comma-Separated Values, значення, розділені комами
CTA	– Cyber Threat Alliance, альянс з протидії кіберзагрозам
DNS	– Domain Name System, система доменних імен
DDoS	– Distributed Denial of Service, розподілена відмова в обслуговуванні
EDR	– Endpoint Detection and Response, виявлення та реагування на загрози на кінцевих точках
ENISA	– European Union Agency for Cybersecurity, Агентство ЄС з кібербезпеки
FOMO	– Fear of Missing Out, страх втратити можливість
FP	– False Positive, хибне спрацювання
FTP	– File Transfer Protocol, протокол передавання файлів
HFACS	– Human Factors Analysis and Classification System, система аналізу та класифікації людських факторів
HTTP	– HyperText Transfer Protocol, протокол передавання гіпертексту
IoA	– Indicators of Attack, індикатори атаки
IoC	– Indicators of Compromise, індикатори компрометації
LLM	– Large Language Model, велика мовна модель

LOF – Local Outlier Factor, локальний фактор викиду

LSTM – Long Short-Term Memory, довга короткочасна пам’ять

MFA – Multi-Factor Authentication, багатофакторна автентифікація

ML – Machine Learning, машинне навчання

OSINT – Open-Source Intelligence, розвідка з відкритих джерел

ROC – Receiver Operating Characteristic, операційна характеристика приймача

SVM – Support Vector Machine, метод опорних векторів

SOC – Security Operations Center, центр операцій безпеки

SOAR – Security Orchestration, Automation and Response, оркестрація, автоматизація та реагування на інциденти безпеки

Sysmon – System Monitor, системний монітор

TPB – Theory of Planned Behavior, теорія запланованої поведінки

TTD – Time to Detect, час до виявлення

UEBA – User and Entity Behavior Analytics, аналітика поведінки користувачів та об’єктів

UBA – User Behavior Analytics, аналітика поведінки користувачів

URL – Uniform Resource Locator, уніфікований локатор ресурсу

VPN – Virtual Private Network, віртуальна приватна мережа

XDR – Extended Detection and Response, розширене виявлення та реагування

АГЗ – Активні групи загроз

ПЗ – Програмне забезпечення

ПК – Персональний комп’ютер

СІ – Соціальна інженерія

США – Сполучені Штати Америки

ШІ – Штучний інтелект

ВСТУП

Соціальна інженерія залишається одним із найнебезпечніших видів кіберзагроз, оскільки зловмисники орієнтуються не на технічні вразливості, а на людський фактор. У сучасних організаціях саме користувачі виступають першою лінією оборони, проте водночас є і найвразливішою ланкою, що зумовлює високий рівень успішності фішингових атак та маніпулятивних впливів. Розмаїття технік психологічного тиску, динамічний характер атак та зростаюча автоматизація інструментів злочинців формують потребу у системному вивченні механізмів впливу на користувачів і виробленні практичних підходів до підвищення стійкості.

Проблематика дослідження полягає у відсутності єдиної, комплексної методики оцінювання та протидії соціально-інженерним атакам, яка б одночасно охоплювала аналіз фішингових листів, поведінкові індикатори атаки, психологічні механізми маніпуляції та практичні засоби організаційного й технічного захисту. Незважаючи на наявність окремих рекомендацій і технічних рішень, більшість організацій стикаються з труднощами у формуванні цілісної стратегії, а успішні атаки продовжують спричиняти порушення роботи сервісів, витоки даних і фінансові збитки.

Актуальність дослідження зумовлена тим, що фішингові атаки залишаються ключовим вектором компрометації у понад 70% інцидентів, а їх складність і націленість зростають щороку. Потреба у практичних, відтворюваних і адаптивних методиках стійкості є критично важливою як для комерційних структур, так і для державних установ. Зростання дистанційної роботи, активне використання електронної пошти, месенджерів і хмарних сервісів створює нові умови для маніпулятивних впливів, що робить тему дослідження особливо затребуваною.

Метою дослідження є формування комплексної системи аналізу та протидії соціальній інженерії, яка поєднує моделювання фішингових листів,

дослідження поведінкових індикаторів атак, аналіз технічних журналів і розробку практичних методик стійкості.

Для досягнення поставленої мети застосовано методи контент-аналізу, емпіричне моделювання фішингових сценаріїв, статистичний аналіз поведінкових профілів, порівняння технік виявлення, а також методи експертного узагальнення для формування рекомендацій.

Результати дослідження будуть корисними фахівцям із кібербезпеки, адміністраторам інформаційних систем, аналітикам SOC, керівникам IT-підрозділів, а також організаціям, які впроваджують політики інформаційної безпеки.

Практична частина роботи може бути використана як методичний матеріал для внутрішніх тренінгів, побудови систем раннього виявлення загроз, аудитів безпеки та розробки стратегії протидії соціальній інженерії на рівні підприємств та державних структур.

1 СТАТИСТИКА, СТАНДАРТИ ТА МЕХАНІЗМИ СОЦІАЛЬНОЇ ІНЖЕНЕРІЇ

Соціальна інженерія (СІ) як феномен кіберзлочинності представляє собою комплексну систему атак, спрямованих на експлуатацію людської психіки, довіри та природних поведінкових стереотипів замість використання суто технічних вразливостей. На відміну від традиційних хакерських атак, які спрямовані на пошук недоліків у програмному забезпеченні, соціальна інженерія цілеспрямовано маніпулює емоціями, когнітивними упередженнями та соціальними нормами людини для досягнення несанкціонованого доступу до інформаційних систем, конфіденційних даних або матеріальних ресурсів.

1.1 Статистичні тенденції та ефективність соціальної інженерії

Згідно з щорічним звітом FBI IC3 за 2024 рік, кількість звернень про Інтернет-злочини сягнула 859532 випадків, а сумарні заявлені втрати перевищили $16 млрд, що на 33% більше порівняно з 2023 роком [1].

Найбільша кількість скарг була надіслана за фішинг, а саме 193407 звернень, далі вимагання з більше, ніж 86 тис. та компрометація даних (65 тис.). Водночас, найбільші фінансові збитки спричиняють шахрайські інвестиції (переважно пов’язані з криптовалютою) – понад $6,5 млрд. за 2024 рік. Для порівняння, інші типи соціально-інженерних атак теж завдали значних втрат: бізнес-Email-компрометація (BEC) призвела до близько $2,8 млрд. збитків при 21442 інцидентах, а шахрайство технічної підтримки – $1,46 млрд. при ~36 тис. випадків, що представлено у табл. 1.1 [1].

Тенденції підтверджують і інші міжнародні джерела. Згідно з дослідженням Verizon DBIR 2025 [2], до 68% випадків витоку даних причетний «людський фактор» – тобто помилки персоналу або успішні соціальні інженерні маніпуляції. Зокрема, майже 32% усіх зламів пов’язані з фішингом як початковим вектором атаки, що лише трохи менше, ніж у попередньому році, що може свідчити про поступову адаптацію компаній, але фішинг усе ще залишається одним із найпоширеніших методів злому. Для порівняння, звіт IBM

Security [3] показує, що у 2023 році фішинг був лідером серед способів проникнення у корпоративні мережі, фігуруючи у 41% проаналізованих інцидентів.

Таблиця 1.1 – Лідери за сумою фінансових втрат

Ранг	Тип інциденту	Скарги, 24	Ранг	Тип інциденту	Втрати, 24
1	Фішинг / спуфінг	193407	1	Інвестиційне шахрайство (крипто)	$6,57 млрд.
2	Здирництво (вимагання)	86415	2	Бізнес-Email-компрометація (BEC)	$2,77 млрд.
3	Витік персональних даних	64882	3	Шахрайство техпідтримки	$1,46 млрд.
4	Недоставка/невиплата (шахрайство з товарами)	49572	4	Компрометація персональних даних	$1,45 млрд.
5	Інвестиційні шахрайства	47919	5	Непоставка/невиплата товарів	$785 млн.

Хоча точні оцінки різняться за джерелами, усі аналітики сходяться на тому, що атаки соціальної інженерії продовжують домінувати, як основна причина успішних кібератак.

Вартість інцидентів соціальної інженерії також невпинно зростає. Середня збитковість порушення безпеки, спричиненого фішингом, оцінюється у 4,8-4,9 млн. (зважаючи на сукупні витрати на реагування, простій, репутаційні втрати тощо) за даними звіту IBM «Cost of a Data Breach 2024» [4]. Цей показник є одним із найвищих серед усіх векторів атак: для порівняння, компрометація облікових даних, як початковий вектор обходиться у ~4,8 млн., хмарні конфігурації – ~4,6 млн., а внутрішні зловмисники – ~4,9 млн.

Таким чином, фішинг є не лише надзвичайно поширеним, але й одним із найдорожчих типів атак для організацій. Відповідно до звіту IBM, у 2024 році фішинг навіть випередив викрадені облікові дані та став найпоширенішим початковим способом злому інформаційних систем – приблизно 16% усіх порушень безпеки починались із фішингу (проти 15% від викрадених паролів).

Детальніші метрики підтверджують ефективність CI у подоланні захисту організацій. Verizon зазначає, що середній показник клікабельності фішингових email-повідомлень складає 2,7%, тобто близько 1 з 37 цільових отримувачів переходить за шкідливим посиланням. Медіанний час реакції надто малий –

медіана часу до кліку – лише 21 секунда після отримання листа. Іншими словами, вже за лічені секунди після успішної доставки фішингового повідомлення співробітник може скомпрометувати систему. Водночас медіанний час, за який користувачі усвідомлюють атаку та повідомляють про підозрілий лист, становить 28 хвилин. За цей час зловмисники часто вже встигають розвинути атаку. Цікавим є те, що навчання персоналу здатне поліпшити ситуацію, а саме корпоративні програми кібергігієни та тренінги знижують відсоток натискань за фішинговими приманками на 30–38% у порівнянні з ненавченими співробітниками. Проте, попри такі заходи, абсолютна кількість успішних соціальних атак не спадає [1].

На глобальному рівні спостерігається зростання загальної кількості фішингових кампаній. За даними Анти-фішингової робочої групи (APWG) [5], лише за першу половину 2025 року було зафіксовано близько 2,13 мільйона фішингових атак (сумарно за Q1 та Q2 2025), що на 13% більше, ніж у попередньому кварталі. Кількість виявлених фішингових сайтів також збільшується. У 2024 році в світі щомісяця активувалось понад 80 тисяч нових фішингових веб-сторінок (на 22% більше, ніж роком раніше). При цьому зловмисники все частіше націлюються на конкретні галузі: найбільше фішингових атак проти корпоративних користувачів у 2024 році було спрямовано на логістичний сектор (25,3% від усіх цілей), сферу подорожей (20,4%), Інтернет-сервіси (16,4%), виробництво (12,8%) та фінансові послуги (10,3%). Такий розподіл свідчить про те, що зловмисники обирають жертв там, де є доступ до платежів, конфіденційних даних або порушення роботи може завдати найбільшої шкоди.

Характерно, що канали доставки соціально-інженерних атак розширюються. Традиційно основним каналом залишаються електронні листи: наприклад, 94% організацій повідомили про фішингові атаки електронною поштою у 2024 році. Проте зростає частка атак через SMS та месенджери, а також вішингу, про що буде вказано далі.

За даними дослідження Egress, 96% успішних фішингових інцидентів призводять до негативних наслідків для бізнесу, причому все частіше фішинг

комбінується з іншими методами впливу. Зловмисники застосовують цілі кампанії, що охоплюють кілька каналів одночасно: листи, дублювання в SMS та дзвінки – така багатоканальна соціальна інженерія набирає обертів, метою якої є збільшення шансів на контакт, обходячи заходи безпеки.

Окремо слід зазначити небезпечну тенденцію зростання кількості атак типу Business Email Compromise (BEC) – складних багатоходових шахрайств, у яких зловмисники видають себе за керівників або бізнес-партнерів і обманом змушують компанії здійснювати великі грошові перекази. За 2022–2024 роки загальні втрати від BEC у світі перевищили 8,5 млрд доларів. Лише у 2024 році більш, ніж 63% опитаних організацій постраждали від хоча б однієї спроби цієї атаки [1]. Попри відносно невелику кількість інцидентів BEC (лише 2,5% від загального числа звернень до FBI IC3), саме вони становлять непропорційно великий фінансовий ризик, а саме 2-ге місце за сумою втрат.

Середній збиток від однієї операції набагато перевищує збитки від пересічного фішинг інциденту, оскільки такі атаки націлені на великі транзакції. Наприклад, у 2023 році в США було зафіксовано понад 21 тис. випадків BEC із сукупними втратами близько 2,7 млрд. доларів.

Зважаючи на це, організації приділяють все більше уваги протидії BEC, впроваджуються процедури багаторівневого підтвердження платежів, навчання бухгалтерів та менеджерів ознакам обману, використовуються технічні засоби перевірки доменів відправників тощо.

Також важливо буде зазначити, що ландшафт атак зазнав як еволюційних змін, так і появи якісно нових векторів, обумовлених геополітичними подіями та технологічними інноваціями. Згідно з узагальненими даними ENISA за період липень 2024 – червень 2025 р., фішинг (у широкому розумінні, включно з email-фішингом, вішингом, смішингом та шкідливим рекламним середовищем) став причиною близько 60% успішних зломів, тоді як експлуатація вразливостей – 21,3% випадків (рис. 1.1) [6].

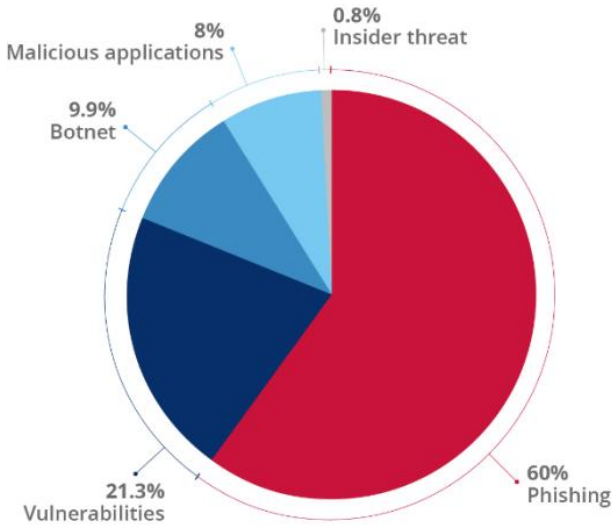

Рисунок 1.1 – Розподіл початкових векторів проникнення у кіберінцидентах [1]

Типові техніки соціальної інженерії у цей період залишаються на подив ефективними. Відповідно до MITRE ATT&CK [7], найбільш розповсюджені методи, що застосовуються зловмисниками для початкового доступу, – це різновиди фішингу: Spearphishing Attachment (T1566.001) – цільове надсилання шкідливих вкладень, Spearphishing Link (T1566.002) – надсилання спеціально сформованих посилань, та Phishing via Service (T1566.003) – використання сторонніх платформ (соцмереж, месенджерів) для встановлення контакту.

Окрім email-повідомлень, все більшої популярності набувають інші вектори доставки: SMS-фішинг (smishing), шахрайські повідомлення в месенджерах WhatsApp, Telegram, Viber та прямі дзвінки жертвам (vishing). Зокрема, у 2024 році понад 75% цільових атак на бізнес включали елемент телефонного або текстового соціального інжинірингу. Така багатоканальна тактика створює ілюзію легітимності: жертва може отримати, скажімо, офіційного на вигляд листа, а слідом – дзвінок «від служби підтримки», що посилається на цей лист, тим самим підштовхуючи до виконання потрібних дій.

Активні групи загроз (АГЗ) у 2024–2025 роках значною мірою продовжили відпрацьовувати прийоми СІ, часто адаптуючи їх до поточної воєнно-політичної обстановки. З початком повномасштабної війни в Україні 2022 року помітно активізувалися російські державні групи APT, які здійснюють кібершпіонаж та

дезінформаційні операції. Війна також породила явище «фейктивізму» – коли державні хакери маскуються під активістів. ENISA у своєму огляді 2025 року [7] відзначає злиття почерків різних груп: державні та про-державні актори дедалі частіше імітують стилі атак, притаманні кіберзлочинцям чи хактивістам. У березні 2022 р. було здійснено резонансну атаку: на телеканалах України хакери запустили сфабриковане відеозвернення Президента Зеленського із закликом скласти зброю, хоча такий грубий монтаж швидко викрили, сам прецедент демонструє новий рівень соціальної інженерії у кібервійні – поєднання технічного злому інфраструктури з психологічною операцією, націленою на широку аудиторію. Згодом подібні методи можуть бути використані у більш цілеорієнтований спосіб.

Північнокорейські АГЗ у 2024 р. теж яскраво проявили себе, зосередившись на фінансово мотивованих атаках із застосуванням СІ. Група Lazarus, спонсорована КНДР, відома зламуванням криптовалютних бірж та банків, вдалася до нової тактики під кодовою назвою «ClickFake». Вона полягала у маскуванні під ІТ-рекрутерів на LinkedIn. Lazarus розсилала фальшиві запрошення на роботу спеціалістам у криптоіндустрії, спілкувалася з жертвами від імені відомих компаній, а на етапі «співбесіди» надсилала файл із шкідливим ПЗ замість тестового завдання. В цілому, у 2023–2024 рр. КНДР подвоїла зусилля з кіберкрадіжок криптовалют для фінансування ракетної програми, і соціальна інженерія була ключовим інструментом цих операцій [8].

Оборонні відомства та корпоративний сектор також фіксують появу нових СІ технік. Один із трендів 2024 року – атаки методом MFA prompt bombing. Зловмисник, дізнавшись пароль користувача, багаторазово надсилає пуш-запити на його телефон до тих пір, поки втомлена або неуважна жертва не натисне Approve. Підлітковим кіберугрупуванням Scattered Spider був здійснений гучний злам казино, вони спершу видурили у співробітника ІТ-підтримки облікові дані та скинули йому MFA, видавши себе за іншого працівника, після чого проникли в мережу і розгорнули програму-вимагач. Таким чином, атаки через help desk стали новим вектором, Verizon зазначає зростання цього виду у 2025 році [9].

