ISSN 2519-2310

CS&CS Journal

KARAZIN UNIVERSITY
CLASSICS AHEAD OF TIME

2(24) 2023

COMPUTER SCIENCE AND CYBERSECURITY

КОМП'ЮТЕРНІ НАУКИ ТА КІБЕРБЕЗПЕКА

ISSN 2519-2310

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені В.Н.КАРАЗІНА
V.N. KARAZIN KHARKIV NATIONAL UNIVERSITY

КОМП'ЮТЕРНІ НАУКИ ТА КІБЕРБЕЗПЕКА
COMPUTER SCIENCE AND CYBERSECURITY
(CS&CS)

Issue2(24) 2023

Заснований 2015 року

Міжнародний електронний науково-теоретичний журнал
International electronic scientific journal

The journal publishes research articles on theoretical, scientific and technical problems of effective facilities development for computer information-communication systems and information security question based, on advanced mathematical methods, information technologies and technical means.

The journal is published every six months.
Approved for placement on the Internet by Academic Council of the Karazin Kharkiv National University (December 25, 2023, Protocol No.23).

The journal has Digital Object Identifier:**10.26565/2519-2310** (Online).

Editorial office:
V.N. Karazin Kharkiv National University
Svobody Sq., 6, office 315a, Kharkiv, 61022, Ukraine*(North building of University, 3th floor)*

E-mail: *cscsjournal@karazin.ua*
Web-page: *http://periodicals.karazin.ua/cscs(Open Journal System)*

Published articles have been internally and externally peer reviewed

В журналі публікуються наукові статті з теоретичних і науково-технічних проблем, що пов'язані зі створенням ефективних засобів комп'ютерних інформаційно-комунікаційних систем та питань захисту інформації, на основі передових математичних методів, інформаційних технологій і технічних засобів.

Журнал виходить кожні півроку.
Схвалено до розміщення в мережі Інтернет Вченою радою Харківського національного університету імені В.Н. Каразіна (25.12.2023 р., Протокол № 23).

DOI (Онлайн): **10.26565/2519-2310**.

Редакція:
Харківський національний університет імені В.Н. Каразіна
пл. Свободи, 6, офіс 315а, Харків, 61022, Україна*(Північний корпус університету, 3 поверх)*

Електронна пошта: *cscsjournal@karazin.ua*
Веб-сторінка:*http://periodicals.karazin.ua/cscs(Open Journal System)*

Опубліковані статті пройшли внутрішнє та зовнішнє рецензування.

ЗМІСТ
TABLE OF CONTENTS

DOI : 10.26565/2519-2310-2023-2-01
УДК 004.8: 342.9

ПРОБЛЕМНІ ПИТАННЯ ТЕХНОЛОГІЇ МАШИННОГО НАВЧАННЯ В ПРАВООХОРОННІЙ ДІЯЛЬНОСТІ

Дмитро Узлов[1], Володимир Струков[2], Владислав Гуділін[3], Олексій Власов[4]

[1]Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків, 61022, Україна,
e-mail: dmytro.uzlov@karazin.ua, ORCID: https://orcid.org/0000-0003-3308-424X
[2]Харківський національний університет внутрішніх справ, пр. Льва Ландау, 27, Харків, 61080, Україна,
e-mail: struk_vm@ukr.net, ORCID: https://orcid.org/0000-0003-4722-3159
[3]Харківський національний університет внутрішніх справ, пр. Льва Ландау, 27, Харків, 61080, Україна,
e-mail: vgudilin7@gmail.com, ORCID: https://orcid.org/0000-0002-3844-1448
[4]Харківський національний університет радіоелектроніки, пр. Науки, 14, Харків, 61166, Україна,
e-mail: moonreactor@gmail.com, ORCID:https://orcid.org/0000-0003-1619-0032

Надійшла до редакції 17 жовтня 2023 р. Переглянута 19 листопада 2023 р. Прийнята 20 грудня 2023 р.

Анотація:Правоохоронні органи все частіше використовують технології прогнозування та автоматизації, де основною технологією часто є застосування методів машинного навчання (ML). У статті розглядається проблема підзвітності та відповідальності правоохоронних органів і посадових осіб в контексті застосування моделей машинного навчання ML. Автори вказують, що підзвітність є ключовим елементом демократичної правоохоронної діяльності, але використання прогнозного програмного забезпечення може створювати проблеми у забезпеченні цієї підзвітності. Стаття обговорює, що застосування ML може призвести до завуалювання відповідальності та ускладнення підзвітності у «мультиагентних структурах», що об'єднують людей і обчислювальні інструменти. Особлива увагу приділяється непрозорості алгоритмів прикладних прогнозних моделей та автоматизованих систем прийняття рішень, що стає джерелом непорозумінь і обережності щодо їх використання. Автори висувають питання щодо того, як можна забезпечити ефективний контроль та повну звітність, коли ключові компоненти процесу прийняття рішень залишаються невідомими для широкої громадськості, посадових осіб та навіть розробників моделей. У статті стверджується, що важливі питання, пов'язані з моделями рішень ML, можуть бути розглянуті без детального знання алгоритму навчання, що дає змогу експертам правоохоронної діяльності, які не займаються ML, вивчати їх у формі інтелектуального контролю. Експерти, які не займаються ML, можуть і повинні переглядати навчені моделі ML. Автори надають «набір інструментів» в формі запитань про три елементи моделі прийняття рішень, які можуть бути якісно досліджені експертами, які не є спеціалістами з машинного навчання: навчальні дані, навчальна мета та антиципаційна оцінка результатів. Такий підхід розширює можливості цих експертів у вигляді об'єктивної оцінки використання моделей ML у правоохоронних завданнях. Основна ідея полягає в тому, що навіть без глибоких технічних знань експерти можуть аналізувати та переглядати моделі ML, розкриваючи їхню ефективність через призму власного досвіду. Даний підхід сприяє порозумінню використання технологій машинного навчання в рамках правоохоронної діяльності, розширюючи потенціал відповідних експертів, не пов'язаних з ML.

Ключові слова:*машинне навчання, штучний інтелект, аналіз даних.*

1.Вступ

Правоохоронні органи все частіше застосовують досягнення інформаційних технологій та штучного інтелекту, щоб намагатися передбачити події та автоматизувати обробку даних що виникають в процесі правоохоронної діяльності. У цьому правоохоронна діяльність схожа на багато інших галузей – управління авто, прогнозування погоди, вирішення заявок на кредит тощо. Прогностична аналітика підтримує управління ризиками у сфері управління безпекою [1]. Лондон, Лос-Анджелес, Мюнхен, Новий Орлеан, Філадельфія, Цюрих та Харків – це приклади міст, де поліція використовує або тестувала інтелектуальне поліцейське програмне забезпечення, яке має на меті або передбачити, де можуть статися злочини, або хто, ймовірно, вчинить злочин у майбутньому. Машинне навчання (ML) є ключовою технологією, яка лежить в основі багатьох із цих програм. Програмне забезпечення для машинного навчання може раціоналізувати трудомісткі завдання обробки даних, таких як аналіз великого обсягу документів, оприлюднених у ході розслідування, та класифікація їх за категоріями [2]. Разом з цим

підзвітність поліції викликала занепокоєння що використання моделей ML робить людей неспроможними відповідати за рішення що були прийняті на їх основі [3]. Щоб спростувати подібні занепокоєння, необхідно зробити процеси прийняття рішень, що ґрунтуються на результатах використання моделей ML, доступними для контролю.

Прогнозну правоохоронну діяльність можна розглядати як окрему техніку під ширшою парасолькою правоохоронної діяльності, керованої аналітикою (ILP). ILP виник як практична управлінська програма для прийняття рішень, щодо правоохоронних послуг на основі об'єктивного аналізу даних [4]. Систематичний збір і аналіз розвідувальних даних мають на меті підвищити як ефективність протидії злочинності, забезпечуючи як більш точне визначення цілей, так і економічну ефективність [5]. У правоохоронній діяльності з прогнозуванням, як і в ILP, аналіз і рішення централізовані та раціоналізовані; прогностична правоохоронна діяльність підкреслює об'єктивний, науковий вибір стратегій та тактик і надає перевагу централізованому, раціоналізованому прийняттю рішень на основі аналізу даних.

2.Підзвітність правоохоронних органів

Підзвітність та відповідальність правоохоронних організацій і посадових осіб є ключовим компонентом демократичної правоохоронної діяльності, і вже давно є предметом занепокоєння дослідників і практиків правоохоронних органів [6]. З точки зору позиції правоохоронних сил у демократичній системі, підзвітність може означати політичний контроль над поліцією або співпрацю між поліцією та урядом, згідно з якою поліція повинна надавати пояснення про прийнятті рішення.

Застосування прогнозного програмного забезпечення або програмного забезпечення для автоматизації для підтримки прийняття рішень може фундаментально поставити під сумнів здатність посадових осіб та організацій звітувати про процеси прийняття рішень, а також завуалювати відповідальність у «мультиагентних структурах», що складаються з людей і обчислювальних інструментів. Непрозорість «алгоритмів» прикладних прогнозних моделей або автоматизованих систем прийняття рішень залишається основною причиною занепокоєння щодо їх використання [7]. Існує занепокоєння, що алгоритми «є непрозорими» в тому сенсі, що одержувачі вихідних даних роботи алгоритму ML*(класифікація, кластеризація, прогноз)*, рідко мають конкретне уявлення про те, як і чому конкретна класифікація, кластеризація або прогноз були отримані на основі вхідних даних [7].

Коли один або більше елементів процесу прийняття рішень незрозумілі, будь-яка з вищезгаданих концепцій підзвітності ставиться під сумнів. Модель ML, як правило, вбудована в програмне забезпечення, працює як «чорна скринька», де вхідні дані *(наприклад, геопросторові дані, щодо злочинності та/чи демографії)* обробляються у вихідні дані *(наприклад, прогноз чи класифікацію)* за допомогою обчислень, які залишаються невидимими для кінцевого користувача. Незважаючи на те, що цей процес, по суті, не є незрозумілим, він практично незрозумілий для не експертів, і може зробити основу незрозумілості щодо обґрунтування прийняття рішень.

Виникають питання: як може існувати ефективний політичний контроль над прийняттям рішень, якщо ключовий компонент у формуванні прийняття рішень фактично невідомий? Як поліція може повністю звітувати про свої рішення, якщо вони частково спиралися на аналіз, який вони самі не в змозі пояснити? В цьому сенсі ***прозорість*** розглядається як частина ідеального вирішення проблем використання ML для підзвітного прийняття рішень. Для досягнення прозорості інформація має бути доступною та зрозумілою [8]. Однак це складно, коли йдеться про напівавтоматизовані інтелектуальні системи. Разом з цим, підзвітність може бути

можливою без повної прозорості *(наприклад, розкриття вихідного коду)* шляхом розробки підзвітності в програмному забезпеченні.

3. Машинне навчання і правоохоронна діяльність

Незважаючи на те, що машини вже досить давно можуть навчатися на основі даних, за останні десятиліття машини стали здатними навчатися та досягати успіху в когнітивних завданнях, таких як позначення об'єктів на зображеннях і визначення слів за звуками. Одним із технологічних застосувань цього було автоматизоване розпізнавання номерних знаків *(APNR)*. Системи *APNR*, встановлені на правоохоронних транспортних засобах, полегшили поліцейський моніторинг правопорушників. Ці розробки відбулися завдяки поєднанню нових алгоритмів навчання *(деякі розроблені з 1950-х років і раніше)*, більшій обчислювальній потужності та розробці коду для ефективного використання обчислювальної потужності машини для вирішення проблем навчання [9]. На додаток до можливості навчання когнітивним завданням, ще однією не менш важливою розробкою ML є винайдення алгоритмів навчання, які можуть наближено створювати складні функції та вибирати важливі характеристики без перенавчання моделі відносно навчальної вибірки. Ці вдосконалення алгоритму дозволили машині навчатися з наборів даних із тисячами позначених функцій, щоб вона могла вибирати функції (змінні) і функціональну форму, яка, ймовірно, добре працюватиме під час прогнозування нових зразків.

Це означає, що змінні, які використовуються в моделях машинного навчання, не обов'язково вибираються фахівцями в галузі, а скоріше самим алгоритмом машинного навчання, і що рішення приймаються не на основі теорій, розроблених людьми, а більше з точки зору того, *«що працює»* в терміни прогнозної сили ML. Не дивно, що ці нові можливості зробили моделі з машинним навчанням дедалі кориснішими для прийнятті рішень на практиці. Моделі ML були використані, наприклад, Управлінням з боротьби з серйозними шахрайствами Великобританії *(UK Serious Fraud Office)* для виявлення юридично конфіденційних матеріалів серед мільйонів розкритих документів у розслідуванні [10], а Норвезьким органом інспекції праці для прогнозування робочих місць з високим ризиком порушень для перевірки агентством.

Обговорюючи, чи використовувати ML у процесі прийняття рішень правоохоронними органами, важливо порівнювати ML не з ідеальним процесом прийняття рішень, а з прийняттям рішень людьми. Машини приймають рішення в неоптимальних середовищах на основі непереконливих, незбагненних і оманливих доказів. Щоразу, коли прийняття рішень призводить до несправедливих результатів, процеси може бути важко відстежити, і «тяжко буває просто визначити, хто повинен нести відповідальність за заподіяну шкоду» [11]. Однак це фундаментальна проблема прийняття рішень як така, а не унікальна для рішень, які приймаються або підтримуються машинами.

Люди чудово навчаються на основі когнітивних даних. Слухаючи звуки, дивлячись на обличчя та спостерігаючи за навколишнім середовищем, ми розрізняємо склади, слова, речення та значення. Ми можемо встановити зв'язок між усмішкою, саркастичним тоном, буквальним значенням речення та тим, що мав намір сказати мовець. Ми можемо читати книги та новини, розмовляти з людьми і робити складні висновки. Комп'ютери все ще не так повно використовують когнітивні дані, як це роблять люди. І в той час як люди зазвичай накопичують лише весь чуттєвий діапазон своїх переживань, то певні дані (наприклад, зображення, звук, відео певного виду) зазвичай збирають з метою навчання комп'ютерів.

Важлива відмінність між машинним і людським навчанням полягає в тому, що ML базується на відомих алгоритмах. За визначенням, алгоритм – це набір інструкцій, які описують порядок дій виконавця, щоб досягти результату розв'язання задачі за скінченну кількість дій;

система правил виконання дискретного процесу, яка досягає поставленої мети за скінченний час. Люди, звичайно, також мають процедури для вирішення проблем у скінченну кількість кроків, які часто включають повторення операції. Однак, навіть людина, яка їх використовує, не завжди може знати або розуміти ці процедури.

Ми знаємо, які алгоритми використовують машини *(ми записуємо їх на мовах програмування)*, і ми можемо контролювати дані, з яких вони навчилися (ми можемо в будь-який момент скинути їх налаштування, ввести певні навчальні дані в модель або припинити навчання). Навчання та наступні рішення що приймаються машиною, в принципі, більш прозорі ніж ті що приймаються людьми. Зрештою, ми не писали код для навчання людини, і ми мало контролюємо вхідні дані, які люди використовували у своєму навчанні та прийнятті рішення. Отже, є певна іронія в тому, що одна з головних критичних зауважень щодо використання машинного навчання при прийнятті рішень полягає в тому, що машинні рішення є непрозорими.

Одне з можливих пояснень цієї невідповідності полягає в тому, що можна відносно просто запитати людей, як вони прийшли до своїх рішень. Було б розумно очікувати, що начальник поліції пояснить факти, інтерпретації та пріоритети, що стоять за його прийняттям рішень. Набагато важче дати подібні пояснення того, чому машина змоделювала саме такі результати своєї роботи; а у багатьох випадках може бути навіть важко описати це простою мовою. Непрозорість навчання машин може, в принципі, бути нижчою, ніж у людей, але на практиці вона вища. Як люди, ми краще підготовлені до того, щоб запитувати інших людей, як вони дійшли своїх висновків, ніж допитувати модель машини.

Ця непрозорість, хоч і зрозуміла, викликає занепокоєння, оскільки може призвести до «позбавлення від відповідальності» людей у змішаних системах *«чоловік-машина»* [8]. Вихідні дані для машини можуть здаватися *«де-суб'єктивованими»* і, таким чином, інтерпретуватися кінцевими користувачами, як більш об'єктивні, ніж вони є насправді, бо на справді, вони цілком залежать від даних навчальної вибірки, яку формує людина, а значить тут є фактор суб'єктивності. В цьому сенсі, може бути корисним структурувати обговорення між експертами з ML та іншими профільними експертами навколо трьох елементів, які відображають цей тип перевірки:

1. Вимоги к даним, які використовуються для навчання нейронної мережі за допомогою ML;
2. Мета навчання нейронної мережі;
3. Як результати впливають на подальші навчальні дані.

Ці елементи не експерти з ML можуть зрозуміти та оцінити.

Корисним припущенням для експертів, які не займаються ML, під час обговорення моделей ML є припущення, що алгоритм навчання, обраний експертом ML, є оптимальним для досягнення встановленої мети за допомогою заданих даних. Незважаючи на те, що це припущення багато разів хибне, воно має перевагу, оскільки робить більшу частину складності машинного навчання, наприклад знання того, як функціонують рекурентні нейронні мережі, неактуальними. Вважається, що це припущення може знизити планку для нефахівців щодо вступу в дискусію з експертами з ML і сприяти плідній дискусії.

Оптимальний у цьому контексті не є нормативним терміном і існує ключова відмінність між поняттями оптимального та доброго. Обчислення та статистика пропонують можливість економічно ефективного тестування величезної кількості можливих моделей. Метою алгоритму ML є визначення оптимальних параметрів для досягнення визначеної мети навчання, нехтуючи такими речами, як етичні проблеми, пов'язані з поліцейською діяльністю, якщо вони явно не реалізовані та запрограмовані [12]. Оптимізація означає вибір параметрів, які роблять

найточніші прогнози, враховуючи використані дані та навчання, щоб досягти найкращої продуктивності.

4. Питання про справедливість і обґрунтованість: інструментарій

Суспільство зацікавлено в запобіганні злочинності та ефективній поліцейській діяльності, але також зацікавлено в тому, щоб стратегії правоохоронних органів, включаючи рішення щодо розгортання та стеження, були ефективними, чесними та справедливими. Це вимагає розуміння, оцінки та управління [3].