Окремо слід розглянути роль штучного інтелекту у сучасних загрозах соціальної інженерії. 2024 рік позначився стрімким поширенням доступних інструментів генеративного ШІ. Зловмисники активно застосовують великі мовні моделі для покращення якості фішингових повідомлень, зменшуючи кількість граматичних та стилістичних помилок, які раніше допомагали відрізнити підробку. У результаті подібні шахрайські листи стали набагато переконливішими. Verizon теж відзначає, що кількість фішингових кампаній з використанням генеративного ШІ зросла на 22% за рік. Більше того, з'явився цілий «підприємницький» напрям у даркнеті. Продаються послуги Phishing-as-a-Service з наповненням, згенерованим ШІ, і навіть цілі набори, що включають реалістичний скрипт розмови для кол-центрів та підроблені сайти з динамічно згенерованим контентом, який підлаштовується під жертву.

Загалом, у 2024-2025 роках ландшафт СІ став ще більш різноманітним і складним. Кіберзлочинці комбінують перевірені часом техніки з новими інструментами, збільшуючи ефективність. Державні хакери активно використовують соціальну інженерію для шпигунства та інформаційних операцій, впливаючи на громадську думку через фейки. З іншого боку, фінансово мотивовані угрупування вдосконалюють соціальні атаки задля отримання прибутку – відточують схеми BEC, атакують ланцюги постачань і партнерів, освоюють нові сценарії обману.

1.2 Економіка соціальної інженерії

Даркнет у сучасних умовах формує багаторівневу тіньову економічну інфраструктуру, у межах якої інструменти та сервіси для атак СІ продаються за тими ж ринковими законами, що й легальні цифрові продукти. Попит, ризик, конкуренція та репутація продавців визначають кінцеву вартість інструментів. У екосистемі соціальна інженерія виступає не лише технікою, а фактично комерційним товаром, який можна придбати, налаштувати або орендувати, як було згадано раніше.

Одним із найпоширеніших продуктів у даркнеті залишаються фішингові набори. Вони містять повні комплекти скриптів і шаблонів, які дозволяють

навіть некваліфікованим користувачам організувати фішингову кампанію без знання програмування. За даними Group-IB, ще у 2019 році середня ціна такого набору досягала приблизно 300 доларів і демонструвала зростання на 149% порівняно з попереднім роком. У 2024-2025 роках вартість найпоширеніших варіантів коливалася від 50 до 900 доларів залежно від складності та якості виконання. Окрему категорію представляли набори, що використовували ШІ для автоматичної персоналізації листів та фішингових сторінок. Такі комплекти продавалися за суму від 50 до 500 доларів і забезпечували значно вищу ефективність через адаптивність вмісту.

Аналітики KeepnetLabs [10] у 2024 році зафіксували зростання кількості фішингових наборів у даркнеті приблизно на 50%. На окремих форумах було виявлено понад 16200 унікальних наборів, що вказує на високий рівень конкуренції та масштаб ринку.

Даркнет є основною платформою для торгівлі краденими обліковими даними, які формують окремий сегмент економіки CI. Інформація про банківські картки з PIN-кодом коштує в середньому близько 20 доларів. Доступи через RDP до корпоративних комп'ютерів продаються у діапазоні від 150 до 400 доларів. Облікові дані банківських рахунків коштують від 200 до 1000 доларів, тоді як пакети особистої інформації (fullz) мають ціну від 1 до 20 доларів. Неактивні або частково активні облікові записи також становлять комерційну цінність, і їхня вартість зазвичай перебуває в межах від 1 до 50 доларів залежно від призначення сервісу.

У даркнеті поширений і продаж спеціалізованих кримінальних послуг, які доповнюють інструменти CI. Добовий DDoS-удар зі швидкістю 20000-50000 запитів за секунду коштує приблизно від 200 доларів. Послуги з відновлення або підбору паролів оцінюються на рівні 100-500 доларів. Розширений OSINT-аналіз жертви коштує від 500 до 2000 доларів. Індивідуальна розробка фішингового набору може досягати 1000-5000 доларів. Створення голосового deepfake-клонування для атак вішингового типу зазвичай коштує 500-2000 доларів.

Найяскравішим проявом економічної ефективності CI є рентабельність атак формату BEC. Середня собівартість однієї операції становить близько 170

доларів, що включає приблизно 100 доларів за фішинговий набір, 50 доларів умовної вартості OSINT-роботи та 20 доларів на інфраструктурні елементи. За даними Proofpoint, середній фінансовий збиток від одного BEC-інциденту у США в 2024 році становив 137132 долари.

Рентабельність може перевищувати 800000%, а кожен долар витрат приносить злочинцю приблизно 8000 доларів прибутку. За умови десяти спроб атак на місяць навіть мінімальна успішність у 10 відсотків дає змогу отримувати понад 135000 доларів щомісячного доходу, що у річному вимірі перевищує 1,6 млн. доларів при загальних витратах близько 20400 доларів.

Схеми CEO fraud демонструють не менш вражаючу економічну ефективність. Їхня собівартість зазвичай перебуває в межах від 100 до 300 доларів, однак рівень успішності часто вищий завдяки поєднанню ефектів авторитету та терміновості, що провокує негайну реакцію співробітників без додаткової перевірки.

Масовий фішинг є ще дешевшим з погляду собівартості. Розсилка 100000 електронних листів зазвичай коштує близько 50 доларів. За типової статистики успішності (2-5% переходів та 0,5-1% введення даних) зловмисник може отримати до 1000 дійсних облікових записів. За середньої ціни 10 доларів за обліковий запис загальний прибуток складає близько 10000 доларів при рентабельності, що перевищує 20000%. Масштабованість і автоматизація є ключовими чинниками, що підтримують високу рентабельність CI. Завдяки ШІ злочинці скорочують підготовку персоналізованих фішингових листів до кількох хвилин, а автоматизовані системи виконують розвідку, синтез текстів, створення фальшивих голосових записів та відстеження результатів атак. Таким чином, одна особа здатна запускати сотні або тисячі однотипних атак на добу при мінімальних витратах.

У підсумку даркнет сформував повноцінну кримінальну економіку CI, у якій поєднання низької собівартості, високої прибутковості та легкості масштабування робить соціально-інженерні атаки одним із найефективніших та найприбутковіших напрямів кіберзлочинності.

1.3 Теорії людських помилок і когнітивних вразливостей

Людський фактор залишається ключовим елементом, який визначає вразливість організацій до атак СІ. Технічні недоліки можуть бути усунені шляхом оновлень та змін у конфігурації систем, проте помилки, зумовлені особливостями людської поведінки, мають постійний характер, оскільки походять із природних обмежень когнітивних процесів. Саме тому жодна інфраструктура не може бути повністю захищеною від СІ, доки взаємодія людини з інформаційними системами залишається невід'ємною частиною операційної діяльності.

1.3.1 Модель «швейцарського сиру»

Однією з найвпливовіших теоретичних моделей, що пояснюють природу людських помилок у складних соціотехнічних системах, є концепція, розроблена Джеймсом Різоном у його праці 1990 року «The Contribution of Latent Human Failures to the Breakdown of Complex Systems» [11]. Запропонована ним модель, що отримала назву моделі «швейцарського сиру», надала якісно нове розуміння того, як виникають інциденти безпеки у складних організаційних структурах, у яких критичні функції залежать від взаємодії людини та технологій.

Згідно з цією моделлю, система безпеки організації складається з кількох взаємопов'язаних шарів захисту, кожен з яких можна представити як окрему «скибку» швейцарського сиру. У кожному шарі існують власні вразливості, що метафорично зображуються як «дірки». За нормальних обставин ці вразливості не збігаються між собою, і система зберігає здатність блокувати інциденти. Проте час від часу виникають ситуації, коли вразливості в усіх шарах проєктуються одна на одну, утворюючи так звану «траєкторію можливості аварії» – шлях, через який загроза може пройти через усі рівні захисту без жодного бар'єра. Саме такий збіг є критичним моментом, що призводить до виникнення інциденту.

Ключовим висновком Різона є твердження, що жоден серйозний інцидент безпеки не має однієї єдиної причини. Будь-яка аварія або вдала атака СІ є

результатом системної сукупності порушень у різних шарах організації, що має особливо важливе значення для розуміння атак СІ, оскільки успішність таких атак ніколи не зводиться лише до того, що певний співробітник натиснув на фішингове посилання. Успіх атаки, як правило, є наслідком взаємодії низки чинників, серед яких недостатня підготовка персоналу, слабкі технічні контролі, дефіцит ресурсів, недосконалі політики організації та наявність зовнішніх стресорів, що впливають на людину в момент прийняття рішення.

Модель Різона поділяє всі неправомірності на два категоріальні типи – активні та латентні. Активні неправомірності є тими помилками, що проявляються безпосередньо перед інцидентом і стають його «видимими» причинами. У контексті СІ типовим прикладом такої помилки є клік співробітника на фішингове посилання або відкриття шкідливого вкладення. Латентні неправомірності мають значно глибший і системний характер. Довготривалі структурні проблеми, які формують середовище, у якому активна помилка стає можливою. До латентних умов можна зарахувати недостатній рівень навчання співробітників щодо виявлення фішингу, відсутність надійних поштових фільтрів, перевантаження персоналу та постійний часовий тиск, недосконалі процедури безпеки, відсутність інструментів для повідомлення про підозрілі повідомлення, а також організаційну культуру, що не приділяє достатньої уваги питанням інформаційної безпеки. Подібні латентні проблеми можуть існувати роками, залишаючись непомітними, доки збіг обставин не призведе до великого інциденту.

Практичне застосування моделі швейцарського сиру до СІ добре ілюструється сценаріями успішних атак, зокрема BEC або CEO fraud. Наприклад, у випадку з атакою видавання себе за керівника організації у ролі першого шару виступає навчання співробітників. Якщо організація забезпечує лише загальне ознайомлення з фішингом, але не навчає розпізнавати специфічні ознаки CEO fraud, виникає перша «дірка». Другим шаром є технічний контроль: якщо система фільтрації пошти не перевіряє внутрішні доменні адреси або не аналізує подібність імен відправників, зловмисник може використати майже ідентичну адресу для надсилання повідомлення. Третій шар – це організаційні процеси. За

відсутності процедури двоканальної верифікації фінансових операцій співробітник, який отримав підроблений «терміновий наказ», не матиме механізму для підтвердження його достовірності. Четвертий шар стосується культури безпеки: якщо швидкість виконання завдань переважає над перевіркою їхньої легітимності, співробітники рідко ставлять уточнюючі запитання. У момент, коли всі ці вразливості збігаються, атака успішно проходить крізь систему захисту.

Попри свою значущість, модель «швейцарського сиру» отримала критику за недостатню здатність відобразити динамічну природу організаційних систем. Сам Різон у 2005 році розширив свою модель, запропонувавши більш деталізовану класифікацію типів помилок.

Найбільш відомим розвитком цієї концепції стала система Human Factors Analysis and Classification System (HFACS), створена Скоттом Шаппеллом і Даґом Віґманном [12]. HFACS пропонує чотирирівневу структуру аналізу людських помилок. На першому рівні розташовуються небезпечні дії операторів, що безпосередньо призводять до інциденту; на другому – передумови цих дій, пов’язані зі станом людини, компетенціями, втомою або зовнішніми стресорами; на третьому – помилки керівництва, що створюють сприятливі умови для порушень; на четвертому – глибинні організаційні фактори, які формують загальне середовище функціонування системи.

HFACS підсилює модель Різона тим, що забезпечує ширший інструментарій для аналізу першопричин інцидентів. Система дозволяє не лише виявляти активні та латентні помилки, а й визначати конкретні рівні, на яких організація може впровадити втручання, усунути вразливості та запобігти повторенню подібних ситуацій. Таким чином, інтеграція моделей Різона та HFACS забезпечує комплексне розуміння людського фактору у безпеці й дозволяє організаціям моделювати ризики СІ не лише через поведінку окремої особи, а як результат системної взаємодії технологічних, організаційних та когнітивних чинників.

1.3.2 Теорія когнітивних упереджень

Когнітивні упередження становлять систематичні помилки у судженні, які виникають тоді, коли людина використовує спрощені «психологічні ярлики» для пришвидшеної обробки інформації. На відміну від випадкових помилок, такі упередження мають передбачуваний і повторюваний характер, оскільки особи з подібними рисами особистості в аналогічних ситуаціях демонструють однакові типи помилок.

Сучасні дослідження доводять, що значна частина рішень ухвалюється за допомогою швидкої, інтуїтивної системи мислення (умовно система 1), тоді як повільна, аналітична система (умовно система 2) залучається значно рідше. За умов часових обмежень, високого когнітивного навантаження або емоційного тиску людина ще більшою мірою покладається саме на інтуїтивну обробку інформації, що закономірно підвищує ймовірність виникнення упереджень [13].

Одним із найважливіших феноменів у контексті СІ є упередження авторитету. Люди мають схильність приписувати вищу надійність і компетентність особам, які сприймаються як авторитетні, та знижувати рівень критичної перевірки їхніх вимог. Класичні експерименти Стенлі Мілґрама [14] продемонстрували, що приблизно 65% учасників виявляли високий рівень покори авторитету, навіть тоді, коли власними очима спостерігали очевидну неправомірність дій.

Іншим поширеним явищем є упередження підтвердження. Люди прагнуть шукати, помічати та інтерпретувати інформацію таким чином, щоб вона відповідала їхнім уже сформованим переконанням, і водночас ігнорують або применшують значення даних, які суперечать цим представленням. У контексті фішингових атак це означає, що, якщо працівник уже очікує повідомлення від банку щодо нібито «проблеми з рахунком», то фішинговий лист, який імітує таке повідомлення, сприймається як природне підтвердження його очікувань. Унаслідок цього знижується імовірність того, що він ретельно перевірятиме адресу відправника, текст повідомлення чи вкладення.

Важливу роль відіграє також упередження доступності. Його сутність полягає в тому, що люди оцінюють імовірність події залежно від того, наскільки легко їм згадати подібні випадки. Якщо співробітник нещодавно чув про масштабний витік даних або бачив новини про кібератаку, він може переоцінювати загрозу саме такого сценарію у своїй організації. Атаки СІ можуть використати це, надсилаючи листи, які нібито повідомляють про «терміновий витік даних» або необхідність «негайного скидання пароля», що змушує жертву діяти імпульсивно.

Окрему групу становлять упередження, пов’язані із соціальним контекстом, зокрема упередження соціального доказу. Люди мають тенденцію вважати поведінку більш прийнятною або правильною, якщо вірять, що «інші вже так зробили». У фішингових повідомленнях цей механізм експлуатується формулюваннями на кшталт «інші менеджери вже погодили цю транзакцію» або «відділ фінансів уже виконав аналогічну операцію». Такі повідомлення створюють ілюзію групової норми, у межах якої відмова чи додаткові запитання здаються небажаними.

Не менш значущим є упередження якорювання. Перша отримана інформація стає для людини своєрідним «якорем», який впливає на подальше сприйняття і прийняття рішень. У випадку BEC атаки перший лист може містити запит на надзвичайно велику суму переказу. Навіть, якщо цей запит буде відхилено, наступні прохання про менші суми підсвідомо здаватимуться співробітнику більш «розумними» у порівнянні з початковим «якорем», що знижує рівень настороженості.

Упередження оптимізму впливає на те, як люди оцінюють власну вразливість. Багато людей схильні вважати, що негативні події більш імовірні для інших, ніж для них самих. Дослідження демонструють, що, наприклад, 88% водіїв оцінюють власні навички керування як вищі за середні, що статистично неможливо. Аналогічним чином, за результатами опитувань, приблизно 74% користувачів вважають себе більш обізнаними у виявленні фішингових листів, ніж середньостатистичний користувач [15]. Такий розрив між представленням

про власну компетентність і реальною поведінкою призводить до недооцінки ризиків і відмови від додаткових заходів безпеки.

Окремо варто згадати упередження, пов’язане зі страхом щось упустити. Люди демонструють підвищену схильність до імпульсивних рішень тоді, коли бояться пропустити важливу можливість чи вигоду. Зловмисні сценарії, у яких використовується риторика «обмеженої пропозиції», «останнього шансу» або «негайної знижки», впливають на це упередження і стимулюють швидке прийняття рішень без належної перевірки.

Важливо наголосити, що когнітивні упередження у реальних умовах рідко проявляються ізольовано. Найчастіше вони накладаються одне на одне та взаємно підсилюються. Атака, яка одночасно апелює до авторитету і створює відчуття терміновості, буде значно ефективнішою, ніж та, що базується лише на одному принципі. Додатковий ризик створює високий рівень когнітивного навантаження. Якщо співробітник обробляє велику кількість листів, виконує декілька завдань паралельно та відчуває часовий тиск, його здатність до критичного аналізу різко знижується, а ймовірність того, що рішення буде прийнято автоматично, керуючись упередженнями, суттєво зростає, що робить когнітивні упередження одним із центральних механізмів, які експлуатує CI, і підкреслює важливість їх урахування при проєктуванні програм навчання та заходів з підвищення обізнаності.