Загалом кажучи, рішення можна критикувати з огляду на два різні питання: обґрунтованість рішення та справедливість рішення. Щоб розглянути валідність (*відповідність*) моделі, ми запитуємо: чи призвело рішення до запланованого результату? Щоб оцінити валідність, рецензенту потрібно буде розглянути: - чи модель навчання відображає фактичну ефективність на основі узгодженого показника ефективності, або сама метрика ефективності вимірює те, що ми мали намір виміряти. Оскільки цілі навчання можуть бути досить абстрактними та суперечливими (*наприклад, ціль зменшення рівня злочинності*), обсяг питань валідності, ймовірно, буде за межами предметної області для розуміння програмістів і статистиків. Однак, навіть досить «вузькі» питання, такі як упередженість відбору в навчальних даних, може бути легше викрита експертами, котрі не займаються ML і які можуть знати, наприклад, як збираються відповідні відомості. Так наприклад – інформація, швидше за все, буде зафіксована поліцейськими, якщо вони вважатимуть її корисною для успішного розкриття або запобігання злочину.

Перевірка справедливості рішення, прийнятого на основі «людської» або машинної моделі, передбачає запитання, чи були запланований результат і засоби його досягнення хорошими? Оцінка справедливості є нормативним завданням. У цьому контексті це означає, що мета навчання, процес, який покращує навчання, і засоби для досягнення успіху в навчанні визначені демократично легітимним шляхом. Забезпечення можливості відкритих і демократичних дебатів є як вимогою, так і частиною вирішення проблеми справедливості.

Далі наведено набір питань, які неексперти можуть поставити розробникам моделей з машинним навчанням, сподіваючись отримати зрозумілі відповіді. Відповіді у формі «проте ми врахували це в нашій моделі» вимагають рішень моделювання, які можна було б висловити явно, і ці рішення повинні бути застереженням для всіх, хто використовує модель. Інструментарій поділено на розділи з питаннями про дані, про навчання та про антиципаційну оцінку результатів. Мета інструментарію – надати можливість експертам, які не займаються ML, вести дебати з експертами з ML.

5. Дані для навчання MLмоделей

Злочини частіше за все фіксуються поліцією і лише зареєстровані злочини стають даними про злочини. Таким чином, статистика злочинності проходить процес відбору. Перша стадія процесу – законодавча; це коли певні діяння криміналізовані. В подальшому дані накопичуються в правоохоронних інформаційних системах. Дані категоризуються, частково структуруються та захищаються. Суспільство має доступ лише до частини відомостей про злочини. Більша частина даних є закритою від суспільства. Збір даних є суб'єктивним і залежіть від суб'єкта що їх збирає (*специфічного підрозділу*). Частина злочинів є латентною (прихованою) і не попадає в системи поліцейського обліку через те, що деякі злочини не повідомляються або не розкриваються громадськістю та поліцією.

В таких умовах, формування даних для навчання моделей ML зустрічається з проблемами:

- репрезентативність вибірки;
- актуальність;
- неупередженість даних.

Упередженість правоохоронних практик інколи можуть впливати на дані, створені поліцією. Дослідження *Human Rights Data Analysis Group* наводить показовий приклад [13]. Дослідження змоделювало прогнози правоохоронної діяльності з використанням алгоритму ML*«PredPol»* [14] на основі правоохоронних даних щодо боротьби з наркотиками в Окленді, Каліфорнія, а потім порівняло прогнози з моделями вживання наркотиків, оціненими на основі даних національного опитування про вживання наркотиків і здоров'я. Було виявлено, що за результатами роботи алгоритму *«PredPol»* «темношкірі люди африканського походження будуть об'єктом поліції по боротьбі з наркотиками приблизно вдвічі частіше», незважаючи на оцінки, які показують приблизно однакові рівні вживання наркотиків [13]. Люди з низьким рівнем доходу та не білошкірі, окрім темношкірих людей африканського походження, також будуть непропорційною мішенню, тобто надмірною цілю поліції.

Цей приклад упередженості показує, як вхідні дані, які використовуються для навчання машин і людей, можуть призвести до недійсних моделей і несправедливої практики. У цьому випадку недійсною моделлю або переконанням є те, що націлювання на житлові райони темношкірих людей є розумним способом поведінки поліції, незважаючи на те, що моделі вживання наркотиків свідчать про те, що в житлових районах темношкірих людей не повинно бути вищих випадків вживання наркотиків. Результатом є несправедлива правоохоронна практика, згідно з якою темношкірі громадяни та райони піддаються більшому нагляду, ніж білошкірі громадяни, незважаючи на відсутність об'єктивної основи в расових моделях злочинів, пов'язаних із наркотиками.

Таким чином, неексперти з ML, повинні задати наступні питання стосовно даних, що планується використовувати для навчання моделі ML:

Блок А:
- які вхідні дані використовуються?
- який набір використовувався для навчання моделі?
- який набір використовується для тестування продуктивності?
- коли, ким, як і де були зібрані дані?

Блок Б:
- чи є іменовані ознаки (змінні)?
- якщо так, то які вони і які найбільше впливають на результати?
- які операції з ними відбуваються та як вимірюються результати?

Блок В:
- - чи охоплюють вхідні дані функції *(прямо чи опосередковано)*, що не використовуються для прийняття рішення? Наприклад, чи пов'язані будь-які вхідні характеристики зі статтю таким чином, що модельні рішення відрізняються, якщо ви чоловік чи жінка?

Блок Г:
- чи дані репрезентативні, як впливають на результат роботи моделі? Наприклад, чи була модель перевірена в умовах, де вона застосована?
- які найбільш очевидні відмінності між умовами навчання та поточною роботою моделі?
- чи потрібно вносити якісь корективи для окремих груп даних чи результатів?

Блок Ґ:

- як збираються дані? Наприклад, чи їх збирали з наміром використовувати для таких рішень?
- чи знаємо ми про будь-які упередження відбору *(або через задум, або через практичні проблеми)* щодо збору даних?
- хто збирає дані?

6 Мета навчання. Постановка питань

Будь-яке навчання має мету. У моделях ML цілі можуть бути більш або менш явними. Незалежно від того, чи є навчання з вчителем, або ні, можна й доцільно запитати, якою є головна мета навчання та яке конкретне правило чи вимірювання використовується як еталон для визначення того, чи навчається модель.

Моделі ML оптимізуються відповідно до конкретних цілей навчання, які необхідно реалізувати та виміряти [15]. Оскільки деякі типи результатів легше виміряти, ніж інші, моделям ML властива упередженість щодо вибору навчальних цілей, які найлегше виміряти. Результати, які вже були виміряні, наприклад, місце арешту, стають привабливішими, ніж невиміряні результати, такі як реакція громадян на тактику поліції. Коли властива упередженість переноситься з машинних моделей на фактичне прийняття рішень, наслідки можуть бути різноманітними, як показує дослідження *HRDAG* [13].

Коли навчальна ціль є спірною або реалізуються далеко від ідеальної моделі, то прогнози таких моделей ML слід застосовувати з обережністю. Надзвичайний приклад можна знайти у By та Чжана [16], які стверджують, що їхня модель ML може автоматично ідентифікувати злочинців лише за характеристиками обличчя та «емпірично встановити достовірність автоматизованого висновку про злочинність за обличчям, незважаючи на історичні суперечки навколо цієї лінії дослідження». Тут модель не відокремлює злочинців від НЕ злочинців, а скоріше фотографії засуджених і підозрюваних із серії фотографій документів, отриманих з мережі Інтернет. Самі автори погоджуються з критиками, які стверджують, що різниця в соціально-економічному статусі в двох наборах може пояснити, чому моделі вдається розділити набори [16]. В цьому випадку мета щодо розпізнавання злочинців за обличчям є хибною з точки зору експертів в області криміналістики, і скоріш відображує розподіл неблагополучної частини населення та, відповідно, благополучної.

Існують два очевидних «занепокоєння» при розгляді використання моделі ML у процесах прийняття рішень:

- чи реалізована ціль закладена в моделі ML, та забезпечена ефективність у порівнянні з більш загальною та всеохоплюючою метою навчання?
- чи не створює операційна мета небажані побічні ефекти?

Крім того, варто окреслити ще одне занепокоєння, котре полягає в тому, що модель ML оптимізує багато, але не всі аспекти головної навчальної цілі [17]. При розробці моделі ML та оцінювання даних, які використовуються для навчання, деякі аспекти можуть бути втрачені. Обговорюючи головну мету процесуML та те, якими мають бути основні цілі навчання, можна визначити елементи, стосовно яких модель ML не оптимізується, і вжити відповідних заходів. Коли модель ML оптимізує лише деякі з встановлених цілей, необхідно бути обережними стосовно того, щоб модель ML вирішувала дії безпосередньо [18].

Таким чином, неексперти з ML, повинні задати наступні питання стосовно мети навчання моделі ML:

- яка основна мета навчання? Наприклад, чого б суспільство, хотіли досягти, приймаючи ці рішення?

- які конкретні правила та метрики використовуються як еталонні для визначення того, чи навчається модель? Наприклад, що таке залежна(і) змінна(и)?
- яке правило подібності використовується?
- які параметри навчання мають більшу вагу на роботу моделі ніж інші?
- як це правило реалізується та вимірюється?
- чи є згода щодо мети навчання?
- чи є конкретна ціль навчання повним описом того, чого ML має досягти?
- оптимізація дій або прийняття рішень щодо цієї навчальної цілі відніме зусиль або знову допоможе почати активно працювати?

7.Антиципаційна оцінка результатів, постановка питань

Наші моделі, машинні чи розумові, впливають на світ, коли ми використовуємо їх для прийняття рішень. У поліцейській діяльності головне, звичайно, зробити певний вплив на соціум. Прогнозний аналіз призначений для проактивної діяльності, «щоб визначити ймовірні цілі для втручання поліції». Рішення, дії, аналізи, політика, а також територіальний та історичний контексти сприяють формуванню сучасних концепцій у практиці правоохоронної діяльності. На відміну від, скажімо, фізики, поліцейські рішення впливають на соціальні системи. Ми використовуємо антиципаційний підхід, щоб позначити це розуміння.

Основна суть цього підходу полягає в тому, що експерти в поліцейській діяльності, які мають величезний досвід в питаннях організації процесів функціонування правоохоронної системи та результатів її роботи, спеціальний досвід *(слідчі, оперативні робітники, криміналісти та інші спеціалісти)*, мають так би мовити колективний розум, що сформований як результат сумаризації навчання та досвіду, кожного мозку спеціаліста на даних, якими він оперує на протязі професійної діяльності. Тобто при постановці задачі для машинного навчання, вони заздалегідь передбачають результат в рамках їх компетенції.

> *Вираз «колективний розум» використовується вже кілька десятиліть, але став важливим і популярним із приходом нових комунікаційних технологій. Він може викликати асоціації з груповою свідомістю або надприродними явищами, але технічно орієнтовані люди зазвичай розуміють під цим отримання нового знання з об'єднаних уподобань, поведінки та уявлень певної групи людей.*
>
> *Звичайно, колективний розум був можливим і до появи Інтернету. Для того, щоб збирати дані від розрізнених груп людей, об'єднувати їх та аналізувати, Всесвітня павутина зовсім не потрібна. До найважливіших форм подібних досліджень входять соціологічні опитування та переписи. Отримання відповідей від великої кількості людей дозволяє робити про групу такі статистичні висновки, які на основі поодиноких даних зробити неможливо. Породження нових знань виходячи з даних, отриманих від незалежних респондентів, – це і є суть колективного розуму*[19].

Різниця між природною нейронною мережею людини та штучною машини полягає в тому, що людина вчиться довгий проміжок часу на обмежених наборах даних, а машина короткий час на великих. Але людина має більш якісний набір даних, тому що постійно отримує зворотний зв'язок на протязі всього часу навчання з іншими спеціалістами. Мозок людини не здатний обробляти занадто великі обсяги даних, а машина здатна, але алгоритм обробки є алгоритмом роботи людського мозку. Спеціалісти з правоохоронної діяльності, можуть оцінити результати роботи моделі машинного навчання в правоохоронних задачах на основі свого досвіду, що не можуть зробити спеціалісти з машинного навчання.

Антиципаційний підхід допоможе вирішити три головних занепокоєння щодо застосування ML для прийняття рішень:

- по-перше, оцінити дані що застосовуються за їх повнотою, репрезентативністю, актуальністю та відповідності постановці задачі;
- по-друге, оцінити навчальні шаблони, та шаблони отримані в результаті роботи щодо наявності причинно-наслідкових зав'язків;
- по-третє, оцінити чи не суперечать отримані результати роботи алгоритмів машинного навчання історичній практиці.

Неексперти в ML але в експерти в поліцейській діяльності, повинні отримати відповіді на питання:

- чи може рішення машини, вплинути на пізніші дані навчання?
- чи модель машини представляє причинно-наслідковий зв'язок, чи це прагматичне рішення?
- чи модель спирається на кореляції, які, ймовірно, лише покращують ефективність через історичну практику?
- чи результати не суперечать історичній практиці?

8. Висновки

1. Оскільки поліцейські департаменти прагнуть одночасно запобігти, як заподіянню шкоди, так і ощадливо витрачати ресурси, то вони все частіше впроваджують проактивну політику і методи. Однак використання інструментів прогнозування вимагає ретельного розгляду і спільної експертизи, як експертами в ML, так і експертами з правоохоронної діяльності, що не є експертами з ML. Тільки сумісна робота, стосовно даних, цілей і конструктивної моделі використання ML, від початку подачі даних до етапу досягнення цілей, дасть можливість правильно оцінити результати роботи, використовувати результати в прийнятті рішень та створювати коректну ініціативу, щодо нормативного врегулювання використання моделей ML в поліцейській діяльності. Питання про мету використання технології ML у правоохоронній діяльності є, як моральними, так і політичними.

2. Головна мета полягає в тому, щоб розширити можливості нетехнічних експертів і зацікавлених сторін та заохотити їхню участь у роботі, щодо можливостей застосування ML у поліцейській діяльності, а також у процесах розробки профільної моделі ML. Така участь є не лише технічно й морально необхідною, але й можливою.

3.Сформовано перелік питань, які доцільно розглянути під час проведення сумісних обговорень і відповідних тематичних експертиз, щодо можливостей застосування технологій ML в поліцейській діяльності та пов'язаних із нею сферах боротьби зі злочинністю. В цьому сенсі питання, щодо коректності формування вихідних даних, мети навчання та того, як саме використовувані модельні рішення впливають на подальші дані, є 3-ма головними векторами можливих досліджень, які неексперти можуть переосмислити та завчасно скоригувати.

References

[1] Halterlein, J. and Ostermeier, L. (2018). 'Special Issue: Predictive Security Technologies'. European Journal for Security Research 3(2): 91–94.DOI: 10.1007/s41125-018-0034-z

[2] Hughes, D. (2017). 'Robot Investigators 'Could Be Used to Examine Documents in Criminal Cases'. The Independent (14 December 2017). https://www.independent.co.uk (accessed 20 November 2018).

[3] Bennett Moses, L. and Chan, J. (2016). 'Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability'. Policing and Society 28 (7): 1–17.DOI: 10.1080/10439463.2016.1253695

[4] Ratcliffe, J. H. (2016). Intelligence-Led Policing. Abingdon, Oxon; New York: Routledge. DOI: 10.4324/9781315717579

[5] Innes, M. and Sheptycki, J. W. (2004). 'From Detection to Disruption: Intelligence and the Changing Logic of Police Crime Control in the United Kingdom'. International Criminal Justice Review 14(1): 1–24.DOI: 10.1177/105756770401400101

[6] Goldstein, J. (1960). 'Police Discretion Not to Invoke the Criminal Process: Low-Visibility Decisions in the Administration of Justice'. The Yale Law Journal 69(4): 543–594.

[7] Burrell, J. (2016). 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms'. Big Data & Society 3(1): 1–12.DOI: 10.1177/2053951715622512
[8] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016). 'The Ethics of Algorithms: Mapping the Debate'. Big Data & Society 3(2): 1–21.DOI: 10.1177/2053951716679679
[9] Strukov V.M., Uzlov D.Yu. et al. Information Technologies in Law Enforcement. Part 1. High-Tech Trends in Law Enforcement of Foreign Countries. Kharkiv: LLC 'DISA PLUS', 2020. [In Ukrainian]
[10] Hughes, D. (2017). 'Robot Investigators 'Could Be Used to Examine Documents in Criminal Cases'. The Independent (14 December 2017). https://www.independent.co.uk (accessed 20 November 2018).
[11] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016). 'The Ethics of Algorithms: Mapping the Debate'. Big Data & Society 3(2): 1–21.DOI: 10.1177/2053951716679679
[12] Norwegian Board of Technology (2018). Artificial Intelligence: Opportunities, Challenges and a Plan for Norway. Oslo: Norwegian Board of Technology. DOI: 10.5617/nmi.9950
[13] Lum, K. and Isaac, W. (2016). 'To Predict and Serve?'. Significance 13(5): 14–19.DOI: 10.1111/j.1740-9713.2016.00960.x
[14] Mohler, G. O., Short, M. B., Malinowski, S. et al. (2015). 'Randomized Controlled Field Trials of Predictive Policing'. Journal of the American Statistical Association 110(512): 1399–1411.DOI: 10.1080/01621459.2015.1077710
[15] Zhang, Lemoine, B. and Mitchell, M. (2018). Mitigating Unwanted Biases with Adversarial Learning. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, ACM. pp. 335–340.DOI: 10.1145/3278721.3278779
[16] Wu, X. and Zhang, X. (2016). 'Responses to Critiques on Machine Learning of Criminality Perceptions' (Addendum of arXiv: 1611.04135). ArXiv: 1611.04135 [Cs]. http://arxiv.org/abs/1611.04135 (accessed 8 January 2019).
[17] Beck, C. and McCue, C. (2009). 'Predictive Policing: What Can we Learn from Wal-Mart and Amazon about Fighting Crime in a Recession?'. Police Chief 76 (11): 18–24
[18] Sklansky, D. A. (2008). Democracy and the Police. Stanford, CA: Stanford University Press. DOI: 10.1515/9780804763226
[19] Toby Segaran (2008) Programming Collective Intelligence. Published by O'Reilly Media Inc., ISBN-10:0-596-52932-5.