1.3.3 Теорія планованої поведінки

Теорія планованої поведінки (TPB), запропонована Іджеком Ажєном у 1985 році [16, 17], є однією з ключових психологічних моделей, що пояснює механізми прийняття рішень та чинники, які визначають ймовірність здійснення певної поведінки. Вона ґрунтується на припущенні, що поведінка людини не є випадковою або хаотичною, а формується під впливом трьох взаємопов’язаних детермінант. Першою детермінантою є ставлення до поведінки, тобто загальна оцінка індивіда щодо того, чи є певна дія бажаною, доцільною або, навпаки, ризикованою та небажаною. Другою є суб’єктивні норми, що відображають

соціальні очікування і сприйняття того, як значущі інші оцінюють цю поведінку. Третьою є сприйманий контроль поведінки, який визначає ступінь, у якому людина відчуває власну здатність успішно здійснити певну дію. У сукупності ці три елементи формують поведінковий намір, який, за твердженням моделі, є найточнішим предиктором реальної поведінки.

Застосування ТРВ у контексті атак СІ, зокрема фішингу, дає змогу пояснити, чому окремі працівники організацій виявляють більшу або меншу схильність взаємодіяти з підозрілими повідомленнями. Якщо розглянути ситуацію, у якій співробітник отримує потенційно шкідливий лист, роль першої детермінанти – ставлення до поведінки – полягає в тому, як він оцінює ризики. Коли людина вважає, що натискання на посилання може становити загрозу безпеці, її намір взаємодіяти з повідомленням істотно знижується. Однак, якщо вона сприймає лист як легітимний або вважає дію терміновою та необхідною, ймовірність взаємодії, навпаки, зростає.

Суб'єктивні норми також відіграють критичну роль. У середовищах, де працівники демонструють зневажливе ставлення до кіберзагроз та схильні вважати фішинг незначною проблемою, формується культура уникнення перевірки повідомлень і, відповідно, низький рівень обережності. Натомість у колективах, де існує практика активного обговорення інцидентів і заохочується повідомлення про підозрілі листи, співробітники демонструють вищу готовність до критичного аналізу інформації та своєчасного реагування.

Третім фактором є сприйманий контроль поведінки, який стосується впевненості індивіда у власній здатності правильно розпізнати загрозу. Якщо людина вважає, що їй бракує навичок для ідентифікації фішингових повідомлень, вона частіше приймає імпульсивні рішення та менше довіряє власним здібностям. Натомість відчуття компетентності та достатнього рівня контролю – навіть за наявності стресу або інтенсивного робочого навантаження – сприяють більш обережній поведінці, ретельному аналізу повідомлень та схильності уникати небезпечних дій.

Теорія планованої поведінки має важливе практичне значення для розуміння причин успішності атак СІ в організаціях. Вона демонструє, що проста

передача знань працівникам не забезпечує необхідного рівня безпеки. Навіть, якщо людина може теоретично розпізнати ознаки фішингу (позитивний вплив першої детермінанти), це не гарантує належної реакції, якщо організаційна культура не підтримує повідомлення про загрози, а працівник не відчуває достатнього контролю над ситуацією. Таким чином, ефективна протидія СІ потребує не лише навчання, а й комплексного впливу на формування соціальних норм у колективі та підвищення впевненості співробітників у власних можливостях ухвалювати безпечні рішення.

1.3.4 Нормалізація відхилення

Феномен нормалізації відхилення описує процес, за якого порушення вимог безпеки поступово стає звичною та прийнятною практикою в організації, якщо таке порушення не супроводжується негайними негативними наслідками. Цей термін отримав широке поширення завдяки працям Дайани Воган та Джеймса Різона, які досліджували катастрофи космічних шатлів Challenger та Columbia [18]. Згодом концепцію почали застосовувати до різних сфер, зокрема до кібербезпеки та протидії атакам СІ, оскільки механізм формування поведінкових відхилень в організаціях виявився універсальним.

Початкова стадія нормалізації зазвичай передбачає незначне порушення політики безпеки, що часто виправдовується браком часу, ресурсними обмеженнями або переконанням, що один невеликий виняток не створює суттєвого ризику. Оскільки після такого порушення не виникає негайних негативних наслідків, співробітники починають сприймати відхилення як нешкідливе. Упродовж певного часу подібна поведінка повторюється дедалі частіше, набуваючи характеру норми. Нові працівники, які потрапляють у вже сформовану поведінкову модель, переймають її як допустиму або навіть «правильну», оскільки вона видається природною частиною усталених робочих процедур. На більш пізніх етапах ці порушення перетворюються на елемент внутрішньої культури: дотримання політик починає розглядатися як щось необов’язкове, а ті, хто намагається діяти відповідно до вимог безпеки, іноді

стикаються з тиском колективу або звинуваченнями у тому, що вони «ускладнюють роботу» [19].

Наслідки такого процесу є особливо небезпечними для кібербезпеки. Атаки СІ часто спираються не на одну конкретну помилку, а на сукупність невеликих, але систематичних відхилень від вимог. Організація може роками ігнорувати окремі елементи безпеки – пропускати двоканальну верифікацію платежів, не перевіряти правильність електронних адрес, використовувати спрощені процедури доступу або залишати відкритими певні порти задля пришвидшення роботи – і водночас не відчувати реальних наслідків. Така тривала відсутність інцидентів створює хибне відчуття безпеки. Коли ж відбувається серйозна атака, співробітники часто заявляють, що не очікували подібного сценарію, хоча насправді організація протягом тривалого часу поступово демонтувала власні захисні механізми.

Запобігання нормалізації відхилення потребує цілеспрямованого втручання з боку керівництва та формування сильної культури безпеки. Лідерство організації повинно демонструвати чітку прихильність до дотримання політик навіть тоді, коли це уповільнює робочі процеси. Співробітники повинні мати можливість повідомляти про відхилення без ризику санкцій чи критики, що формує культуру прозорості та відповідальності. Регулярний перегляд та переатестація політик безпеки сприяють підтриманню актуальності вимог та зменшують ризик того, що працівники забуватимуть про їх існування. Важливо також пам’ятати, що тривала відсутність інцидентів не є свідченням того, що ризики відсутні; навпаки, досвід численних аварій і кібератак показує, що саме така відсутність негативних подій створює передумови для найсерйозніших порушень.

1.4 Психологічні механізми маніпуляції

Психологічні механізми, які лежать в основі успішних атак соціальної інженерії, не є універсальними. Люди з різними демографічними характеристиками, особистісними рисами та професійним досвідом мають неоднакову вразливість до однакових тактик маніпуляції.

1.4.1 Демографічні характеристики

Молоді люди (18-25 років) демонструють 65% показник натискання на шахрайські посилання порівняно з іншими віковими групами. Однак важливо розуміти, що високий відсоток часто пояснюється не недостатністю критичного мислення, а вищою активністю в Інтернеті та більшою часткою часу (в середньому 5 годин 48 хвилин), проведеною в цифровому середовищі. Молоді люди отримують більше фішингових повідомлень і тому мають більше можливостей для помилок [20].

До психологічних властивостей молодих людей можна включити: страх пропустити щось важливе (FOMO), що підсилює сприйнятливість до пропозицій із лімітом часу чи формулюваннями на кшталт «останній шанс». Висока залежність від соціальних доказів спричиняє те, що повідомлення на основі принципу «інші вже зробили» впливають на них значно сильніше. Крім того, природна допитливість спонукає молодих людей переходити за невідомими посиланнями та відкривати файли, прагнучи дізнатися більше.

На людей середнього віку (25-40 років) припадає 40-50% натискань. Група часто знаходиться на позиціях з помірною відповідальністю у компаніях, тому вони отримують менше уваги від зловмисників, ніж фінансові працівники. До психологічних вразливостей відносять авторитет, термінову надзвичайність та довіру на попередніх взаємодіях.

Люди старшого віку (40-50+ років), насправді, більш вразливі до атак фішингу, ніж молоді люди. Однак причини відрізняються. Люди старшого віку демонструють 50-60% показник натискань на фішингові посилання, особливо для атак, які видаються як офіційні повідомлення від важливих установ. Низький рівень цифрової грамотності, довіра, послаблена пам’ять та увага з емоційною вразливістю впливає на отримані результати [21].

Важливо також розглянути, як гендерний фактор впливає на рівень впливу СІ. Дослідження показують, що жінки більш уразливі до атак, ніж чоловіки. Показник натискань на фішингові посилання для жінок становить 53-55%, порівняно з 40-45% для чоловіків. До чинників відносять вищий рівень агрегації,

а саме готовність до комунікації, емпатія та довіра, низький рівень скептицизму, обов’язок.

Чоловіки показують нижчий рівень вразливості до традиційного фішингу, як було вказано раніше, однак вони більш уразливі до атак, які апелюють до їхнього статусу, компетентності та честі. До чинників відносять авторитет, технічну впевненість та конкуренцію.

Не менш значущим є й те, що рівень освіти безпосередньо впливає на сприйнятливість до соціальної інженерії. Людям із середньою освітою, як правило, важче виявити підозрілі ознаки в повідомленнях, зрозуміти технічні нюанси чи розпізнати невідповідності в термінології. Узагальнення демографічних та доповнення професійними факторам представлено у табл. 1.2.

Таблиця 1.2 – Категоризація за демографічними та професійними показниками

Вік/стать	Рівень вразливості	Показник натискань	Категорія	Рівень вразливості	Показник натискань
Вік 18-25	Високий	+65% до інших груп	Низька освіта	Середньо-висока	45-50%
Вік 25-40	Середній	40-50%	Середня освіта	Середня	35-40%
Вік 40-50	Середньо-високий	45-55%	Висока освіта	Низька-середня	25-35%
Вік 50+	Високий	50-60%	Фінансовий персонал	Дуже високий	28-30% відкриття BEC
Жінки	Високий	53-55%	IT персонал	Низький	15-20%
Чоловіки	Середній	40-45%	Адміністративний персонал	Високий	40-50%

Водночас варто підкреслити й протилежний аспект: хоча вища освіта зазвичай розвиває критичне мислення та аналітичні навички, це не гарантує обізнаності у сфері кібербезпеки. Тому навіть освічені користувачі можуть залишатися вразливими, якщо їхня підготовка не охоплює безпекові питання.

1.4.2 Класифікація маніпулятивних прийомів

Для глибшого дослідження потрібно виокремити низку основних психологічних маніпуляцій, що найчастіше застосовуються зловмисниками для впливу на людей: страх, терміновість, авторитет, соціальний доказ, взаємність,

знайомість/приязнь, цікавість або вигода, довіра. Кожен з цих прийомів базується на викликанні відповідного емоційного тригера та проілюстрований типовими сценаріями атак [22].

Маніпуляція страхом передбачає залякування жертви та створення відчуття загрози або неминучих негативних наслідків. Зловмисник свідомо провокує паніку чи тривогу, щоб спонукати людину поспіхом виконати потрібні дії. Під впливом страху користувач менш схильний перевіряти достовірність повідомлення і може виконати вказівки зловмисника, аби уникнути уявної небезпеки.

Принцип створення терміновості полягає у штучному обмеженні часу на обмірковування, що змушує жертву діяти негайно. Зловмисники часто використовують фрази на кшталт «дійте зараз», «залишилося 10 хвилин» або погрожують невідкладними наслідками, щоб викликати у жертви відчуття поспіху. Такий поспіх дезактивує раціональний аналіз ситуації.

Принцип взаємності апелює до відчуття обов'язку «віддячити» за отриману послугу або подарунок. У СІ цей прийом реалізується так: спочатку зловмисник «дарує» жертві щось корисне або проявляє послугу, щоб викликати підсвідоме почуття вдячності, а згодом просить про відповідну поступку. Прикладом може бути розсилка «безкоштовного інструменту для безпеки» чи вигідної поради (насправді – зі шкідливим ПЗ), після чого жертві пропонується встановити цей інструмент або надати свої дані для отримання обіцяної вигоди.

Принцип соціального доказу або соціального підтвердження полягає в тому, що люди схильні орієнтуватися на дії та реакції інших, особливо в неоднозначних ситуаціях. Зловмисники використовують це, створюючи оманливе враження, ніби «всі так роблять» або, що бажана дія вже підтримана групою людей, рівних жертві.

Завоювання довіри – ключовий аспект складніших соціально-інженерних схем, таких як confidence scam. На відміну від швидких атак через страх чи терміновість, тут шахрай може тривалий час підтримувати спілкування, демонструючи уважність і компетентність, аби жертва перестала його підозрювати. Наприклад, аферист видає себе за фінансового консультанта або

технічного експерта, спочатку надає дрібні корисні поради або допомогу (необов'язково шкідливі), поступово отримує дрібні підтвердження довіри від жертви. Згодом, міцно «ввійшовши в довіру», такий соціальний інженер може безперешкодно просити конфіденційні дані або гроші, оскільки жертва більше не очікує обману з його боку. Ефект довіри часто виступає мультиплікатором для інших маніпуляцій. Для повноти представлення різновидів психологічного впливу табл. 1.3 систематизує розширений спектр маніпулятивних стратегій, охоплюючи як уже розглянуті в підрозділі, так і додаткові механізми.

Таблиця 1.3 – Зв'язок маніпуляцій, тригера та атаки

Маніпуляція	Емоційний тригер	Сценарій атаки
Залякування (апеляція до страху)	Страх, паніка	Фішингове повідомлення з погрозою (втрата доступу, штраф тощо), що змушує негайно діяти під страхом негативних наслідків.
Створення терміновості	Поспіх, терміновість	Лист або дзвінок з ультимативною вимогою негайної дії («негайно змініть пароль», «терміново зателефонуйте»), аби жертва діяла без роздумів.
Використання авторитету	Повага до влади, покора	Шахрай видає себе за авторитетну особу (керівника, чиновника, службу підтримки) і розпоряджається виконати певні дії.
Соціальний доказ	Конформізм, стадний інстинкт	Посилання на «всі так роблять» – фальшиве підтвердження від імені колег чи інших користувачів (наприклад, повідомлення, що «інші вже виконали цю дію»), аби жертва наслідувала їх приклад.
Принцип взаємності	Обов'язок віддячити	Попередньо надана «послуга» або подарунок (фіктивна допомога, безкоштовний «бонус»), після чого жертву просять відповісти взаємністю (наприклад, надати дані чи доступ, встановити файл).
Ефект приязні	Довіра до «свого», симпатія	Встановлення неформального контакту, дружнього тону. Зловмисник вдає колегу або знайомого, використовує особисті деталі, щоб викликати симпатію і знизити пильність, перш ніж запросити дані чи послугу.
Приманка на цікавість/вигоду	Цікавість, жадібність	Привабливa пропозиція або спокуслива приманка: виграш, вигода, сенсаційна інформація. Жертву спонукають перейти за посиланням, відкрити файл чи пристрій («знайдений» USB) із цікавості або заради обіцяної вигоди.
Завоювання довіри	Довіра, відчуття безпеки	Поступове вибудовування репутації надійної особи. Довготривалий контакт або серія взаємодій, після яких жертва вже не сумнівається в чесності зловмисника і виконує його прохання.

Узагальнюючи викладене, можна стверджувати, що психологічні техніки маніпуляції є центральним компонентом соціальної інженерії, а їхня ефективність визначається поєднанням емоційних тригерів та індивідуальних характеристик користувачів. Розуміння цих закономірностей є критичним кроком на шляху до створення більш стійких систем кібербезпеки та мінімізації впливу людського фактору.

Наведений у розділі матеріал дозволяє зробити висновок, що соціальна інженерія перетворилася на один із ключових та найбільш результативних інструментів сучасної кіберзлочинності. Статистичні дані FBI IC3, Verizon DBIR, IBM, ENISA та інших аналітичних центрів демонструють не лише стабільне зростання кількості інцидентів, пов’язаних із фішингом, BEC, вішингом та іншими видами маніпулятивних атак, але й вкрай високий рівень їх економічної ефективності. Фішинг і надалі залишається домінуючим початковим вектором проникнення, тоді як BEC та інвестиційні шахрайства формують непропорційно великі фінансові збитки, що вимірюються мільярдами доларів щороку. Середня вартість одного інциденту, пов’язаного із CI, є співставною або вищою, ніж у випадку експлуатації суто технічних вразливостей, що підкреслює системний характер загрози.

Паралельно з кількісним зростанням атак відбувається якісна еволюція кримінальної інфраструктури, що їх підтримує. Даркнет сформував повноцінну економіку соціальної інженерії, де фішингові набори, викрадені облікові дані, OSINT-послуги, deepfake-інструменти та повноцінні «Phishing-as-a-Service» моделі продаються й орендуються за ринковими правилами. Надзвичайно високий показник рентабельності, зокрема для BEC та масових фішингових кампаній, у поєднанні з можливістю масштабування та автоматизації (включно з використанням генеративного ШІ), робить CI одним із найбільш привабливих напрямів кіберзлочинної діяльності.

Разом з тим, аналіз теорій людських помилок і когнітивних вразливостей показує, що у центрі успіху CI лежить не стільки технічна складова, скільки системна експлуатація людського фактору. Модель «швейцарського сиру» Дж. Різона та її розвиток у вигляді HFACS демонструють, що успішні атаки

зазвичай є наслідком поєднання активних та латентних помилок – від індивідуальних дій окремого співробітника до глибинних організаційних недоліків, таких як слабка культура безпеки, дефіцит ресурсів або формальне ставлення до політик. Теорія когнітивних упереджень пояснює, як авторитет, соціальний доказ, терміновість, оптимізм, FOMO та інші ефекти систематично спотворюють сприйняття ризиків і ухвалення рішень. Теорія планованої поведінки, у свою чергу, показує, що, навіть, наявність знань про загрози не гарантує безпечної поведінки, якщо не змінені суб'єктивні норми та відчуття контролю над ситуацією. Феномен нормалізації відхилення додатково підкреслює, що небезпечні практики можуть непомітно закріплюватися всередині організації і роками не сприйматися як загроза, доки не стануть ланкою у ланцюгу катастрофічного інциденту.