Submitted October 17, 2023; Revised November 19, 2023; Accepted December 20, 2023

Authors:
Dmytro Uzlov, Acting Dean of the Faculty of Computer Science, Candidate of Technical Sciences, Associate Professor, V. N. Karazin Kharkiv National University, Ukraine.
E-mail:dmytro.uzlov@karazin.ua
ORCID: https://orcid.org/0000-0003-3308-424X

Volodymyr Strukov, professor of the Department of Cybersecurity and DATA technologies, Candidate of Technical Sciences, Associate Professor, Kharkiv National University of Internal Affairs, Ukraine.
E-mail:struk_vm@ukr.net
ORCID: https://orcid.org/0000-0003-4722-3159

Vladyslav Hudilin, master's degree (cybersecurity), Kharkiv National University of Internal Affairs, Ukraine
E-mail:vgudilin7@gmail.com
ORCID: https://orcid.org/0000-0002-3844-1448

Oleksii Vlasov, Ph.Dcadidate, Kharkiv National University of Radio Electronics, Ukraine.
E-mail:moonreactor@gmail.com
ORCID: https://orcid.org/0000-0003-1619-0032

Problematic issues of machine learning technology in law enforcement.

Abstract. Law enforcement agencies increasingly use predictive and automation technologies where the core technology is often a machine learning (ML) model. The article considers the problem of accountability and responsibility of law enforcement agencies and officials connected with using of ML models. The authors point out that accountability is a key element of democratic law enforcement, but using of the predictive software can create challenges in ensuring that accountability. The article discusses how the application of ML can lead to obfuscation of responsibility and complicating accountability in «multi-agent structures» that combine humans and computational tools. Special attention is paid to the opacity of predictive algorithms and automated decision-making systems. It becomes a source of misunderstandings and caution regarding their use. The authors raise questions about how effective oversight and full reporting can be ensured when key components of the decision-making systems remain unknown to the general public, officials, and even developers of the models. The paper argues that important questions related to ML decision models can be solved without detailed knowledge of the machine learning algorithms, allowing non-ML law enforcement experts to study them in a form of intelligent control. Non-ML experts can and should review trained ML models. The authors provide a «toolkit» in the form of questions about three elements of the ML-based decision models that can be qualitatively explored by non-machine learning experts: training data, training goal, and anticipatory outcome evaluation. This approach expands the capabilities of these experts in the form of an objective assessment of the use of ML models in law enforcement tasks. This will allow them to evaluate effectiveness of the models through the prism of their own experience. The basic idea is that even without deep technical knowledge, law enforcement experts can analyze and review ML models. This approach promotes understanding of the use of machine learning technologies in law enforcement, expanding the potential of non-ML law enforcement experts.

Keywords*: Machine Learning, Artificial Intelligence, Data Analysis.*

DOI: 10.26565/2519-2310-2023-2-02
UDC004.056.55

USING ZK-SNARK TO SOLVE BLOCKCHAIN SCALABILITY PROBLEM

Kuznetsova Kateryna [1,2], Yezhov Anton [2]

[1]V.N. Karazin Kharkiv National University, Kharkiv, 61022, Ukraine
e-mail: kate7smith12@gmail.com, ORCID: https://orcid.org/0000-0002-5605-9293
[2]Zpoken, OU, Harju maakond, Tallinn, Kesklinnalinnaosa, Sakala tn 7-2, 10141, Estonia, https://zpoken.io/
anton.yezhov@zpoken.io

Submitted October 18, 2023; Revised November 24, 2023; Accepted December 25, 2023

Abstract: The paper elucidates the fundamental concepts of blockchain technology and its essential parameters, delving into the contemporary scalability challenges faced by blockchain networks. It studies existing directions and compares well-known protocols to propose the solution for the blockchain scalability problem. The main goal of this research is to propose a promising method to solve the scalability problem in blockchain technology. This proposed solution should be universal and applicable in different systems. We chose zero-knowledge proof technology as a promising direction for detailed study. We used protocols, based on this technology, to develop a validation system for a linked chain of blocks. Presented experimental results substantiate the prospects of this direction for solving the scalability problems of modern blockchain systems. The relevance of the chosen topic is determined by the mass introduction of blockchain systems in various areas of human life. As it happens to every network, the volume of information that must be continuously processed increases. This challenge demands to develop solutions to improve systems, making them flexible in working with millions of users. At the same time, it is still important to maintain the security and confidentiality of the information and keep the decentralized organization of the data exchange process in the updated systems. Therefore, in the modern blockchain industry, the predominant challenge revolves around discovering models and methods to overcome the scalability hurdle, facilitating the widespread implementation of blockchain applications on a full scale.

Keywords: *blockchain, blockchain trilemma, blockchain scalability problem, Zero-knowledge proofs, ZK-SNARK, PLONK, FRI.*

1.Introduction

Blockchain is the distributed ledger technology that promises to transform industries with its immutable, transparent and decentralized mechanism for recording transactions. However, the inherent problem of scalability in blockchain creates a significant barrier preventing its widespread adoption. The main goal of this research is to propose a promising method to solve the scalability problem. During the research we made an overview of blockchain technology concepts, focusing particularly on the issue of scalability. We also studied well-known directions, analyzing their advantages and highlighting the challenges and risks they face to focus on most relevant areas.

A chosen direction for in-depth study is zk-SNARKs. It plays a key role in enhancing privacy and security in decentralized networks, it increases the integrity of systems while protecting user identities and transaction details. We developed numerous schemes for constructing proofs of computational integrity, including recursive proof generation and verification processes, using the Rust programming language and *PLONK & FRI* protocols within the *Plonky2* framework. The work also presents the results of computational complexity and efficiency of the proposed schemes. Experimental analysis covers scenarios involving stand-alone and aggregated proofs for single and multiple data blocks. The results highlight the trade-off between the complexity of proof generation and the speed of verification, emphasizing the potential advantages of recursive proofs.

This comprehensive study aims to contribute valuable information to the current blockchain scalability discourse, paving the way for more scalable and efficient blockchain systems.

2.Overview on blockchain technology & scalability problem

Blockchain represents a tamper-resistant digital ledger without a central repository and usually without a central decision-making center, which is implemented by linking information into a

continuous chain of blocks. Connection between blocks is provided by a cryptographic mechanism through the calculation of the hash of the previous block.

Such a system has to provide a sufficient level of security and anonymity, i.e. preserve the right to privacy [1].

In addition to security, it must adhere to decentralization, i.e. not be governed by a single decision-making center that solves reliability issues, because a centralized structure with a potentially single point of failure always attracts attackers. Decentralization makes a ground for censorship, establishes the principles of democratic decision-making, provides freedom of speech and independence of thought.

Moreover, the demand for the system's services is growing, and it must continuously develop. The number of users increases and their demands become more complex, but at the same time the service time should not increase significantly. This is a difficult requirement, because it is not always possible to provide scalability extensively, i.e. by only increasing the number of computers.

Thus, we have three main requirements: security, decentralization and scalability, which are formulated in the well-known blockchain trilemma proposed by Vitalik Buterin, the co-founder of *Ethereum*, shown in fig. 1 [1].

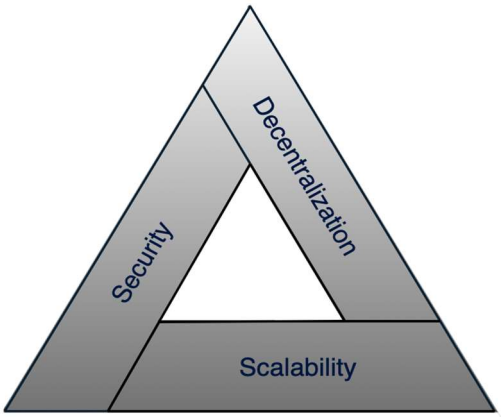


Fig. 1 – Blockchain trilemma

Decisions made about the blockchain trilemma have significant implications for blockchain network design and performance. *Bitcoin*, for example, prioritizes decentralization and security over scalability, resulting in shorter transaction confirmation times, which excludes the majority of users. Modern practice has become the use of *Bitcoin* as a savings account rather than an instant payment system. *Ethereum* has explored various strategies, including moving to *Ethereum 2.0*, to improve scalability while maintaining some level of decentralization and security.

In particular, scalability affects two other characteristics, namely that increasing the size of the network potentially centralizes control, as large amounts of data will attract new users to join. Additionally, increasing block sizes have security implications, as large blocks can slow down the propagation of data across the network, ultimately potentially making it easier for miners to manipulate the blockchain.

Therefore, since scalability can potentially be the root cause of instability of security and decentralization, this paper examines exactly this characteristic in its dynamics, evaluates potential threats to blockchain systems, considers existing approaches to optimize the amount of calculations in the blockchain network, and proposes an accelerated method of verifying network blocks.

3. Studying existing approaches on solving scalability problem

Solving the blockchain scalability issue is critical to the widespread adoption of blockchain technology. Various projects offer solutions that facilitate the use of blockchain networks [2].

Layer-2. Layer-2 solutions are techniques that work on top of the main blockchain, allowing off-chain transactions. They aim to significantly increase transaction throughput and lower transaction costs while maintaining the security and decentralization of the underlying blockchain.

On the other hand, users are responsible for the security of their decisions. Mismanagement can lead to loss of funds. This additional responsibility increases the complexity of implementation.

Centralization also may be a concern in the early stages of Layer-2 implementation, as some nodes or channels may influence more than others.