Окремий пласт аналізу стосується психологічних механізмів маніпуляції та їхньої варіативності залежно від демографічних, особистісних та професійних характеристик користувачів. Продемонстровано, що різні групи користувачів (за віком, статтю, рівнем освіти, родом діяльності) мають відмінні профілі вразливості, а соціальні інженери адаптують свої сценарії під конкретні цільові аудиторії. Страх, терміновість, апеляція до авторитету, соціальний доказ, взаємність, приязнь, цікавість та вигода виступають базовими емоційними тригерами, на яких будуються переважна більшість атак, тоді як завоювання довіри дозволяє вибудовувати тривалі, високоризикові шахрайські схеми. Систематизація цих прийомів демонструє, що СІ є не хаотичним набором трюків, а добре структурованою системою психологічного впливу.

У сукупності розглянуті статистичні дані, економічні параметри, теоретичні моделі людських помилок та класифікація маніпулятивних технік дають підстави стверджувати, що соціальна інженерія є одночасно технічною, організаційною та психологічною проблемою. Отже, побудова ефективних механізмів протидії неможлива без інтегрованого підходу, який враховує не лише вдосконалення технічних засобів захисту, а й глибинну роботу з людським фактором, що створює підґрунтя для подальшого дослідження.

2 АНАЛІЗ ІНДИКАТОРІВ СОЦІАЛЬНО-ІНЖЕНЕРНИХ АТАК ТА ПОВЕДІНКОВИХ МЕХАНІЗМІВ ВИЯВЛЕННЯ ЗАГРОЗ

2.1 Логіка побудови атаки соціальної інженерії в реальному середовищі

У реальних умовах атака СІ розгортається поетапно, утворюючи свій «kill chain» (рис. 2.1) – послідовність дій зловмисника від планування до досягнення мети. Узагальнена модель життєвого циклу СІ-атаки включає шість основних етапів: розвідка, встановлення первинного контакту, залучення, спонукання жертви до потрібних дій, виконання та використання отриманого доступу.

Рисунок 2.1 – Модель життєвого циклу [23]

На початковому етапі зловмисник визначає ціль атаки та формує максимально повний інформаційний профіль. Ефективність соціальної інженерії значною мірою залежить від підготовки, оскільки атака фактично починається задовго до першого контакту з жертвою – з аналізу її оточення, посадових обов'язків, поведінкових звичок та можливих слабких місць. Для цього використовуються методи збору відкритих даних: соціальні мережі,

корпоративні ресурси, загальнодоступні матеріали, попередні витоки інформації. Шахраї детально вивчають особисті та професійні профілі, комунікаційні зв'язки та інтереси жертви, щоб сформувати максимально правдоподібний сценарій майбутньої взаємодії. Одночасно визначаються ключові психологічні важелі – страх, владність, цікавість, жадібність, – які можуть забезпечити високий рівень впливу. На основі отриманих даних готується реалістична легенда та обирається відповідний канал взаємодії.

У наступній фазі зловмисник ініціює контакт із ціллю. Оптимальний канал комунікації обирається залежно від звичної поведінки жертви, а саме електронна пошта, телефон, месенджери або особиста взаємодія. Основна мета – створити перше враження легітимності та не викликати підозри. Завдяки попередньо зібраним даним зловмисник може видавати себе за співробітника іншого відділу, технічну підтримку, бізнес-партнера або клієнта – залежно від того, який образ найкраще відповідає очікуванням жертви. Навіть перше повідомлення може містити штучно створену терміновість або формальність, що підсилює правдоподібність історії.

На етапі розвитку взаємодії зловмисник цілеспрямовано будує довіру. Він демонструє професійну обізнаність, дружність або просить про невелику допомогу – залежно від того, як саме сформовано легенду. Часто використовуються прийоми справжньої «вбудованості» у контекст жертви. Згадки про внутрішню інформацію, нібито спільні події минулого, знайомі імена, спільні інтереси створюють стійку ілюзію легітимності запиту. З погляду психології, тут працюють механізми симпатії, соціального доказу та ефект знайомства. У корпоративних середовищах поширеною технікою є підробка або продовження наявного ланцюжка службової переписки, що створює відчуття повної нормальності взаємодії. Основне завдання цієї фази – створити в жертви відчуття безпеки й комфорту, щоб будь-які подальші запити з боку зловмисника сприймалися як природні.

Далі настає момент, коли зловмисник починає безпосередньо тиснути на жертв – центральна стадія атаки, під час якої довіра, сформована раніше, використовується як інструмент примусу. Успішні прийоми включають

створення терміновості, нагнітання страху або викликання сильних емоцій. Зловмисник може повідомити про нібито негайну загрозу, таку як блокування облікового запису, збій у системі, ризик для колеги або для роботи, що змушує жертву діяти імпульсивно. Такі повідомлення апелюють до інстинкту уникнення втрат та знижують здатність критично оцінювати ситуацію. Інша стратегія – гра на співчутті чи відповідальності. Також активно використовуються когнітивні упередження: автоматизм дій, ефект авторитету, ефект дефіциту. Так, масові атаки із QR-кодами використовують рутинність поведінки співробітників, а обманні листи від імені керівництва створюють тиск службового обов'язку.

Після цього жертва здійснює потрібну зловмиснику дію. Можливий перехід за шкідливим посиланням, відкриття зараженого файлу, передавання конфіденційної інформації (паролів, кодів підтвердження), здійснення фінансової транзакції або використання зовнішнього носія. На цьому етапі досягається головна мета атаки. З технічної точки зору саме ця фаза часто формує найбільше цифрових артефактів: переходи за підозрілими URL, відкриття файлів, виклики шкідливих процесів, введення даних на сторонніх сторінках. Багато SOC-рішень можуть реєструвати такі події та формувати сигнали тривоги, однак робити це потрібно дуже швидко – у реальних інцидентах час між переглядом фішингового листа та введенням даних часто становить лише кілька десятків секунд.

Завершальна стадія – закріплення успіху атаки та отримання вигоди. Якщо ціллю було викрадення даних, відбувається їхнє виведення за периметр, таких як завантаження баз клієнтів, конфіденційних документів або архівів. Якщо метою був доступ до систем – може здійснюватися встановлення бекдору або виконання деструктивних дій, таких як шифрування або видалення інформації. У випадку фінансових атак кошти остаточно переводяться на рахунки злочинців. На цьому етапі зловмисник зазвичай уже діє автономно, без додаткової взаємодії з жертвою. Технічно ця фаза характеризується різкими аномаліями. Саме на фінальній стадії атака найчастіше стає видимою для захисних систем – але в більшості випадків уже запізно.

З точки зору виявлення в корпоративних системах, ранні фази є найменш помітними, оскільки проходять у соціальному або зовнішньому інформаційному просторі. Дії зловмисника в ці моменти або не фіксуються технічно, або виглядають як звичайна активність користувачів та співробітників. Більшої видимості набуває проміжна та фінальна активність. На останніх етапах події добре лягають у загальні моделі поведінки зловмисника та можуть бути виявлені системами моніторингу, засобами аналізу поведінки користувачів та мережевими датчиками. Якщо початкові кроки соціальної інженерії є здебільшого «невидимими», то кінцеві кроки вже нагадують класичний кіберінцидент, який залишає багатий технічний слід для подальшого аналізу.

На підтвердження попередньо викладеного тексту доречно навести приклади синтетичні дані, які можуть відображатися у технічних журналах і артефактах. До першого прикладу відноситься фішингова атака із завантаженням шкідливого ПЗ. У ході дії цієї атаки жертва отримує лист із вкладенням «Invoice_3022.docm», відкриває його, макрос завантажує бекдор. Отримані артефакти у журналі поштового шлюзу – запис про доставку листа з зовнішньої адреси, зі збігом по фішинговому шаблону. В журналах проксі-сервера – вихідне HTTP-з'єднання з ПК користувача до підозрілого хоста (наприклад: http://malicious-site.xyz/load.exe). На станції в Sysmon/EDR – подія запуску процесу winword.exe – створення дочірнього процесу powershell.exe зі спробою завантажити виконуваний файл. Такі журнали чітко вказують на етап Дія, спровокований CI-атакою: користувач активував шкідливий контент.

У другому прикладі BEC-атака, де зловмисник видає себе за CFO та просить переслати конфіденційний фінансовий звіт. Жертва відповідає та надсилає документ. На поштовому сервері артефакт відображатиметься, як ланцюжок листування, де вихідна адреса CFO ледь відрізняється (oleksandr.maks@companya.com замість справжнього oleksandr.maks@company.com) – індикатор імперсонації домену. У заголовках email – відсутність SPF/DKIM збігу для домену відправника (технічний IoC). Вміст листа містить нетипову лексику для даного керівника (аналітика стилю могла б це виявити) та нагальні формулювання: «Конфіденційно – відправити

негайно». Після цього, у фаєрволі можлива подія зовнішнього підключення до FTP, куди було завантажено викрадений звіт (ексфільтрація). У цьому сценарії маніпуляція авторитету + терміновості проявилася через текст листів, а компрометація – через передачу даних.

Отже, найбільш «видимими» фазами соціально-інженерної атаки є дія та ексфільтрація. Саме на них варто робити акцент при побудові систем моніторингу – виявляти аномалії користувацької активності, нестандартні переходи, логіни, передачу даних. Ранішні фази потребують розвитку проактивних та навчальних заходів, а також аналізу непрямих ознак.

2.2 Індикатори атак IoC та IoA, що базуються на психологічних маніпуляціях

У попередньому розділі було розглянуто психологічні механізми, які лежать в основі маніпулятивних дій зловмисників, зокрема вплив емоційного тиску, принципів переконання та когнітивних упереджень. На даному етапі основну увагу приділено тому, як ці маніпулятивні впливи відображаються у технічних артефактах інформаційних систем. У межах такого аналізу розрізняють два типи індикаторів: IoC, що позначають уже здійснену компрометацію, наприклад, при виявлені фішингових сайтів чи гешів шкідливих файлів; та IoA, які вказують на активність зловмисника ще до настання негативних наслідків і відображають поведінкові аномалії, характерні для соціально-інженерних атак [24].

У контексті електронної пошти, яка залишається основним каналом розповсюдження CI-атак, індикатори умовно поділяються на елементи заголовків листа та вміст повідомлення. До технічних ознак належать невідповідність між полями «From» і «Reply-To», відсутність коректної SPF/DKIM-верифікації або нетипові маршрути доставки. У текстовому наповненні значущими індикаторами є використання словосполучень, що створюють штучну терміновість, надмірна емоційність, граматичні порушення чи стилістична невідповідність характерній манері автора. Час надсилання повідомлення також може виступати ознакою атаки, зокрема коли лист

надходить у нетиповий період, що не відповідає звичайному робочому циклу адресата. Серед поведінкових IoC виділяють шкідливі вкладення (файли зі скриптами або виконуваними компонентами) та посилання на маловідомі чи масковані домени.

У смішингу та повідомленнях у месенджерах індикаторами виступають аномальні номери відправника, надмірно короткі та емоційно насичені формулювання, а також використання скорочених URL-посилань. Додаткову підозру викликають щойно створені облікові записи, що маскуються під існуючі контакти, та синхронна розсилка однотипних повідомлень кільком співробітникам.

У голосових атаках технічні ознаки обмежені можливостями телекомунікаційної інфраструктури, однак можуть включати повторювані виклики з невідомих або аномальних номерів, а також характерний зв’язок між дзвінком і подальшими діями користувача в IT-системах. У таких випадках індикатором стає не сам дзвінок, а послідовність подій, таких як зміна пароля, встановлення ПЗ віддаленого доступу або підтвердження MFA одразу після розмови.

Багатоканальні CI-кампанії характеризуються часовою синхронністю подій у різних каналах. Індикаторами тут виступають взаємопов’язані повідомлення (лист + дзвінок), групові розсилки або узгодженість сценаріїв, що потребують кореляційного аналізу між різнорідними журналами подій. Сукупність таких сигналів вказує на скоординований характер атаки та підвищує ймовірність її успішної ідентифікації.

Загалом наведені індикатори демонструють, що хоча соціальна інженерія базується на психологічному впливі, її реалізація неминуче породжує специфічні цифрові прояви, які можуть бути використані для побудови правил детекції та поведінкових моделей реагування. Саме ці аспекти стають основою подальшого аналізу.

На основі аналізу індикаторів у багатьох відомих інцидентах, можна скласти профіль їх частоти. Наприклад, рис. 2.2 відображає умовну heatmap

частоти появи різних типів IoC/IoA у вибірці зі 100 фішингових інцидентів зі звітів Verizon, APWG, IBM та ін.).

Рисунок 2.2 – Heatmap частоти різних індикаторів CI-атак у вибірці інцидентів

З графіка помітно, що всі атаки з використанням психологічних маніпуляцій залишають по собі певні аномальні сліди. Однак розподіл цих індикаторів відрізняється. Масові кампанії дають більше «грубих» IoC, а цільові атаки – тонші IoA. Виявлення і кореляція індикаторів є завданням систем кібербезпеки, що підводить нас до наступного розділу – аналізу поведінкових патернів користувачів та підходів до детекції, де буде показано, як технічні індикатори поєднуються з поведінковими для ухвалення рішень щодо загрози.

2.3 Аналіз поведінкових патернів користувачів (UEBA/UBA)

Підхід UEBA (User and Entity Behavior Analytics) [25] ґрунтується на формуванні статистичної моделі звичної поведінки користувача та подальшому виявленні відхилень, що можуть свідчити про соціально-інженерний вплив. Його логіка базується на трьох послідовних елементах: базовій лінії, аномалії та ризиковій події.

На етапі побудови базової лінії формується поведінковий профіль, у якому фіксуються типові патерни роботи конкретної особи: часові інтервали входу до систем, використовувані пристрої, частота переходів за посиланнями, характер

електронного листування та інші стабільні ознаки. Наприклад, для окремого співробітника нормою може бути робота в межах робочого дня, мінімальна взаємодія з зовнішніми доменами та стабільний набір дій у пошті.

Аномалією вважається будь-яка дія, що суттєво відхиляється від усталеного патерну. Якщо користувач, який зазвичай не взаємодіє з зовнішніми листами, пізно ввечері переходить за невідомим посиланням або завантажує неспецифічні файли, така поведінка автоматично отримує підвищений рівень аномальності. Чим більше вона віддалена від базової моделі, тим вищий ризиковий бал присвоює система.

Ризиковою подією вважається аномалія, пов’язана з типовими ознаками, як перехід за незнайомими посиланнями, введенням облікових даних на новому ресурсі, взаємодією з підозрілими контактами або ненормальними викликами. Таким чином UEBA перетворює поведінкові відхилення на сигнали можливих IoA, навіть у тих випадках, коли традиційні засоби ще не здатні класифікувати дію як компрометаційну.

Важливим компонентом такого аналізу є врахування індивідуальних та демографічних особливостей користувачів. Для різних категорій працівників характерні різні моделі активності, і їхнє порівняння дає змогу виявляти порушення симетрії. Він також дає змогу зафіксувати відхилення на рівні функціональної ролі. Поведінковий профіль користувача може включати часові особливості логінів, характер взаємодії з листами, типові обсяги доступу до внутрішніх систем та рівень залучення до зовнішніх комунікацій.

Подальший аналіз SOC може встановити зв’язок таких відхилень із попередньою маніпуляцією. У такому випадку UEBA дає змогу своєчасно виявити ланцюг подій, які окремо могли б залишитися непоміченими. Його застосування також корисне для оцінювання впливу упереджень поведінки. Система може фіксувати не лише технічні, а й опосередковані прояви маніпуляції, такі як різке прискорення реакції на листи від осіб із високим «авторитетним» значенням або підвищену вразливість у певні часові періоди, що корелюють із усталеними звичками користувача.

Як приклад UEBA-поведінкового профілю можна навести користувача «Alex Doe», співробітника відділу продажів віком 37 років. За результатами місячного спостереження його типова активність характеризується наступними ознаками:

- робочі входи в систему у будні дні з 9:00 до 18:00 з офісної мережі;
- періодичні підключення через VPN з дому 1-2 рази на тиждень близько 20:00;
- переважно внутрішнє листування з приблизно п’ятьма зовнішніми листами на день;
- переходи за 10-12 посиланнями на тиждень виключно на домени відомих партнерів;
- середній час відкриття листа після отримання близько 10 хвилин;
- відсутність використання адміністративних привілеїв.

В один із днів фіксується відхилення від цього профілю: о 23:15 здійснено вхід через VPN з нового, раніше не зареєстрованого пристрою, виконано масове завантаження даних із CRM та відправлено електронний лист на зовнішню адресу *@gmail.com із вкладеним CSV-файлом; о 23:45 від імені користувача надсилаються повідомлення кільком колегам із нетиповим проханням перейти за посиланням.

У процесі UEBA-аналізу система фіксує низку аномалій:

- вхід у нетиповий час із нового пристрою (ризиковий бал 80/100);
- масове завантаження даних, нехарактерне для попередньої поведінки (90/100);
- перша зафіксована передача даних на зовнішню адресу (95/100);
- розсилка нестандартних листів колегам (85/100);

Сукупний ризиковий бал досягає критичного рівня, формується сповіщення про ймовірну компрометацію облікового запису внаслідок соціально-інженерної атаки. Подальший аналіз з боку підтверджує, що напередодні користувач отримав лист із підробленою сторінкою авторизації, ввів

облікові дані, після чого сесія була перехоплена зловмисником, а інцидент локалізовано.

Загалом UEBA доповнює традиційні механізми виявлення, оскільки дозволяє пов’язати окремі технічні ознаки з повноцінною поведінковою картиною, забезпечуючи раннє виявлення як індикаторів компрометації, так і індикаторів самої атаки.

2.4 Методи виявлення соціально-інженерних атак

Протидія атакам потребує спеціалізованих підходів до виявлення, оскільки вони поєднують технічний та поведінкові виміри. У цьому підрозділі увага зосереджується на методах детекції, тобто на розпізнаванні атак у режимі реального або наближеного часу з позиції захисника. До основних класів належать аналіз текстових і контентних патернів, виявлення аномальної поведінки користувачів, мультимодальний підхід, застосування методів машинного навчання, порівняння їх ефективності та інтегровані системи, які приймають рішення на основі сукупності індикаторів [26].

Аналіз текстових і контентних патернів є одним із найдавніших, але все ще актуальних підходів до протидії фішингу. Він охоплює аналіз тексту повідомлень та URL-адрес на предмет підозрілих характеристик і значно еволюціонував порівняно з примітивним пошуком окремих ключових слів на кшталт «password». Ентропійний аналіз URL ґрунтується на припущенні, що фішингові посилання часто містять довгі випадкові рядки або нетипові структури доменів, тому обчислюється ентропія доменного сегмента чи шляху і, у разі перевищення порогу, посилання розглядається як потенційно згенероване. У дослідженнях Yamana Lab (Waseda Univ.) запропоновано методику оцінювання ентропії нелатинських символів в URL для виявлення фішингу, що обходить класичні детектори через використання Unicode-символів. Водночас складний або довгий URL не завжди означає атаку, наприклад, посилання на Google Drive, тому така ознака комбінується з репутаційною інформацією про домен та час його реєстрації.

N-gram аналіз тексту використовується для виявлення стилістичних аномалій. Будується модель легітимних корпоративних листів, що описується частотами біграм і триграм слів або символів, а далі оцінюється, наскільки розподіл n-грам у новому повідомленні відхиляється від звичного. Значні відхилення, наприклад, надмір уживання слів «urgent», «wire transfer» у контексті, де вони зазвичай майже не зустрічаються, можуть вказувати на фішинговий характер повідомлення. Подібна логіка застосовується до URL, коли аналізуються послідовності символів у доменних іменах для виявлення typosquatting, наприклад, відмінність між paypal та paypa1 на один символ. Техніка виявлення аномалій за n-грамами добре відома у системах IDS і придатна для фільтрації фішингового трафіку.

Stylometry або аналіз стилю є більш просунутим напрямом, що використовує такі ознаки, як середня довжина речень, характерні формули привітання й підпису, структура тексту та форматування. У випадку BEC-атак, де зловмисник отримує доступ до реальної поштової скриньки і продовжує листування, можна зіставляти стиль нових повідомлень з історичними. Якщо від імені керівника з'являється лист із проханням здійснити фінансовий переказ, але у тексті відсутні характерні для нього елементи підпису або змінено звичні мовні обороти, це може слугувати сигналом. Деякі рішення аналізують увесь ланцюжок листування і виявляють точки, де різко змінюється тон, структура або мова спілкування.

Поєднання URL reputation та лексичного аналізу дозволяє одночасно оперувати базами відомих шкідливих посилань і оцінювати нові. URL розкладається на токени, наприклад, login-office365.notreal.ua у вигляді [login, office365, notreal], після чого зіставляється з легітимними доменами. Якщо «office365» з'являється в домені, що не належить Microsoft, це є сильною ознакою фішингу. Окремо перевіряється схожість доменів із відомими брендами за відстанями Левенштейна чи Jaro-Winkler, наприклад, company.com з кириличною «а» проти справжнього company.com.

ML-класифікація контенту є логічним продовженням цих підходів. Сучасні рішення часто використовують моделі, які навчаються на корпусах

фішингових та легітимних листів і виявляють складні, нелінійні поєднання ознак. Класичні алгоритми SVM та Random Forest у дослідженнях середини 2010-х років демонстрували точність на рівні 90-99% на тестових наборах, зокрема для класифікації фішингових URL проти нормальних. Сьогодні дедалі ширше застосовуються нейронні мережі, зокрема CNN та LSTM для аналізу тексту, а також градієнтний бустинг, який працює з комплексом ознак, що включає заголовки, основний текст і характеристики вкладень.

Контент-аналіз має низку суттєвих переваг. Він не залежить від профілю конкретного користувача, може працювати вже при першому контакті та здатен блокувати явно шкідливі листи ще до будь-якої взаємодії з ними. Водночас його обмеження пов’язані з новими або високореалістичними патернами атак, коли текст не містить очевидних підозрілих елементів, а також із адаптацією зловмисників, які переходять до надто коротких або повністю графічних повідомлень. Обробка таких випадків потребує або додаткових технологій, наприклад OCR у реальному часі, або комбінування з іншими методами, тож контент-аналіз розглядається як необхідний, але недостатній елемент захисту.

Виявлення аномальної поведінки ґрунтується на поведінковому аналізі і, як показано в підрозділі 2.3, дозволяє фіксувати непрямі ознаки атак, спираючись на зміни у діях користувачів. З практичної точки зору застосовуються декілька груп методів. Правила на основі часу та контексту реалізують прості експертні евристики, наприклад, виявлення входів у позаробочий час із нетипових локацій, що може свідчити про компрометацію MFA через vishing, або фіксація ситуацій, коли користувач клацає на посилання менш ніж через 30 секунд після відкриття листа. До таких правил також належать сценарії, у яких протягом короткого проміжку приходять email та SMS із однаковим кодом чи посиланням, що вказує на багатоканальну атаку.

Статистичні моделі та пороги будуються на індивідуальному baseline. Для кожного користувача розраховується середня кількість зовнішніх посилань, за якими він переходить протягом тижня. Якщо цей показник раптово збільшується, наприклад, удвічі, подія позначається як тригер. Аналогічно, якщо медіанний час відповіді для користувача зазвичай становить 1 годину, а у

конкретному випадку він відповідає на підозрілий лист за 2 хвилини, це розглядається як аномалія. Неконтрольовані методи ML, такі як Isolation Forest або автоенкодери, дозволяють виявляти відхилення у багатовимірних профілях активності без попереднього маркування даних. Isolation Forest ізолює рідкісні, «не схожі на інших» точки за сукупністю метрик, включаючи час входу, IP-адресу, кількість переходів за URL і обсяги переданих даних. Автоенкодери навчаються відтворювати нормальну поведінку, а значне зростання похибки реконструкції сигналізує про аномальну послідовність дій.

На практиці поведінковий аналіз застосовується у різних сценаріях. Час доби використовується як простий, але дієвий критерій: введення корпоративного пароля о 3-й годині ночі на сторонньому сайті вважається нетиповим, тож фіксація запиту до login.microsoftonline.com через proxy або DNS-логи в нестандартний час може стати підставою для додаткової перевірки автентичності. Тип пристрою також виступає індикатором, якщо користувач зазвичай працює лише з Windows-ПК, а система раптом бачить авторизацію з MacOS чи Android, що може свідчити про використання іншого середовища зловмисником. Системи EDR та MDM здатні виявляти підключення нових незареєстрованих пристроїв до облікових записів, а в рамках Zero Trust-політик такі події часто одразу блокуються.

Швидкість реакції є ще одним сигналом. Надто швидке відкриття вкладення після отримання листа може трактуватися як поведінкова вразливість, а також як індикатор атаки. Поштові клієнти із відповідними розширеннями можуть у таких ситуаціях виводити попередження. Довгострокові відхилення фіксуються у вигляді user risk scoring, коли співробітник, який роками не мав інцидентів, раптом двічі за місяць потрапляє на фішинг або двічі допускає зараження своєї станції, що стає підставою для додаткового навчання. Поведінкові методи добре працюють щодо невідомих раніше атак, однак чутливі до якості базової моделі і можуть породжувати значну кількість хибних спрацьовувань, якщо користувач змінює стиль роботи з легітимних причин. Тому часто використовують послідовність «спочатку контент, потім поведінка»

і активніше реагують тоді, коли контент здається безпечним, але поведінка навколо нього є підозрілою.

Мультимодальна або багатоканальна детекція пов’язана з тим, що сучасні соціальні інженери комбінують e-mail, телефон, SMS та соціальні мережі. Традиційно ці канали контролюються окремо, що ускладнює побудову цілісної картини. Інтегровані платформи XDR/SOAR агрегують дані з поштових серверів, телефонних систем, SMS-шлюзів і журналів SOC та застосовують правила кореляції.

Окремим перспективним напрямом є зіставлення контенту між каналами, коли однаковий код підтвердження або домен з’являється в SMS і email, а також використання NLP для збагачення таких зіставлень. Аналогічно, захисні стратегії можуть використовувати другий канал для верифікації підозрілих запитів, наприклад, out-of-band підтвердження фінансових доручень. Загалом багатоканальна детекція перебуває у фазі активного розвитку, а міжнародні організації, такі як ENISA, відзначають необхідність посилення моніторингу voice-phishing та smishing, особливо з огляду на зростання відповідних атак протягом останніх 2 років.

Методи машинного навчання для детекції застосовуються практично на всіх рівнях. Контрольоване навчання передбачає наявність великої навчальної вибірки «доброякісне проти шкідливого», у випадку фішингу це корпуси email-повідомлень, URL та поведінкових послідовностей. Алгоритм SVM формує гіперплощину, яка розділяє фішингові й нефішингові приклади за обраними ознаками, такими як наявність певних слів, довжина URL, кількість одержувачів листа, наявність вкладень. У роботах 2018 року для класифікації фішингових URL у поєднанні лексичних ознак та характеристик домену було досягнуто точності близько 99%. Random Forest дає змогу одночасно враховувати велику кількість різнорідних ознак і оцінювати їхню важливість, у тому числі поєднуючи контентні, поведінкові та загрозливі сигнали. Такі моделі демонструють високу точність на знайомих шаблонах, однак потребують регулярного оновлення через зміну тактик зловмисників і явище data drift [27, 28].

Неконтрольовані методи, зокрема Isolation Forest, кластеризація та автоенкодери, дозволяють виявляти невідомі раніше аномалії, але зазвичай супроводжуються підвищеним рівнем false positive. Типовий сценарій, коли Isolation Forest позначає 5% найаномальніших подій, серед яких, умовно, лише 0,5% становлять реальні інциденти, що призводить до 4,5% зайвих сповіщень і потребує комбінування з іншими сигналами.

Великі мовні моделі (LLM) застосовуються для глибинного аналізу текстового контенту email і чатів. Вони здатні інтерпретувати сенс повідомлення, ідентифікувати приховані прохання здійснити певні дії та виявляти непрямі ознаки обману. Модель може, наприклад, розпізнати, що лист містить запит на підтвердження облікового запису через перехід за посиланням і оцінити його як фішинговий. Проте генеративні моделі можуть припускатися помилок, їх можливо обдурити спеціально підготовленими текстами, а самі зловмисники використовують ШІ для створення більш переконливих і стилістично бездоганних листів. Тому LLM доцільно інтегрувати в багатокомпонентні системи та не покладатися виключно на їхні рішення.

Порівняльні оцінки методів детекції, наведені у відкритих дослідженнях [1-4], показують, що контент-аналіз на основі правил і regex забезпечує середню точність на рівні 70-80% із низьким рівнем recall та невисоким відсотком хибних спрацьовувань, URL-репутація дає високу точність для відомих загроз (понад 90%) при низькому рівні FP, класичні ML-моделі SVM і RF досягають 95-99% точності на тестових вибірках, але баланс між recall і FP залежить від порогів прийняття рішень. Behavioral UEBA, за оцінками, дозволяє знизити ймовірність пропуску складних атак приблизно на 20-30%, а впровадження ШІ-аналітики в окремих звітах пов’язується зі скороченням середнього часу виявлення інцидентів на десятки днів і економією в мільйони доларів. Комбіновані рішення, такі як Microsoft 365 Defender, декларують блокування 99,9% загроз, однак навіть 0,1% від 100 млн. листів становлять 100 тисяч повідомлень, які проходять перший каркас захисту і частково фільтруються вже поведінковими засобами. Згідно з Verizon DBIR 2024, близько 11% користувачів, які натиснули на фішингове посилання, згодом самостійно повідомили про підозрілий лист, що

демонструє обмеженість суто технічних і поведінкових методів та підкреслює незамінну роль поінформованого користувача як останньої лінії оборони.

У другому розділі магістерської роботи було розглянуто логіку побудови соціально-інженерних атак, їх життєвий цикл, а також основні технічні і поведінкові індикатори, що допомагають виявляти загрози. Соціально-інженерні атаки розгортаються у кілька фаз – від розвідки через встановлення контакту і побудову довіри до безпосередніх дій і ексфільтрації даних. Ранній етап атаки часто не фіксується традиційними технічними засобами і відбувається переважно у соціальному просторі, тоді як пізні етапи з імітацією класичних кіберінцидентів залишають численні цифрові артефакти, які можна виявити системами моніторингу.

Важливим є розмежування індикаторів компрометації, які свідчать про вже здійснені атаки, і індикаторів активності, що утворюють сигнали раннього попередження, і часто виявляються поведінковими аномаліями користувачів. Підхід UEBA реалізує аналіз поведінкових патернів, формує індивідуальні профілі ко

ристувачів і визначає аномалії у режимі реального часу.

Методи машинного навчання, такі як SVM, Random Forest та нейронні мережі, забезпечують високу точність класифікації. Однак вони потребують постійного оновлення та уваги до явища data drift, а також інтеграції з поведінковими та багатоканальними системами.

Загальний тренд у сучасній кібербезпеці – комплексне поєднання технічних методів, поведінкового аналізу, мультимодальної кореляції та освіти користувачів як останньої лінії оборони. Саме така стратегія дозволяє ефективно протистояти соціально-інженерним атакам, що використовують психологічні маніпуляції та складні сценарії.

3 МЕТОДИКИ ПІДВИЩЕННЯ СТІЙКОСТІ

3.1 Розробка експериментальних фішингових листів

Мета експериментального дослідження полягала у розробці та аналізі реалістичних фішингових електронних листів для емпіричної оцінки ефективності різних психологічних маніпуляцій серед різних цільових груп користувачів. Основна гіпотеза передбачала, що результативність психологічної маніпуляції у фішингових повідомленнях залежить від адекватного узгодження конкретної психологічної техніки з демографічними та професійними характеристиками цільової аудиторії. Іншими словами, кожна група користувачів найвразливіша до певних методів впливу, і належне поєднання техніки та аудиторії буде давати максимальний ефект. При цьому додатково очікувалося, що комбінування кількох технік в одному сценарії може посилювати загальний вплив експоненційно, що було зазначено у розділі 1.4.2 [22].

Дизайн експерименту було побудовано за факторіальною схемою 3×5. Було визначено три незалежні сценарії атак, що різнилися між собою контекстом і цільовою аудиторією, та п’ять варіантів фішингових листів у межах кожного сценарію, причому в кожному варіанті домінувала своя основна психологічна техніка (додаток А).

Незалежними змінними виступали ключові психологічні техніки соціальної інженерії, що були докладно розглянуті у розділі 1.2. Залежна змінна визначалася як ефективність маніпулятивного впливу, операціоналізована через комплексний показник успішності листа.

Для вимірювання цього показника враховувалося наступне:

- наявність явних лінгвістичних маркерів обраної техніки;
- стилістична і мовна відповідність листа автентичній діловій комунікації;
- наявність конкретних і правдоподібних елементів, що підвищують реалізм сценарію;

- психологічна інтенсивність емоційного впливу повідомлення на отримувача.

Крім того, було введено декілька контрольних змінних, аби гарантувати коректність проведення експерименту:

- візуальний брендинг та форматування;
- тон спілкування;
- коректні технічні деталі;
- реалістична послідовність подій.

Відповідно до гіпотези та дизайну, на основі аналізу актуальних кіберзагроз (розділ 1.1) [1] було відібрано три типові сценарії фішингових атак з різними цільовими аудиторіями. Враховуючи ці тенденції, а також відмінності у вразливостях різних груп користувачів (табл. 1.2) для експерименту обрано такі сценарії, що охоплюють як корпоративний сегмент, так і масову аудиторію.

Вибір сценаріїв ґрунтувався на статистиці кіберзлочинності з розділу 1.

Сценарій 1 – фішинг на фінансових працівників був спрямований на фахівців віком 30-50 років, які працюють із регуляторними вимогами та мають середньо-високу цифрову грамотність. Враховуючи їхню підвищену чутливість до авторитету державних установ, основними техніками стали авторитет + терміновість, а також вигода у вигляді податкових пільг. Такий підхід логічний, оскільки фінансові відділи належать до найчастіших і найзбитковіших цілей кіберзлочинців.

Сценарій 2 – CEO Fraud для менеджерів середньої ланки використовував три групи тригерів: авторитет керівництва + конфіденційність, особиста вигода та імітація попередньої комунікації.

Сценарій 3 – масовий фішинг для молодих користувачів соціальних мереж був побудований на техніках FOMO, соціального доказу, персоналізації та обіцянки популярності.

Після визначення трьох репрезентативних сценаріїв доцільним є систематизований аналіз психологічних механізмів, які лежать в основі фішингових атак (табл. 3.1).