Off-chain. Off-chain solutions involve conducting transactions and interactions completely outside the main blockchain. These transactions take place off-chain, meaning they are not recorded in the blockchain ledger. Examples of off-grid solutions include payment channels *(Lightning Network for Bitcoin)* and state channels (*Raiden Network for Ethereum*). They enable fast and low-cost transactions between users and can be used for a variety of use cases, including micropayments. However, they create new security and data availability challenges that must be carefully managed for successful deployment and implementation of these solutions.

Sharding. Sharding is an approach to solving scalability issues in blockchain that involves dividing the network into smaller parts called "shards" to process transactions and smart contracts more efficiently. However, this approach creates issues related to security, configuration of communication between segments, and data availability.

Changing consensus algorithm. Some blockchains are moving from energy-intensive consensus algorithms such as Proof of Work (*PoW*) to more efficient and friendly algorithms like Proof of Stake (*PoS*) or Delegated Proof of Stake (*DPoS*). *Ethereum 2.0*, also known as Eth2 or Serenity, is a major upgrade to the Ethereum blockchain that aims to move from *PoW* consensus mechanism to a *PoS*. But changing the consensus algorithm can lead to network forks, if there is no consensus among participants, and may potentially cause confusion and fragmentation of the network.

Zero-Knowledge Proofs. Zero-Knowledge Proofs (*ZKP*) [3-5] play a crucial role in solving scalability issues in blockchain technology. *ZKP* uses advanced cryptographic methods to authenticate transactions without revealing data itself, ensuring secure and tamper-proof transactions. The implementation of *ZKP* makes it easier to verify off-chain transactions, reducing the computational burden on the main blockchain and improving scalability.

ZKP algorithms are of two types: interactive and non-interactive. The first ones work in such a way that the prover and the verifier participate in a reverse interaction where they exchange a series of messages. Non-interactive proofs do not require multiple rounds of interaction. A verification device can generate a single message that can be verified by a verifier.

ZK-SNARK. ZK-SNARK (*Zero-Knowledge Succinct Non-Interactive Argument of Knowledge*) [6]. This is a non-interactive *ZKP* that allows efficient verification of calculations without revealing the details of the calculations.

To sum up, *ZK-SNARK* advantages such as privacy, scalability and security make it a promising direction for improving blockchain technology. They offer robust compact proofs and scalability improvements.

Ongoing research and development in this field improves the existing implementation of zero-knowledge proof technology for wider adoption.

4. Overview on ZK-SNARK

Let's delve into the three technical concepts that underlie all cryptographic proofs: arithmetization, low-degreeness, and cryptographic assumptions.

Arithmetization. In the field of cryptographic proofs, arithmetization involves the transformation of mathematical problems and operations into arithmetic operations performed within finite fields. Essentially, it involves expressing any given statement as an algebraic equation, usually in polynomial form [7].

The choice of arithmetic approach depends on the specific requirements of the cryptographic scheme, including considerations of security, efficiency, and the nature of the problem being solved.

Low-degreeness. Applying low degree polynomials is the process of ensuring that polynomials (*algebraic equations created during arithmetization*) have a degree lower than a specified threshold value. The degree of a polynomial corresponds to the highest degree of the term in this polynomial.

Polynomials of low degree also provide computational efficiency, particularly accelerated verification. This is especially important in blockchain where speed and resource efficiency are critical.

Polynomial commitment scheme (PCS).PCS is a cryptographic protocol designed to efficiently compute polynomials. In this scheme, the prover, one of the involved parties, has the ability to commit a polynomial without revealing its full details. Subsequently, the verifier, the other party, has the opportunity to confirm the properties of the fixed polynomial without gaining access to its full information [8].

Different proof systems use different *PCS* to generate and verify proofs, the most famous are *FRI* and *KZG*.

FRI (*Fast Reed-Solomon Interactive Oracle Proof*) is a cryptographic protocol designed to efficiently fix and verify large polynomials.

In the commitment step, the prover generates the high-degree polynomial commitment using a recursive process in which the original polynomial is broken down into lower-degree components. Then the prover calculates the commitment to each lower-level component, and the process is repeated recursively to the base element.

FRI achieves succinctness by using recursive composition of low-degree polynomial expansions, resulting in a commitment much smaller in size than the commitment of the original high-degree polynomial. This size reduction is critical to the performance of *ZK-SNARK*, where succinctness is a key requirement [9].

KZG polynomial commitment scheme (*named after its original inventors Keith, Zaverukh, and Goldfeder*) is a cryptographic protocol that allows efficient polynomial commitment [10].

KZG allows the prover to fix a polynomial using homomorphic properties, which allows efficient computation of fixed polynomials without revealing them.

Cryptographic assumptions. Cryptographic assumptions are mathematical assumptions or hypotheses that form the basis of the security of cryptographic primitives. These assumptions include the complexity of certain mathematical problems.

Zero-knowledge proofs rely on cryptographic assumptions to ensure the security and reliability of the proof system. *ZK-SNARK* assumes the complexity of certain problem: knowledge of an exponent (*Groth16*), algebraic group model (*PLONK, MARLIN*), elliptic curve cryptography (*Bulletproofs, Halo*), resistance to hash collisions (*STARK, Aurora, etc.*). If these problems are computationally difficult to solve, *ZKP* remains secure.

PLONK. PLONK (*Permutations over Lagrange-bases for Ecumenical Noninteractive arguments of Knowledge*) is a zero-knowledge proof system that made a significant contributions to the ZK-SNARK field. *PLONK* uses *SRS* (*Structured Reference String*) and permutation techniques to increase the computational efficiency of the prover, which simplifies its operation. This approach provides increased flexibility and eliminates the need for trusted configuration.

PLONK is a permutation-based constraint system that offers advantages in certain use cases. On the other hand, *PLONK* may have a larger proof size compared to *MARLIN*or*Groth16*, but this is often compensated by the increased efficiency and performance of the protocol in certain scenarios [11].

Additionally, there is an improved version of *PLONK* called *Turbo PLONK* that is positioned as a universal *SNARK*, which implies versatility and applicability in different scenarios.

In our opinion, this protocol deserves special attention and study, as it is promising due to its increased efficiency, reliability and applicability in various use cases.

5. Development of the block chain verification scheme using ZK-SNARK

This section presents a developed scheme for recursively proving the computational integrity of a chain of linked blocks.

The scheme uses a linked list built through a cryptographic connection. For each block, we generate a proof of the computational integrity of hash and digital signature to prevent data substitution. A chain of proofs is created by aggregating the previous block with the current one. As a result, we can verify that the block hash and signature have been calculated correctly, and the chain of proofs for previous blocks have been verified, i.e. are valid.

For the test case, consider a simplified version where each block contains the following data:

- Unique block number: $Nonce_i$;
- Hash of previous block: h_{i-1};
- Digital signature: $EDS(h_i)$.

For simplification we take $Nonce_i = i$. The result of the n-th hashing is as follows:

$$h_n = H(h_{n-1} \| n) = H(H(h_{n-2} \| n-1) \| n) = \ldots = H(H(H(\ldots H(H(0) \| 1) \ldots \| n-2) \| n-1) \| n) \quad (1)$$

Additionally, each hash is encrypted with a secret key sk, i.e. we form a signature $EDS(h_i, sk)$.

The public key pk is used to verify the signature, i.e. we decrypt EDS_i and check the equality $h_i = D(EDS_i, pk)$.

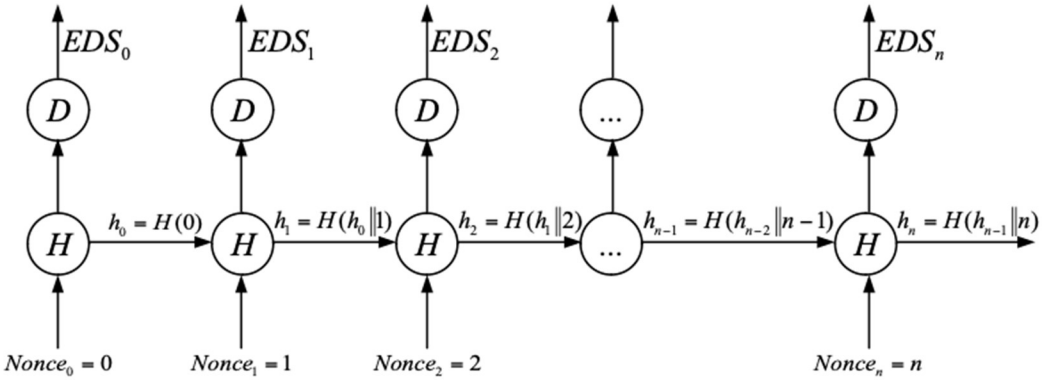


Fig. 2 – A simplified scheme of the linked list with digital signature

To implement a recursive proof of CI of a chain, it is necessary to consistently implement the following tasks:

1) Implement a hash chain $h_0 = H(0)$, $h_i = H(h_{i-1}||i)$, $i = 1,\ldots n$;
2) For each hash $h_0,\ldots h_n$:
   a) Create a circuit $CH_i(xH_i, wH_i)$ of the hash algorithm H, where the public input $xH_i = h_i$ is the result of hashing, the witness $wH_i$is the hash preimage: $wH_0 = 0$, $wH_i = h_{i-1}||i$,$i = 1,\ldots n$;
   b) Form public settings $(SH_{pi}, SH_{vi}) = S(CH_i(xH_i, wH_i))$, where SH_{pi} are public prover settings, SH_{vi} are public verifier settings;
   c) Form a proof CI for hashing $\pi H_i = P(SH_{pi}, xH_i, wH_i)$;
   d) Implement verification algorithm $V(SH_{vi}, xH_i, \pi H_i)$ takes values *{0, 1} (accept or reject)*;
   e) Proof verification, i.e. to make sure that $V(SH_{vi}, xH_i, \pi H_i) = accept$.
3) For each signature $EDS_i = E(h_i, sk)$, $i = 0,\ldots n$:

a) Create a circuit $CD_i(xD_i, wD_i)$ of proof verification $h_i = D(EDS_i, pk)$, where the public input $xD_i = h_i$ is the result of hashing, the witness $wD_i = (EDS_i, pk)$ are the signature and the public key;
b) Form public settings $(SD_{pi}, SD_{vi}) = S(CD_i(xD_i, wD_i))$, where SD_{pi} are public prover settings, SD_{vi} are public verifier settings;
c) Form a proof CI for signature verification $\pi D_i = P(SD_{pi}, xD_i, wD_i)$;
d) Implement proof verification algorithm $V(SD_{vi}, xD_i, \pi D_i)$ takes values *{0, 1} (accept or reject)*;
e) Proof verification, i.e. to make sure that $V(SD_{vi}, xD_i, \pi D_i) = accept$.

4) For every triple of proofs $\prod_{i-1} = P(S_{Pi-1}, X_{i-1}, W_{i-1})$, $\pi H_i = P(SH_{pi}, xH_i, wH_i)$ and $\pi D_i = P(SD_{pi}, xD_i, wD_i)$, $i = 1,...n$:
a) Create a circuit $C_i(X_i, W_i)$ verification algorithm V, where:
$X_i = (V(S_{Vi-1}, X_{i-1}, \prod_{i-1})), V(SH_{vi}, xH_i, \pi H_i), V(SD_{vi}, xD_i, \pi D_i)$, for all $i = 1,...n$.
$W_1 = (\pi_0, h_0, \pi_1, h_1, EDS_1, pk)$, $W_1 = (\prod_{i-1}, X_{i-1}, \pi_i, h_i, EDS_i, pk)$, for all $i = 2,...n$.
b) Form public settings $(S_{Pi}, S_{Vi}) = S(C(X_i, W_i))$, where S_{Pi} are public prover settings, S_{Vi} are public verifier settings;
c) Form a proof of CI $\prod_i = P(S_{Pi}, X_i, W_i)$;
d) Implement proof verification algorithm $V(S_{Vi}, X_i, \prod_i)$ takes values *{0, 1} (accept or reject);*
e) Proof verification, i.e. to make sure that $V(S_{Vi}, X_i, \prod_i) = accept$.

Thus, each proof $\prod_i = P(S_{Pi}, X_i, W_i)$, $i = 1,...n$ is the aggregation of three other proofs:

1) Proof CI of previous chain of linked hashes $\prod_{i-1} = P(S_{Pi-1}, X_{i-1}, W_{i-1})$;
2) Proof CI of current hash $\pi H_i = P(SH_{pi}, xH_i, wH_i)$;
3) Proof CI of current signature verification $\pi D_i = P(SD_{pi}, xD_i, wD_i)$.

Proof $\prod_0 = P(S_{P0}, X_0, W_0)$ is the aggregation of two proofs:

1) Proof CI of current hash $\pi H_0 = P(SH_{p0}, xH_0, wH_0)$;
2) Proof CI of current signature verification $\pi D_0 = P(SD_{p0}, xD_0, wD_0)$.

The scheme of forming a chain of recursive proofs of computational integrity with verification of the correctness of electronic digital signatures is shown in the figure below (fig.3).

Condition fulfillment $V(S_{Vi}, X_i, \prod_i) = accept$ for all $i = 1,...n$ means that the proof verification $\prod_{i-1} = P(S_{Pi-1}, X_{i-1}, W_{i-1})$, $\pi H_i = P(SH_{pi}, xH_i, wH_i)$ and $\pi D_i = P(SD_{pi}, xD_i, wD_i)$ were calculated correctly. If $V(S_{Vi-1}, X_{i-1}, \prod_{i-1}) = accept$, $V(SH_{vi}, xH_i, \pi H_i) = accept$ and $V(SD_{vi}, xD_i, \pi D_i) = accept$, it means that:

1. There is a proof of CI of previous chain, i.e. the verification $V(S_{Vi-2}, X_{i-2}, \prod_{i-2}) = accept$ is computed correctly;
2. There is a proof of CI of current hash, i.e. the value $h_i = H(h_{i-1}||i)$ is computed correctly;
3. There is a proof of CI of current signature, i.e. verification $h_i = D(EDS_i, pk)$ is computed correctly.

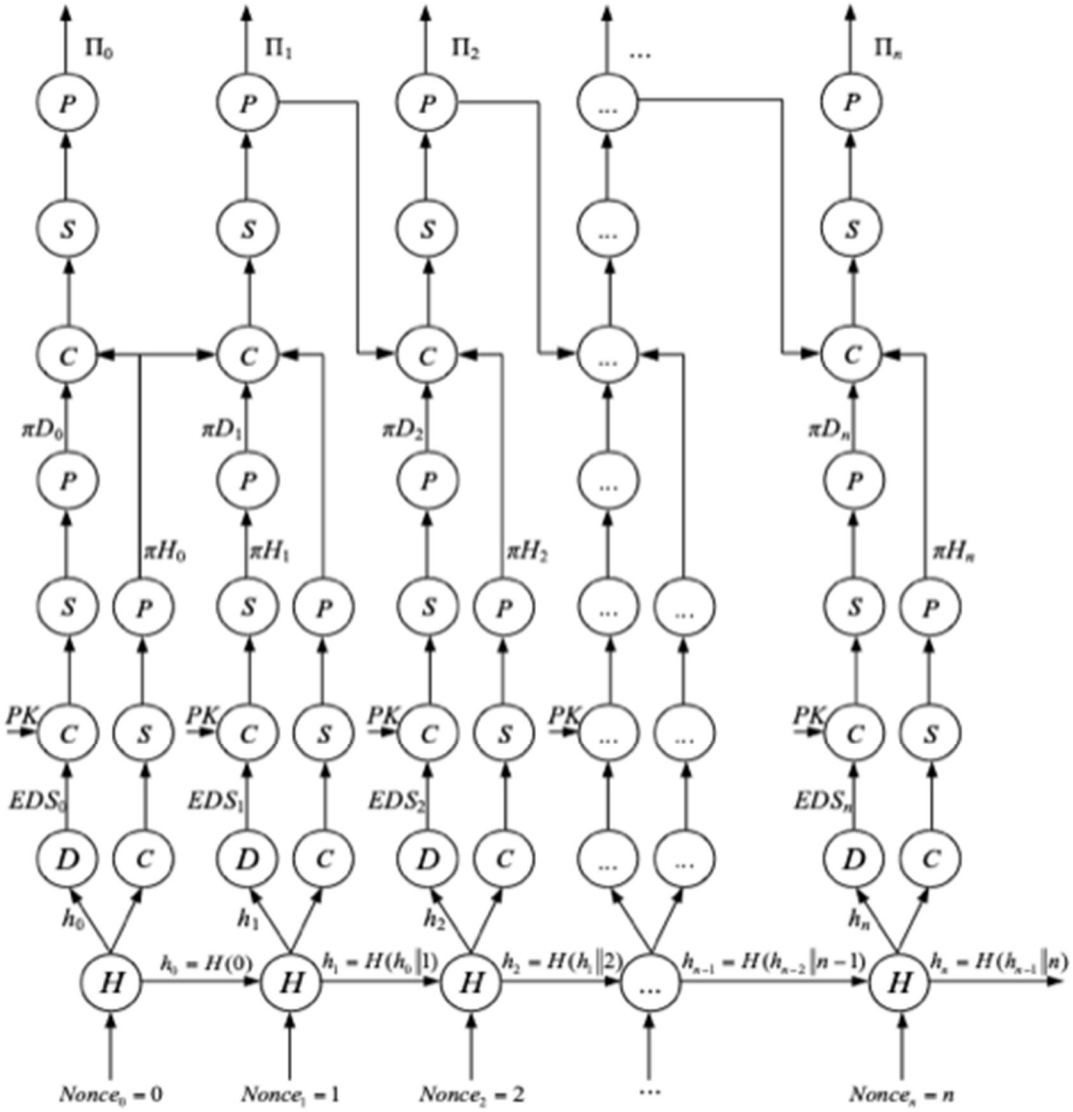


Fig. 3 – Scheme of forming a chain of proofs of CI with digital signature verification

6. Implementation & testing results of the block chain verification system

The implementation is based on the *Plonky2* framework [12], a widely recognized tool in the blockchain space developed by *Polygon Zero* [13]. It is known for its versatility and efficiency, making it an ideal choice for projects that require a high level of security and computational integrity.

Plonky2 implements the PLONK protocol combined with *Turbo PLONK* arithmetization, that optimizes the efficiency and performance of the verification process. As a commitment scheme, *Plonky2* uses *FRI*, which increases the security and verifiability of the proof system by providing a reliable mechanism for verifying the computational integrity on the blockchain.