Таблиця 3.1 – Психологічні техніки та лінгвістичні реалізації

Психологічна техніка	Операціоналізація в листі	Лінгвістичні маркери	Приклад маркування	Сценарії з найвищою ефективністю
Страх	Чітко визначена загроза, зазначення часових меж та можливих наслідків	ВИКОРИСТАННЯ CAPS, слова «НЕГАЙНО», «НАЗАВЖДИ», «КРИТИЧНО», емоджі	«всі операції будуть АВТОМАТИЧНО ЗАБЛОКОВАНІ»	S1V1, S2V1, S3V1
Авторитет	Посилання на офіційні інституції, нормативні акти та посадових осіб	«@organization.gov.ua», «Закон України №…», формальні титули «Директор», «Міністр»	«Міністерство фінансів України», «Департамент кібербезпеки НБУ»	S1V1–V2, S2V2, S2V5
Терміновість	Встановлення конкретних дедлайнів, кількісних обмежень, використання лічильників часу	«до 16:00», «залишилось 47 місць», «6 годин 47 хвилин», «сьогодні»	«У ТЕБЕ ЗАЛИШИЛОСЬ: 6 ГОДИН 47 ХВИЛИН»	S1V1, S2V1, S3V1–V5
Вигода	Пропозиції матеріальної вигоди, точні числові обіцянки	Грошові суми «5€M», «2,4 млн ₴», відсотки «на 150%», емоджі	«5,000,000 євро інвестиції», «бонус 35,000–80,000 грн»	S1V2, S2V2–V3, S3V2–V5
Соціальний доказ	Згадування реальних імен, компаній, кількості людей, які «вже виконали дію»	Імена «Марина К.», компанії «Епіцентр», числа «347», використання галочок ✓	«✓ Марина Коваль – підтвердила дані ✓»	S1V4, S2V4, S3V3–V4
Завоювання довіри	Створення враження попереднього спілкування, персоналізація, конкретні дати	Префікси «Re: Re:», дати «25.11.2024», неформальний тон, часткове використання імені	«Дякую за швидку відповідь вчора»	S1V3, S2V3, S3V4
FOMO + Дефіцит	Наголошення на обмеженій кількості ресурсів, коротких часових рамках та посилення ефекту через соціальний доказ	«47 місць залишилось», «останній шанс», «847 людей чекають», таймери	«Залишилось лише 23 місця! Ваші колеги вже зареєструвалися»	S3V2, S3V5

Кожна техніка впливу, що застосовується у відповідних листах, реалізується через конкретні лінгвістичні маркери, композиційні рішення та стилістичні елементи. Саме операціоналізація таких технік – тобто перехід від психологічного тригера до конкретного мовного патерну – дозволяє оцінити, які прийоми виявились найефективнішими у межах кожного сценарію.

Для подальшої оцінки ефективності сценаріїв було здійснено порівняльний аналіз листів, що входять до кожного варіанта (табл. 3.2, рис. 3.1).

Таблиця 3.2 – Характеристики фішингових листів

ID листа	Сценарій	Основна техніка	Доп. техніка	Довжина листа	Довжина пропозиції	Тип CTA	Складність
S1V1	Фішинг фінанс.	Страх	Терміновість	287	18	Посилання	Низька
S1V2	Фішинг фінанс.	Авторитет	Вигода	324	21	Вкладення	Середня
S1V3	Фішинг фінанс.	Довіра	Реалістичність	298	19	Посилання	Низька
S1V4	Фішинг фінанс.	Соц. доказ	Приязнь	312	20	Посилання	Низька
S1V5	Фішинг фінанс.	Комбінована атака	Всі техніки	451	28	Посилання + SMS	Висока
S2V1	CEO Fraud	Страх	Терміновість	156	10	Посилання (реквізити)	Низька
S2V2	CEO Fraud	Авторитет	Вигода	287	17	Посилання	Низька
S2V3	CEO Fraud	Завоювання довіри	Реалістичність	315	19	Посилання	Низька
S2V4	CEO Fraud	Соціальний доказ	Приязнь	289	18	Посилання	Низька
S2V5	CEO Fraud	Комбінована атака	Всі техніки	387	24	Посилання	Середня
S3V1	Молодь	Страх	Терміновість	198	12	Посилання	Низька
S3V2	Молодь	Цікавість	Вигода	267	16	Посилання	Низька
S3V3	Молодь	Соціальний доказ	Приязнь	245	15	Посилання	Низька
S3V4	Молодь	Завоювання довіри	Реалістичність	276	17	Посилання	Низька
S3V5	Молодь	Комбінована атака	Всі техніки	398	25	Посилання + SMS	Висока

Доцільність такого аналізу полягає у тому, що результат впливу соціально-інженерного повідомлення визначається не лише психологічною технікою, але й композиційною побудовою, обсягом тексту, характером заклику до дії та технічними елементами, які ускладнюють або спрощують виконання шкідливих інструкцій. Відповідно, у таблиці узагальнено характеристики кожного листа: його сценарну належність, домінантну та допоміжну техніку впливу, довжину, тип заклику до дії та рівень технічної складності. Подібне структурування дозволяє визначити, які комбінації параметрів найчастіше зустрічаються в успішних атаках і які з них є найбільш ризиковими для різних аудиторій.

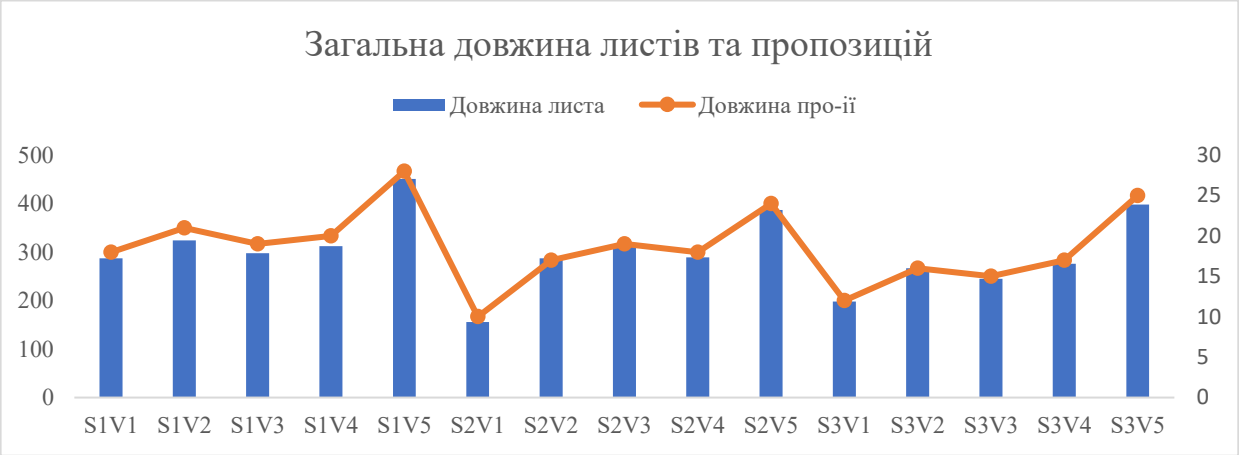

Рисунок 3.1 – Діаграмне представлення даних

Середній об’єм листів, орієнтованих на фінансових працівників, становив 314 слів з варіативністю у межах від 287 до 451 слів. Для менеджерів середньої ланки цей показник був дещо нижчим – 284 слова при діапазоні 156-387 слів. Листи, спрямовані на молодіжну аудиторію, виявилися найкоротшими, із середнім значенням 277 слів та коливанням 198-398 слів.

Основні труднощі, що виникли під час розроблення реалістичних фішингових листів, можна згрупувати навколо трьох моментів.

По-перше, одним із центральних викликів стало забезпечення балансу між правдоподібністю та маніпулятивністю. Необхідно було інтегрувати виразні психологічні техніки таким чином, щоб листи викликали довіру адресата і водночас містили достатньо сильні тригери для впливу. Оптимальним підходом

виявилося рівномірне «розподілення» маніпулятивних сигналів між різними елементами структури повідомлення – темою, підписом, контактами, формулюванням CTA – замість надмірної концентрації у одному фрагменті.

По-друге, складністю становила операціоналізація психологічних явищ, які за своєю природою є тонкими та контекстно залежними. Найбільш чутливою до цього виявилася техніка «завоювання довіри», оскільки вона вимагає відтворення притаманних корпоративній комунікації інтонацій, стилістики та композиційних патернів. Для досягнення цієї мети було проведено аналіз автентичних листів організацій і відтворено характерні структурні та лінгвістичні особливості.

По-третє, важливою задачею стала персоналізація контенту без створення враження витоку персональних даних. Листи мали містити релевантні згадки про типову діяльність, посади чи імена, але без конкретних ідентифікаторів, здатних скомпрометувати приватність респондентів. Реалізація полягала у використання узагальнених, типових для цільових груп імен та посад, що забезпечило водночас персоналізований ефект і відповідність етичним обмеженням.

3.2 Виявлення поведінкових індикаторів атак у корпоративних журналах

Розроблені у попередньому підрозділі фішингові листи стали фундаментальною основою для подальшого етапу роботи, зокрема для аналітичних процедур, пов’язаних із виявленням ознак атаки у технічних журналах. Саме їхня контрольована структура та систематичне варіювання маніпулятивних елементів дозволяють перейти від моделювання соціально-інженерних впливів до формування формальних критеріїв виявлення скомпрометованих користувачів та IoA у поведінкових і лінгвістичних даних. Усі листи містять чітко задокументовані технічні й стилістичні ознаки, що дає змогу використовувати їх як еталонний набір під час побудови та оцінювання алгоритмів, спрямованих на технічне розпізнавання атак.

Основні завдання експериментальної частини включають:

- визначення оптимальних меж виявлення аномалій, які дозволяють мінімізувати помилкові спрацювання при збереженні високої чутливості;

- порівняльне оцінювання різних обчислювальних підходів щодо їх придатності для виявлення окремих класів IoA;
- валідація каталогу поведінкових індикаторів на реальних даних;
- кількісна оцінка часу виявлення від появи ознак аномалії до їх детекції системою.

Для експериментального дослідження було обрано BETH Dataset (Behavioral Engagement Threat Hunting) [31] – відкритий, анотований набір даних, сформований кафедрою комп'ютерної безпеки та призначений для аналізу аномалій у Linux-системах. Датасет містить 1,14 млн. подій, що поділені на тренувальну, валідаційну та тестову множини у пропорції 66,8% / 16,6% / 16,6%, що забезпечує репрезентативність нормальної поведінки та коректність подальшої оцінки моделей. До складу даних входять системні, мережеві, аудиторські та сесійні журнали, зібрані протягом розширеного часового періоду, що дозволяє охопити різноманітні поведінкові патерни.

Для кожного запису було сформовано комплекс ознак, що відображає часовий контекст, активність процесів та індикатори потенційної компрометації користувача чи системи. Часові характеристики включали абсолютні та відносні часові мітки, зокрема hour (для виявлення нетипової активності у проміжку «02:00-06:00», раніше визначеному як IoA003) та dayofweek для аналізу подій поза робочими годинами.

До групи ознак процесної активності увійшли processId, parentProcessId, threadId, а також argsNum, значення якого виявляє кореляцію з компрометованими сесіями. Ознаки помилкової поведінки моделювалися через returnValue, бінарний індикатор hasError, а також похідну метрику errorRate, яка відображає частку неуспішних викликів. Категоріальні поля – processName, eventName, hostName – були закодовані для подальшого використання в алгоритмах машинного навчання. Нарешті, мітки sus та evil слугували показниками підозрілої та фактичної компрометованої активності, де evil використовувалася як цільова функція моделі.

На основі вихідних журналів було сформовано агреговані часові профілі, що узагальнюють поведінку кожного окремого процесу. Для процесу P такий профіль містить: загальну кількість системних викликів, кількість унікальних потоків, сукупність невдалих викликів, середню кількість аргументів, різноманітність типів подій, частоту позначень sus, а також похідні метрики errorRate, suspiciousRate та evilRate.

Отримані дев'ятимірні профілі утворюють компактний аналітичний простір, у якому кожна точка відповідає поведінці конкретного процесу. Саме ці представлення використовувалися як вхідні дані для методів виявлення аномалій (Isolation Forest, LOF) та побудови базових метрик відхилення (z-score).

Для формалізації критеріїв виявлення шкідливої активності було створено узагальнений перелік IoA (табл. 3.3), кожен з яких відповідає певному типу поведінкових або технічних відхилень у журналах системи. Таблиця охоплює індикатори, що базуються на різних джерелах даних – від системних викликів і журналів процесів до часових міток, командних рядків та агрегованих профілів. Кожен IoA має унікальний ідентифікатор, логічне правило спрацювання, типовий приклад індикації та рівень критичності, що визначає його значущість у контексті компрометації.

Таблиця 3.3 – Опис IoA

IoA ID	Назва індикатора	Тип логу	Логічне правило	Приклад індикації	Критичність
IoA001	Масована активність процесу	Process Log	countEvents > 1000	DDoS-подібні патерни	MEDIUM
IoA002	Аномально висока частка помилок	System Log	errorRate > 0.3	Невдалі доступи	MEDIUM
IoA003	Нетипова нічна активність	Timestamp Log	hour ∈ 2-6 AND countEvents > 500	Нічні завантаження	HIGH
IoA004	Паралельне спрацювання sus та evil	Process Log	suspiciousCount > 0 AND evil > 0	Підтверджена компрометація	HIGH
IoA005	Аномальна кількість потоків	Thread Log	countThreads > 50	Масове паралельне виконання	MEDIUM

Продовження таблиці 3.3

IoA006	Критичні помилки системних викликів	System Call Log	errorRate > 0.5	Привілейовані спроби	HIGH
IoA007	Складні командні рядки	Command Line Log	avgArgsPerEvent > 20	Шифрування, обфускація	MEDIUM
IoA008	Різноманітність системних подій	Event Log	uniqueEventTypes > 30	Reconnaissance	MEDIUM
IoA009	Підозріла частка подій	Suspicious Flag Log	suspiciousRate > 0.5	>50% позначено як sus	HIGH
IoA010	Композитна аномалія	Multiple Logs	errorRate > 0.4 AND countEvents > 800 AND suspiciousCount > 50	Потрійна аномалія	HIGH
IoA011	Zero-day-подібна поведінка	Evil Flag Log	evilRate ≠ 0	Рідкісні/невідомі патерни	CRITICAL
IoA012	Статистичне багатовимірне відхилення	Anomaly Log	z-score > 2.5	Сукупне відхилення профілю	LOW

Представлені індикатори IoA охоплюють ключові поведінкові та технічні аномалії, що можуть свідчити про компрометацію системи або виконання шкідливих дій. IoA001–IoA003 фіксують нетипово високу інтенсивність подій та активність у нічний час, характерну для автоматизованих або прихованих атак. IoA004-IoA006 відображають поєднання підозрілих міток та аномальну частку помилкових системних викликів, що є ознаками несанкціонованих операцій. Індикатори IoA007-IoA009 стосуються складних командних рядків, надлишкової різноманітності подій та стійкої підозрілої поведінки. Композитний індикатор IoA010 поєднує одразу декілька критичних відхилень, підвищуючи точність виявлення загрози. IoA011 відповідає за zero-day-подібні аномалії, у яких навіть невелика кількість записів з міткою evil сигналізує про значний ризик. Нарешті, IoA012 визначає статистичні багатовимірні відхилення профілю процесу від норми та слугує базовим механізмом раннього попередження.

Експериментальна порівняльна оцінка охопила чотири методи, обрані за їхню релевантність до задачі виявлення аномалій у поведінкових журналах.

Перший підхід, статистична нормалізація z-score [32], ґрунтується на багатовимірному оцінюванні відхилень від середнього значення і вважає

аномальними ті профілі процесів, у яких максимальне абсолютне значення нормалізованої ознаки перевищує 2.5. Метод є простим і детермінованим, має низьку обчислювальну вартість та добре працює для кількісних ознак, проте його обмеженням є припущення нормальності розподілу та нечутливість до локальної структури даних.

Логічним продовженням стала перевірка можливостей Isolation Forest – ансамблевого алгоритму, що ізолює точки за допомогою випадкових розділень та трактує як аномальні ті, які потребують мінімальної глибини ізоляції. Він не робить статистичних припущень, легко масштабується й стійкий до високої розмірності, але є менш інтерпретабельним і залежним від вибору параметра contamination, що визначає очікувану частку аномалій [33].

Для зіставлення машинних методів з інженерними підходами було застосовано rule-based детекцію, яка використовує каталог IoA як систему логічних правил. Для кожного запису обчислюється сумарний бал на основі спрацьованих індикаторів, і за перевищення встановленого порогу подія вважається аномальною. Такий підхід забезпечує повну прозорість рішень та повну відповідність бізнес-логіці, однак залежить від якості й повноти каталогу IoA та не адаптується автоматично до нових типів атак [34].

Останній метод – LOF у поєднанні з попередньою z-score нормалізацією – дозволяє виявляти як глобальні, так і локальні аномалії завдяки аналізу щільності сусідства у нормалізованому просторі. Попри здатність уловлювати як поодинокі, так і кластери відхилень, він має вищі обчислювальні вимоги й потребує налаштування *k* параметрів та порогових значень [35].

Сукупне використання цих підходів сформувало широку основу для порівняння результатів, де статистичний baseline виконував роль орієнтиру, Isolation Forest демонстрував потенціал сучасних машинних алгоритмів, rule-based підхід відтворював виробничу практику, а LOF забезпечував аналіз локальної структури даних, що разом дозволило комплексно оцінити ефективність виявлення IoA.

Параметри моделей було підібрано таким чином, щоб забезпечити відтворюваність результатів і коректне порівняння різних підходів.

Для статистичного baseline застосовано поріг z-score, рівний 2.5, що відповідає ймовірності появи події приблизно 1.24%, а самі статистичні характеристики – середнє значення та стандартне відхилення – розраховано виключно на тренувальній вибірці.

У методі Isolation Forest використано 100 дерев із параметром contamination, встановленим на рівні 0.2, що означає очікувану частку аномалій приблизно 20%; для забезпечення репродукованості задано фіксоване значення random_state=42, а max_samples обмежено 256 записами на дерево.

Rule-based детекція ґрунтувалася на вагових коефіцієнтах для окремих IoA (4 для критичних, 3 для високих, 2 для середніх та 1 для низьких індикаторів), а поріг активації встановлено як сумарний бал не менше, ніж 3, що дозволяє спрацювати принаймні одному індикатору високої критичності; для комбінованих правил застосовувалися дробові коефіцієнти.

Гібридний метод LOF у поєднанні із z-score базувався на виборі двадцяти найближчих сусідів у нормалізованому просторі, пороговому значенні LOF=1.3 та додатковому м’якому z-score порозі 1.5, що підвищувало чутливість на складних кластерах; режим novelty було активовано для коректної роботи з новими даними без повторного переналаштування моделі.

Оцінювання ефективності кожного методу проводилося на основі стандартних метрик класифікації на тестовому наборі.

Показник precision характеризував частку коректно виявлених аномалій серед усіх спрацьовувань моделі та був особливо важливим для зменшення навантаження на SOC-аналітиків. Метрика recall відображала здатність моделі виявляти реальні аномалії й була критичною з позиції недопущення пропуску атак. Гармонійне середнє precision та recall – F1-score – використовувалося як основний індикатор збалансованості моделі. Додатково аналізувалася площа під ROC-кривою, яка дає змогу оцінити здатність моделі відрізняти аномальні записи від нормальних незалежно від обраного порогу. Для детальнішого розуміння роботи окремих індикаторів було побудовано матриці помилок і обчислено TPR та FPR для кожного IoA окремо.

Серед розглянутих підходів статистичний baseline продемонстрував найбільш збалансовану роботу, забезпечивши найвищу точність виявлення аномалій за рахунок поєднання високого precision і майже максимального recall (рис. 3.2), що відобразилося у найкращому F1-score і найнижчому рівні помилкових спрацювань, що в поєднанні з коротким часом обробки робить baseline ефективною відправною точкою для систем раннього попередження (рис. 3.3).

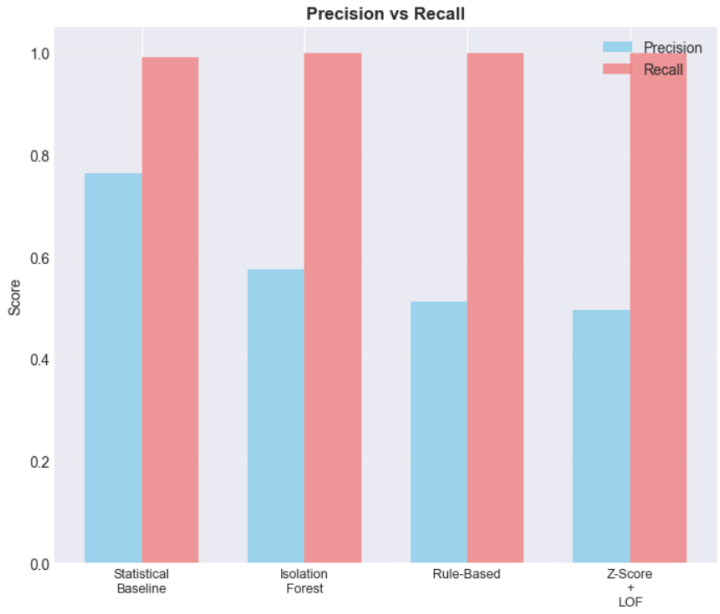

Рисунок 3.2 – Порівняння підходів

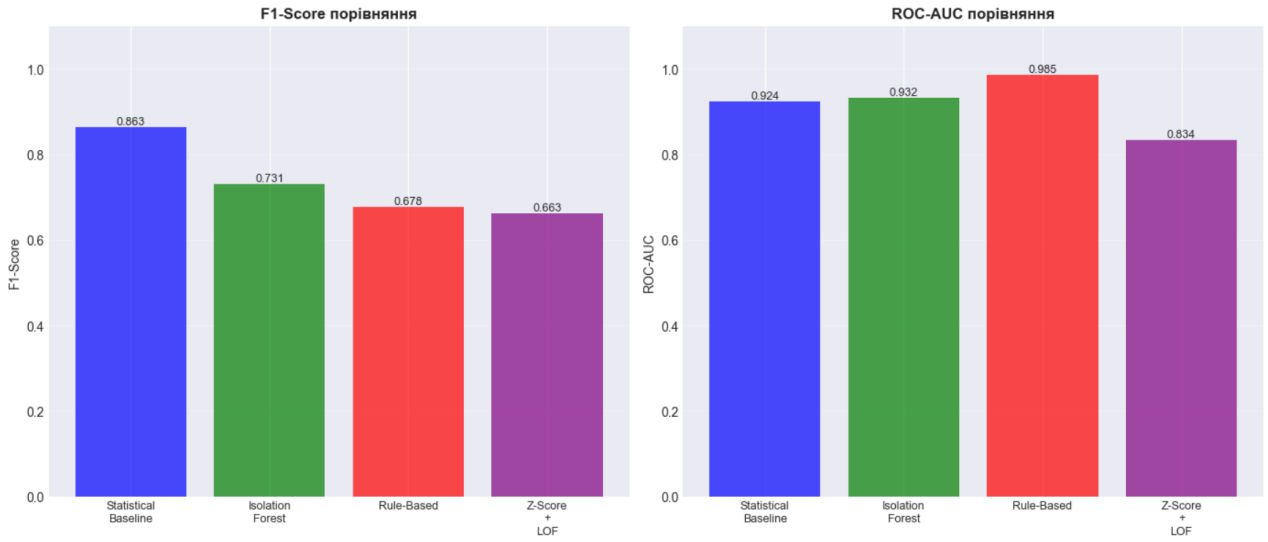

Рисунок 3.3 – F1 та ROC-AUC

Isolation Forest, хоч і виявив усі аномалії, досягнув цього ціною істотного збільшення FPR та зниження precision, однак водночас продемонстрував ROC-

AUC, близький до статистичної моделі, що підтверджує його здатність до якісного ранжування подій (рис. 3.4). Rule-based підхід також забезпечив повний recall і продемонстрував найвищу ROC-криву серед моделей, проте виявився менш точним у перетворенні спрацьовувань на валідні аномалії, що призвело до збільшеного FPR. LOF-модель виявилася найменш ефективною через зниження precision і найвищий FPR, а також істотно більший час обробки, пов'язаний із пошуком найближчих сусідів у багатовимірному просторі.

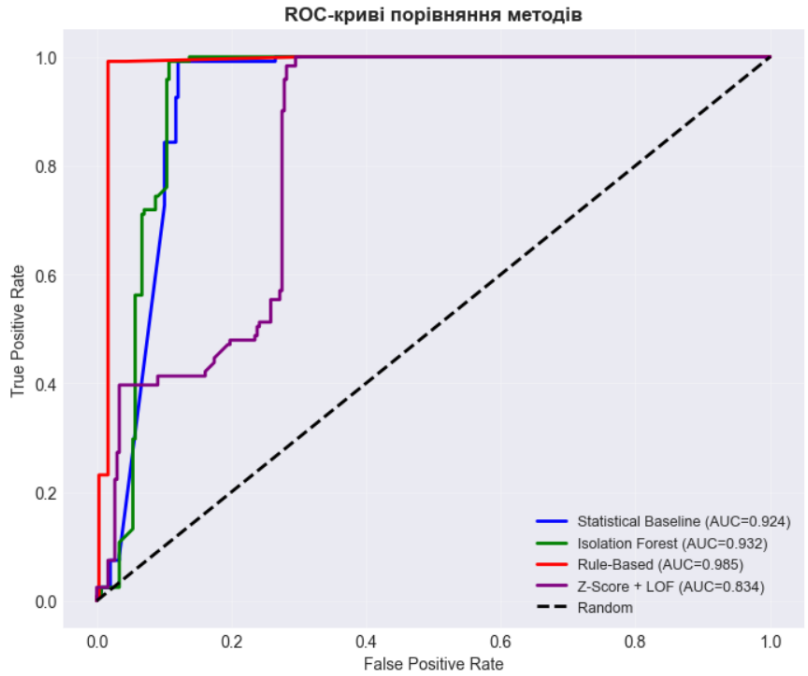

Рисунок 3.4 – ROC-криві порівняння методів

Детальніший аналіз за окремими IoA показав, що моделі найкраще обробляють індикатори з чітко визначеними граничними умовами. Зокрема, поєднання підозрілих і шкідливих міток формує найвиразніший сигнал, а масова активність процесу та надмірна частка підозрілих подій також дають високу якість класифікації. Натомість індикатори зі «м'якими» межами – наприклад, аномальна кількість потоків або composite-аномалії – характеризуються нижчими значеннями F1, що свідчить про складність їхнього розпізнавання на основі загальних моделей.

Аналіз часових характеристик підтвердив, що rule-based підхід є найшвидшим, за ним йдуть baseline та Isolation Forest, тоді як LOF значно поступається іншим методам через підвищену обчислювальну складність:

- Isolation Forest: середній TTD = 42 ms від появи аномалії до виявлення;
- Rule-Based: TTD = 8 ms (найшвидший);
- Statistical Baseline: TTD = 12 ms;
- Z-Score LOF: TTD = 124 ms (повільна через k-NN).

З огляду на критерій корпоративних SOC щодо часу реакції, придатними для реалізації в реальному часі є baseline, rule-based і Isolation Forest. Додаткові графіки матриць помилок (рис. 3.5) та ROC-кривих (рис. 3.4) підтвердили, що baseline демонструє найкращу дискримінацію при низьких значеннях FPR, Isolation Forest і LOF мають подібну форму ROC-кривої, однак із гіршими показниками при конкретних порогах, а rule-based модель характеризується менш диференційованою ROC-кривою через дискретність правил.

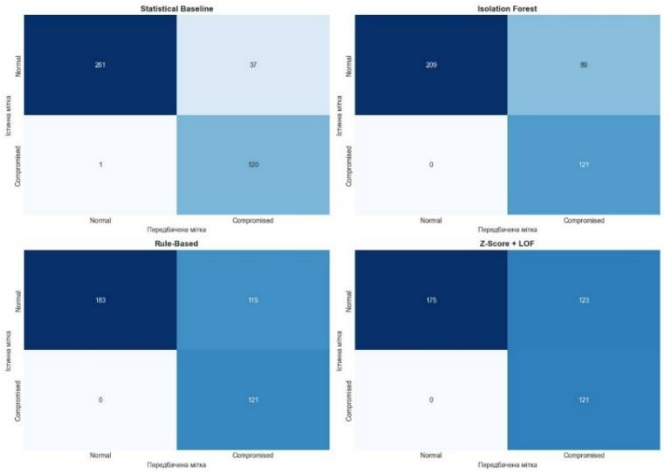

Рисунок 3.5 – Матриці помилок аналізованих методів

3.3 Розробка методик протидії атакам соціальної інженерії та підвищення стійкості організації

На основі теоретичного аналізу механізмів соціальної інженерії, проведеного в розділі 3.1, а також результатів тестування алгоритмів виявлення аномалій на датасеті BETH у розділі 3.2, у цьому розділі сформовано практичні методики підвищення стійкості організацій до фішингових кампаній і атак типу BEC. Запропонована класифікація рівнів зрілості дозволяє оцінити поточний стан захищеності, зіставити витрати та очікувану ефективність заходів і визначити пріоритетні напрями впровадження засобів протидії.

3.3.1 Класифікація рівнів стійкості організації до атак соціальної інженерії

Критичний рівень описує організації, які покладаються на вбудовані антиспам-фільтри й не здійснюють моніторингу подій. Персонал не проходить навчання, а формальні процедури реагування відсутні. Ефективність такого рівня низька, тому що більшість цілеспрямованих листів успішно обходять захист, а час виявлення інцидентів може тривати сотні днів.

Низький рівень передбачає базові сигнатурні засоби – перевірку SPF/DKIM/DMARC, фільтрацію доменів та IP-адрес, антивірусну перевірку вкладень і щорічні тренінги для співробітників. Такі рішення забезпечують кращий контроль масових фішингових атак, але залишаються неефективними проти персоналізованих і нових технік. Високий FPR у сигнатурних методах підтверджує їхню обмеженість.

Середній рівень доповнює базовий захист контекстним аналізом на основі індикаторів атак та початковим поведінковим профілюванням. Упроваджується SIEM, автоматизоване сповіщення про події та регулярні симуляції фішингу. Тестування показало, що статистичний baseline на цьому рівні забезпечує найкраще співвідношення precision і recall, зберігаючи низьку кількість хибних спрацювань.

Високий рівень інтегрує машинне навчання та мультиканальне зіставлення подій: поєднуються дані поштових систем, робочих станцій, VPN, мережевих брандмауерів і систем керування доступом. Такі рішення значно підвищують здатність виявляти складні, багатокрокові атаки. Ефективність методів типу Isolation Forest підтверджує високу чутливість до нетипових поведінкових відхилень.

Оптимальний рівень характеризується використанням адаптивних моделей штучного інтелекту, які автоматично оновлюють поведінкові профілі та пороги виявлення, а також поєднанням кількох методів у єдиному ансамблі. Організація застосовує архітектуру Zero Trust, автоматизоване реагування та розвинені можливості пошуку загроз. За такого рівня досягається мінімальний

час виявлення та високий рівень точності при низькому FPR, що робить захист найбільш стійким до складних і персоналізованих атак.

3.3.2 Навчання й підвищення обізнаності користувачів

Ефективність захисту від соціальної інженерії значною мірою залежить від рівня обізнаності персоналу та вмінь розпізнавати підозрілі повідомлення. Однією з базових методик є регулярне проведення навчальних програм та симуляційних тренінгів з імітацією реальних фішингових атак. Такі заняття включають кейси, адаптовані до галузевих сценаріїв організації, що дозволяє персоналізувати контент і підвищити релевантність навчального матеріалу. Важливим елементом є персоналізований підхід: за гіпотезою, гейміфікація й інтерактивність тренінгів можуть значно поліпшити засвоєння знань. Наприклад, впровадження конкурсів на виявлення фішингових листів чи «кібер-репутаційні рейтинги» підвищують мотивацію співробітників до вивчення проблеми.

Іншим важливим компонентом є створення відкритої культури повідомлення про підозрілі листи без страху покарання. Практика показує, що наявність простого й чіткого механізму звітування в поштовій системі, регулярна розсилка роз’яснювальних матеріалів та кейсів успішного виявлення загроз стимулюють співробітників бути пильнішими [36]. Такі заходи не тільки підвищують швидкість виявлення реальних атак за рахунок додаткового каналу повідомлень, а й сприяють формуванню культури безпеки.

3.3.3 Технічні засоби та процеси

Посилення стійкості передбачає також ширше використання технічних методів запобігання та виявлення фішингових атак. По-перше, необхідно закріпити політики автентифікації та авторизації: для критичних бізнес-процесів рекомендується обов’язкове застосування багатофакторної автентифікації, сегментація доступів за принципом Least-Privilege [37], а також своєчасне оновлення прав доступу під час внутрішніх змін, що ускладнить можливість

зловмисників використати викрадені облікові дані. По-друге, слід інтегрувати просунуті email-фільтри та антиспуфінг рішення – перевірку SPF/DKIM/DMARC у компанії на основі реальних IP-адрес надсилань і блокування аномалій.

На доповнення, розгортання аналітики лінгвістичних патернів допоможе виявляти нові цільові атаки. За гіпотезою, застосування систем машинного навчання натренованих на ознаки фішингових листів у поєднанні з правилами здатне виявляти навіть нетипові варіанти. Варто також розгорнути технологію автоматичної перевірки вкладень і URL на шкідливий зміст у режимі реального часу, а при виявленні аномалій – позначати листи і сповіщати користувача.

Не менш важливо впровадити чіткі організаційні процедури перевірки фінансових та критичних запитів. Наприклад, правило «подвійного підтвердження», де будь-який запит на зміну реквізитів або переказ грошей повинен додатково підтверджуватися за допомогою голосового зв’язку чи іншого незалежного каналу. Така практика знижує ймовірність успіху BEC. Аналогічно, рекомендовано вводити процедуру «four-eyes principle» для важливих рішень, коли два незалежних співробітники перевіряють легітимність запиту. За гіпотезою, комплекс цих процесних контролів спільно із технічними дозволяє знизити рівень успішних атак навіть при опущених помилках персоналу.

3.3.4 Використання аналітики поведінкових індикаторів

Результати підрозділу 3.2 показали, що обробка журналів подій може виявляти непрямі ознаки фішингових атак. Пропонується інтегрувати цю аналітику з email-системою організації. Наприклад, у разі підозрілих записів у журналах під час сесії певного користувача, система може автоматично знижувати довіру до отримувача фішингового листа і позначати майбутні вхідні повідомлення на додаткову перевірку. Зворотній зв’язок від подібних крос-модальних сигналів – нова гіпотеза, яка передбачає тіснішу взаємодію аналітики журналів з виявленням атак соціальної інженерії.

Також доцільно розвивати автоматизовані системи раннього попередження на базі розробленого набору IoA. Наприклад, у разі активації одного чи кількох IoA надсилається оповіщення SOC-оператору з деталями й пріоритетом, що дозволяє прискорити розслідування. Гібридний підхід – комбінування rule-based сигналів з методами машинного навчання – може забезпечити баланс між точністю і адаптивністю. За гіпотезою, подібна система, що поєднує лінгвістичну перевірку листів та поведінкові ознаки в журналах, підвищує чутливість виявлення цільових атак і знижує час реакції на інциденти.

3.3.5 Децентралізоване розповсюдження знань та тестування на ефективність

Для постійного підвищення стійкості організації корисно залучати зовнішні ресурси та спільноти. Наприклад, участь у професійних розсилках та довірених мережах обміну загрозами отримує свіжу інформацію про нові методики фішингу та дозволяє адаптувати захист. Гіпотетично, створення галузевого «фішингового альянсу», в якому кілька організацій анонімно обмінюються досвідом атак і виявленими компрометованими листами, забезпечило б більш широкий обмін ознаками та підвищення колективного імунітету до загроз.

Нарешті, введення метрик кіберстійкості – оцінка часу виявлення та нейтралізації атак, частка успішно заблокованих фішингових листів, частота повідомлень працівників про підозрілі емейли – дозволяє керівництву відстежувати прогрес і визначати пріоритетність вдосконалень. Регулярні тестування на ефективність серед підрозділів та публічні звіти про результати (без викриття інцидентів) створюють прозорий механізм постійного поліпшення.

Підсумовуючи, комбінування навчальних ініціатив, технічних бар’єрів, процедур верифікації та просунутої аналітики формує багатошаровий захисний контур. Введення інноваційних підходів (гейміфікація, міжорганізаційна кооперація, гібридні алгоритми) забезпечує новизну методик і дозволяє

організаціям поступово еволюціонувати до найвищих рівнів, наближаючись до тих, що були описані в класифікації зрілості.

Узагальнюючи, було здійснено всебічне моделювання фішингових сценаріїв із використанням психологічних технік впливу, адаптованих до конкретних типів цільових аудиторій. Отримані результати підтвердили, що ефективність фішингу значною мірою залежить від правильного поєднання маніпулятивної стратегії з характеристиками отримувача. Комбіноване використання декількох технік значно підвищує імовірність успішного впливу.

Було сформовано набір поведінкових індикаторів атак, які дозволяють на формальному рівні виявляти ознаки компрометації системи через фішинг. Експериментальне тестування різних підходів до виявлення аномалій показало, що прості статистичні моделі демонструють високу ефективність за умов правильної агрегації профілів, тоді як машинні методи забезпечують додаткову чутливість до складних або нетипових відхилень.

На основі проведених аналізів запропоновано низку практичних методик для підвищення стійкості організацій до атак соціальної інженерії. Вони охоплюють навчальні програми, технічні засоби перевірки листів, багаторівневу автентифікацію, процедурні обмеження доступу та мультиканальний моніторинг. Окрему увагу приділено створенню культури безпеки та персоналізованим підходам до підготовки співробітників.

Отримані результати підтверджують доцільність комплексного захисту, який поєднує емоційно-поведінкову профілактику, формалізовані індикатори ризику та технологічні засоби раннього виявлення.

ВИСНОВКИ

У першому розділі було проведено комплексний теоретичний огляд соціальної інженерії як виду кіберзагроз, проаналізовано її психологічні, комунікаційні та технічні аспекти. Визначено ключові техніки маніпулятивного впливу, типові моделі поведінки зловмисників та чинники, що підвищують успішність атак. Узагальнення наукових джерел та міжнародних практик дозволило встановити, що саме емоційні тригери, довіра до авторитетних джерел і неусвідомлені когнітивні упередження користувачів формують основу для більшості фішингових сценаріїв. Встановлено також, що людський фактор лишається найбільш вразливою ланкою інформаційної безпеки незалежно від рівня технічного захисту.

У другому розділі зосереджено увагу на аналітичних та поведінкових індикаторах атак, а також на технічних журналах, які дозволяють формалізувати прояви соціальної інженерії у корпоративному середовищі. Було побудовано дев’ятимірний набір індикаторів атаки (IoA), що відображає характерні відхилення в діях користувача та системи після фішингової взаємодії. Проведене порівняння алгоритмів виявлення аномалій показало, що статистичні підходи забезпечують високу точність для простих сценаріїв, тоді як машинні методи підвищують чутливість до комбінованих чи нетипових випадків. Це підтверджує важливість комплексної аналітики, яка не обмежується лише лінгвістичним аналізом листів, а охоплює поведінкові маркери і внутрішні системні події.

У третьому розділі було розроблено практичні методики підвищення стійкості до фішингових атак і маніпуляцій. Опрацьовано моделювання фішингових листів із різними техніками впливу, досліджено реакції користувачів, а також запропоновано комплекс заходів для навчання персоналу, технічного захисту та організаційної профілактики. Результати підтвердили ефективність поєднання симуляцій, гейміфікованих тренінгів, багаторівневого моніторингу та чітких процедур інформаційної безпеки. Було також показано,

що стійкість до соціальної інженерії значно зростає за умов персоналізованого навчання та системного аналізу поведінкових індикаторів.

Узагальнені результати роботи свідчать про доцільність інтеграції психологічних, технічних та організаційних підходів для формування повної моделі протидії соціальній інженерії. Побудовані моделі, структуровані індикатори ризику та запропоновані методики навчання доводять, що найвищого рівня безпеки можна досягти лише за рахунок багаторівневого захисту, що поєднує аналіз контенту, поведінкову аналітику, стійкі фреймворки та підготовку персоналу. Отримані результати можуть бути використані для вдосконалення політик інформаційної безпеки, проведення тренінгів, розробки систем виявлення та реагування, а також як практична основа для подальших досліджень у сфері соціальної інженерії та поведінкової кібербезпеки.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. FBI Internet Crime Complaint Center. (2024). Internet crime report 2024. Режим доступу: https://www.ic3.gov/.

2. Verizon. (2025). 2025 data breach investigations report. Режим доступу: https://www.verizon.com/business/resources/reports/.

3. IBM 2025 Cybersecurity Report: Credential Theft Skyrockets. Режим доступу: https://cyberinsurancenews.org/ibm-2025-cybersecurity-report/.

4. Cost of a Data Breach Report 2025. Режим доступу: https://www.ibm.com/reports/data-breach.

5. Phishing Activity Trends Reports. Режим доступу: https://apwg.org/trendsreports.

6. New Egress report reveals Millennials are the key target, as AI, Quishing, and Multi-Channel attacks top phishing trends. Режим доступу: https://www.egress.com/newsroom/new-egress-report-reveals-millennials-are-the-key-target-as-ai-quishing-and-multi-channel-attacks-top-phishing-trends.

7. European Union Agency for Cybersecurity (ENISA). (2025). Threat landscape for supply chains – July 2024 - June 2025. Режим доступу: https://www.enisa.europa.eu/.

8. MITRE ATT&CK. (2024). Enterprise tactics and techniques: Initial access via phishing. Режим доступу: https://attack.mitre.org/.

9. Chainalysis. (2025, June 4). $2.2 Billion Stolen in Crypto in 2024 but Hacked Volumes Decline. Режим доступу: https://www.chainalysis.com/blog/crypto-hacking-stolen-funds-2025/.

10. Quorum Cyber. (2025). What We Know About Ransomware Group Scattered Spider. Режим доступу: https://www.quorumcyber.com/insights/scattered-spider/.

11. KeepnetLabs. (2024). Annual phishing report 2024: Dark net ecosystem analysis. Режим доступу: https://www.keepnetlabs.com/.

12. Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. Journal of the Operational Research Society, 41(12), 1047–1057.

13. Shappell, S. A., & Wiegmann, D. A. (2000). The Human Factors Analysis and Classification System (HFACS). Режим доступу: https://skybrary.aero/articles/human-factors-analysis-and-classification-system-hfacs.

14. Greene, E. (2020). Cognitive Biases and How They Affect Security Behavior. Journal of Cybersecurity and Information Integrity, 5(1), 45–56.

15. Milgram, S. (1963). Behavioral study of obedience. Journal of Abnormal and Social Psychology, 67(4), 371–378.

16. Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Threats, threats, and stability. Self and Identity, 7(2), 213–220.

17. Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), Action control: From cognition to behavior (pp. 11–39). Springer-Verlag.

18. Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211.

19. Vaughan, D. (1996). The Challenger launch decision: Risky technology, culture, and deviance at NASA. University of Chicago Press.

20. Singer, S., & Vogt, D. S. (2011). Normalization of deviance in organizations: conceptualizations and implementation. In Human factors in organizational design and management (pp. 195–203).

21. Статистика соцмереж в Україні у 2025 році. Режим доступу: https://www.sober-ads.kyiv.ua/blog/statistic-kviten-2025.

22. Звіт Департаменту кіберполіції Національної поліції України (2025) – звіт охоплює тенденції кібершахрайств (фішинг, вішинг), зловживання в інтернеті, вікові та соціальні аспекти ризиків. Режим доступу: https://cyberpolice.gov.ua/news/zvitpro-diyalnist-departamentu-kiberpolicziyi-naczionalnoyi-policziyi-ukrayiny-u--roczi-7074/.

23. Datami. (2025). Як змінюється фішинг у 2025 році – аналіз. Режим доступу: ttps://datami.ee/ua/blog/how-phishing-is-changing-in-2025-analysis-by-datami/.

24. Cyber kill chain. Режим доступу: https://upload.wikimedia.org/wikipedia/commons/1/1d/Intrusion_Kill_Chain_-_v2.png.

25. SANS Institute. (2023). Behavioral Analytics: Using Indicators of Attack to Detect Threats.

26. Режим доступу: https://www.sans.org/white-papers/behavioural-analytics-indicators-of-attack-detect-threats/.

27. ESKA. (2024). Системи аналізу поведінки користувачів та сутностей. Режим доступу: https://eska.global/solutions/ueba.

28. Symantec Enterprise Security. (2024). Using Indicators of Attack (IoA) to Detect Advanced Threats. Режим доступу: https://www.symantec.com/content/dam/symantec/docs/white-papers/aioi-indicators-of-attack-wp-en.pdf.

29. Verma, A., & Wan, S. (2018). Phishing URL detection with lexical and host-based features. Proceedings of the 2018 IEEE International Conference on Intelligence and Security Informatics, 169-174.

30. Alshehri, M. D., Alkhammash, M. A., & Alghamdi, M. A. (2020). Phishing Detection Using Random Forest Classifier and Feature Analysis. IEEE Access, 8, 196965-196973.

31. BETH Dataset. Режим доступу: https://www.kaggle.com/datasets/katehighnam/beth-dataset

32. Z-Score Normalization: Definition and Examples. Режим доступу: https://www.geeksforgeeks.org/data-analysis/z-score-normalization-definition-and-examples

33. Liu, Fei Tony, Ting, Kai Ming, & Zhou, Zhi-Hua. Isolation Forest. ICDM 2008. Режим доступу: https://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/icdm08b.pdf

34. Demystifying the Role of Rule-based Detection in AI Systems for Windows Malware Detection. Режим доступу: https://arxiv.org/html/2508.09652v1

35. M. M. Breunig, et al. LOF: Identifying Density-Based Local Outliers. SIGMOD 2000. Режим доступу: https://dl.acm.org/doi/pdf/10.1145/342009.335388

36. Information security culture and phishing-reporting model: structural equivalence across Germany, UK, and USA. Режим доступу: https://academic.oup.com/cybersecurity/article/11/1/tyaf011/8128837

37. National Institute of Standards and Technology. (2020). Security and Privacy Controls for Information Systems and Organizations (NIST SP 800-53 Rev. 5). U.S. Department of Commerce. Режим доступу: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf

ДОДАТОК А

РОЗРОБЛЕНІ ФІШИНГОВІ ЛИСТИ

S1V1

FROM: **security@nbu-verify.com.ua**

TO: **finansist@company.com**

SUBJECT: НЕГАЙНО! Блокування корпоративних рахунків згідно з Інструкцією НБУ № 34

Шановний бухгалтере,

Увага! Якщо ви не підтвердите свою особу протягом наступних 2 годин, всі операції по корпоративних рахунках будуть АВТОМАТИЧНО ЗАБЛОКОВАНІ згідно з Інструкцією НБУ № 34. Це може призвести до значних фінансових втрат і відповідальності згідно з законодавством. Натисніть на це посилання зараз, щоб уникнути блокування: [Підтвердити особу]

З повагою,

Департамент кібербезпеки НБУ

Контакт: **security@nbu.gov.ua**

S1V2

FROM: info@minfin.gov.ua

TO: accountant@company.com

SUBJECT: Нові податкові пільги від Міністерства фінансів – скористайтеся зараз

Шановний колего,

Міністерство фінансів України повідомляє про введення нових податкових ставок відповідно до Закону України № 3254-IX, які можуть знизити податкове навантаження на 12-25%. Ви можете отримати прискорене відшкодування ПДВ та бонус від держави. Для завантаження офіційної інструкції скористайтесь вкладенням у цьому листі.

З повагою,

Міністерство фінансів України

Відділ підтримки платників податків

S1V3

FROM: support@privatbank-corporate.ua

TO: finance.manager@company.com

SUBJECT: Re: Оновлення тарифів корпоративного обслуговування

Re: Пане Іване,

Дякую за швидку відповідь вчора. У відповідності до нових змін у тарифах міжнародних платежів будь ласка оновіть свої налаштування через особистий кабінет або звертайтесь до підтримки. Це допоможе уникнути затримок із переказами. Посилання на форму оновлення тут: [Оновити тариф]

З повагою,

ПриватБанк, служба корпоративного банкінгу

Телефон: +380 44 123 4567

S1V4

FROM: webinar@dpsu.gov.ua

TO: accountant@company.com

SUBJECT: Запрошення на вебінар: Податкові новації 2025 – уже 147 колег зареєструвались!

Шановний бухгалтере,

Запрошуємо вас на безкоштовний вебінар від ДПСУ щодо нових податкових норм на 2025 рік. Уже 147 ваших колег з компанії зареєструвалися і планують бути присутніми. Роз’яснення від провідних фахівців та сесія запитань-відповідей. Зареєструйтесь за посиланням: [Реєстрація]

З найкращими побажаннями,

Департамент податкової служби України

S1V5

FROM: ceo@makscompany.com

TO: manager@company.com

SUBJECT: [НЕГАЙНО] Втрата контракту через невиконання термінів – Ваше негайне втручання потрібно!

Шановний колего,

Вкрай важливо виконати платіж по контракту №23711 до 16:00 сьогодні, інакше компанія ризикує втратити ключового замовника та 15 млн грн доходу. Будь ласка, підтвердіть виконання операції негайно, дотримуючись інструкцій в доданому посиланні: [Підтвердити переказ]

З повагою,

[Підпис]

Олександр Іванов, CEO

makscompany.com

S2V2

FROM: hr@company.com

TO: manager@company.com

SUBJECT: Річний бонус за своєчасне виконання завдань – інформація від HR

Доброго дня,

Ваші високі результати у цьому році відзначені керівництвом. Ви маєте право на отримання річного бонусу. Для оформлення виплати потрібно заповнити форму за посиланням у вкладеному листі. Поспішайте, термін до 29 листопада.

HR Відділ

makscompany.com

S2V3

FROM: it-sup@company.com

TO: manager@company.com

SUBJECT: Re: Re: Питання щодо замовлення IT обладнання

Привіт,

Дякую за ваш швидкий відгук учора щодо замовлення Dell Latitude 5520. Будь ласка, підтвердіть фіналізацію угоди, натиснувши на це посилання: [Підтвердити покупку]. Нагадую про знижку 35%, акція діє до кінця місяця.

З найкращими побажаннями,

Олена Петрівна, IT support

makscompany.com

S2V4

FROM: cfo@company.com

TO: manager@company.com

SUBJECT: Колеги вже підтвердили переказ бонусів – чекаємо вашої дії

Доброго дня,

Марина К. та Андрій П. вже успішно підтвердили операції по виплаті бонусів за останній квартал. Просимо вас також надати підтвердження, використовуючи цей посилання: [Підтвердити]. У разі питань звертайтесь до мене напряму.

З повагою,

Ігор М., CFO

makscompany.com

S2V5

FROM: investor@venturecapital.eu

TO: manager@company.com

SUBJECT: Термінове рішення щодо інвестицій у розмірі €5 млн – залишається 3 години

Шановний колега,

Міжнародний інвестор Michael Johnson надіслав підтвердження переказу €5,000,000 для розширення вашого проєкту. Однак для завершення операції потрібно виконати деякі верифікаційні дії до 15:00 CET сьогодні. Якщо пропустите цей строк, інвестиція буде скасована. Натисніть тут для підтвердження: [Підтвердити інвестицію]

З повагою,

Команда MaksCompany

S3V1

FROM: security@instagram-account-verify.com

TO: user@studentmail.com

SUBJECT: Ваш Instagram акаунт буде видалено через 24 години – діяти негайно!

Привіт!

Ваш аккаунт Instagram був позначений за підозру у порушенні правил. Якщо не пройдете верифікацію протягом 24 годин, акаунт буде назавжди видалено. Перейдіть за посиланням для підтвердження особистості: [Підтвердити акаунт]

З повагою,

Команда Instagram

S3V2

FROM: support@tiktok-verification-ua.com

TO: user@studentmail.com

SUBJECT: Отримайте безкоштовну верифікацію TikTok! Обмежена пропозиція на $499 вартість

Доброго дня!

Ви були обрані для безкоштовної верифікації свого облікового запису TikTok. Програма верифікації коштує $499, але для вас абсолютно безкоштовно! Пропозиція діє 48 годин. Пройдіть за посиланням щоб скористатися: [Отримати верифікацію]

З найкращими побажаннями,

TikTok Ukraine

S3V3

FROM: promo@spotify-promo-ua.com

TO: user@studentmail.com

SUBJECT: Виграй 12 місяців Spotify Premium – 347 українських студентів вже отримали!

Привіт!

Твій друг Марина К. вже виграла 12 місяців преміум-підписки на Spotify! Ти теж можеш стати переможцем. Натисни [Отримати преміум] та приєднуйся до 347 студентів, які вже скористались цією пропозицією.

Залишайся на хвилі музики!

Команда Spotify Ukraine

S3V4

FROM: hr@rozetka-career.com

TO: user@studentmail.com

SUBJECT: Пропозиція роботи в Rozetka – підтвердження кандидатури

Доброго дня,

Дякуємо за вашу заявку на вакансію в Rozetka. Для завершення оформлення, будь ласка, підтвердіть свої дані за посиланням: [Підтвердити кандидатуру]. Для запитань звертайтеся до HR служби.

З повагою,

HR Відділ Rozetka

Телефон: +380 44 567 89 10

S3V5

ROM: giveaway@mrbeast-giveaway-final.com

TO: user@studentmail.com

SUBJECT: У ТЕБЕ ЗАЛИШИЛОСЬ: 6 ГОДИН 47 ХВИЛИН – Виграй $10,000 від MrBeast!

Привіт, bro!

Ти серед тих обраних, хто може виграти $10,000 від MrBeast! Залишилось тільки 6 годин 47 хвилин, щоб завершити участь! 73 з 100 твоїх друзів вже підтвердили участь – не пропусти свій шанс! Натисни [Учасник] для підтвердження.

Серйозно, не марнуй час!

Команда MrBeast Giveaway