We chose the SHA-256 and the ED25519 protocols for the blockchain test network. SHA-256 is a cryptographic hash function, which is widely used in blockchain technology and other security applications. SHA-256 is designed to take input data (or messages) of any length and produce fixed-size output data, 256 bits long. It is a one-way function, meaning it is computationally infeasible to reverse the process and retrieve the output from its hash.

ED25519 is an elliptic curve digital signature algorithm based on *EdDSA* (*Edwards Curve Digital Signature Algorithm*). ED25519 is designed for high security and performance using the Edwards Curve25519. The public keys generated are 256 bits long, which provides a good balance between

security and computational efficiency. The signatures generated are 512 bits long, providing a secure means of authentication. ED25519 uses the SHA-512 cryptographic hash function to process messages and create digital signatures. The hash function contributes to the security of the algorithm by producing a fixed-size output.

So, the two key components of the linked block chain proof scheme are the SHA-256-based hash validation scheme and the ED25519 signature scheme. In addition, we also need the SHA-512 scheme.

This implementation was tested on a chain of five blocks to analyze time costs and the size of final proof (*results in the table below*). Testing was performed on a 1,900 *GHz* AMD *Ryzen 7 5800U* computer (16).

Table 1 – Time and measurement results for a chain of proofs

№	Time to build a circuit, s	Time to make a proof, s	Proof size, bytes	Verification, s
0	*34,0128646*	*74,3645*	*146348*	*0,0549*
1	*33,7251775*	*98,1495*	*146348*	*0,061*
2	*31,6582847*	*107,0492*	*146348*	*0,1398*
3	*32,4201871*	*103,3622*	*146348*	*0,0922*
4	*34,285282*	*72,3417*	*146348*	*0,1147*

The graph below shows the results of calculating the time for native verification (*or recalculating all hashes*), verifying the proofs generated for each block, and recursive proof.

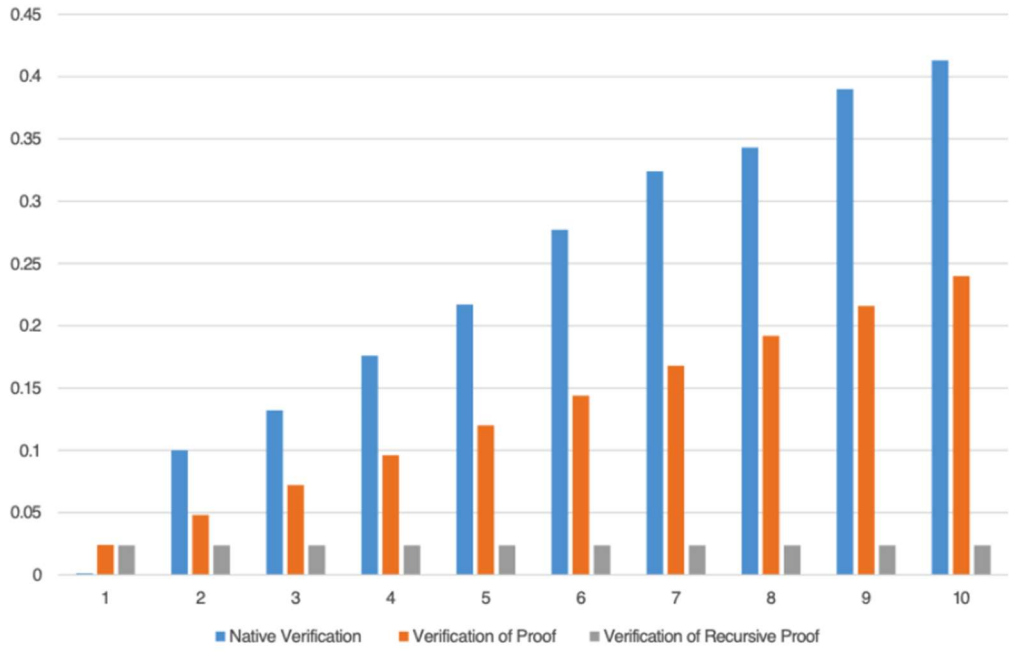


Fig. 4 – Computational complexity of native verification, proof verification, recursive proof verification

We see that even a proof for block significantly reduces the cost of verification. Recursion combines all these proofs into one. Verification is very fast, which, in fact, solves the main scalability problem.

To be more precise, we note that system errors may have occurred in the computing processes. System errors during calculation processes can arise from various sources, which leads to inaccuracies in the results. In addition, there are rounding errors because numbers with an infinite number of decimal places approach a finite representation [14].

Regardless, the experiment demonstrates that zero-knowledge proofs are a strong solution to scalability and privacy issues. This is especially important in large distributed computing projects. An introduced proof system can replace native verification. This significantly speeds up verification and makes the system easier to operate. The given time and measurement estimates show the prospects of this direction. In our opinion, it is necessary to continue researching this technology and implement blockchain verification systems based on it.

3.Conclusions

1. Modern blockchain systems face the challenge of scalability, which refers to increasing the capacity of the blockchain network to handle growing numbers of transactions. The scalability issue arises from the inherent trade-offs between decentralization, security, and scalability, known as the «*blockchain trilemma*».

2. *ZKP* plays a key role in enhancing privacy and security in decentralized networks. It increases the integrity of systems while protecting user identities and transaction details. In addition, *ZKP* simplifies the adaptation process on decentralized platforms. Users can quickly and efficiently establish their digital identity without the cumbersome task of providing large amounts of personal data. *ZKP* adheres to the principle of data minimization. By disclosing only what is essential for verification, it significantly reduces the amount of data in the network. This reduction is a key factor in enhancing security, as less information is exposed to potential attacks.

3. During the study, authors of the work analyzed all existing approaches that solve scalability problem. In our opinion, it is necessary to draw attention to the *PLONK* protocol and the *FRI* commitment scheme.

4. *PLONK* and *FRI* are used in Layer-2 solutions. The load on the main blockchain is reduced, by offloading transaction processing to the second layer, which solves scalability issues. *PLONK* and *FRI* have been implemented in various blockchain projects, demonstrating their versatility and effectiveness in increasing scalability.

5. During the research, authors of the paper developed a scheme for proving the validity of the block chain. Experiments showed a significant reduction in block verification costs. Recursion consolidates these proofs, allowing for quick verification of the entire chain. This effectively solves the main problem of scalability in the conditions of widespread implementation of distributed systems. Moreover, the conducted experiment highlights that zero-knowledge proofs offer an excellent solution to privacy problems, especially in large-scale distributed computing projects.

6. The time and measurement estimates provided highlight promising prospects toward zero-knowledge proofs. According to the team of authors, the continuation of research into this technology is a promising scientific direction.

References

[1] The History & Future of Blockchain Technology. https://www.linkedin.com/pulse/history-future-blockchain-technology-the-coin-times (31.05.2023)

[2] Blockchain Scalability: Exploring Solutionsin Blockchain Space. https://www.linkedin.com/pulse/blockchain-scalability-exploring-solutions (22.08.2023)

[3] Zero-Knowledge proofs. URL: https://en.wikipedia.org/wiki/Zero-knowledge_proof (6.10.2023)

[4] Zero-knowledge proofs – a powerful addition to blockchain. https://blockheadtechnologies.com/zero-knowledge-proofs-a-powerful-addition-to-blockchain/ (6.10.2023)

[5] Comparing General Purpose ZK-SNARKs. https://medium.com/coinmonks/comparing-general-purpose-zk-snarks-51ce124c60bd (2.11.2023)

[6] Eli Ben-Sasson, Alessandro Chiesa. Succinct Non-Interactive Zero Knowledge for a von Neumann Architecture. URL: https://eprint.iacr.org/2013/879.pdf (20.10.2023)ISBN 978-1-931971-15-7
[7] Arithmetization. URL: https://medium.com/starkware/arithmetization-i-15c046390862 (15.10.2023)
[8] Cambrian Explosion of Cryptographic Proofs. https://medium.com/starkware/cambrian-explosion-of-cryptographic-proofs-5740a41cdbd2 (7.10.2023)
[9] V. Buterin. STARKs, Part II: Thank Goodness It's FRI-day. URL: https://vitalik.ca/general/2017/11/22/starks_part_2.html (3.10.2023)
[10] Aniket Kate, Gregory M. Zaverucha, Ian Goldberg. Constant-Size Commitments to Polynomials and Their Applications? URL: https://www.iacr.org/archive/asiacrypt2010/6477178/6477178.pdf (16.10.2023) DOI 10.1007/978-3-642-17373-8_11
[11] V. Buterin. Understanding PLONK. URL: https://vitalik.ca/general/2019/09/22/plonk.html (5.10.2023)
[12] Plonky2. URL: https://github.com/0xPolygonZero/plonky2/tree/main (9.10.2023)
[13] Polygon Zero. URL: https://polygon.technology/blog/polygon-announces-the-worlds-first-zero-knowledge-zk-scaling-solution-fully-compatible-with-ethereum (10.10.2023)
[14] Kateryna Kuznetsova, Solving blockchains valability problem using ZK-SNARK technology. Masterwork: 125–Cybersecurity / Kateryna Kuznetsova; KarazinKharkivNationalUniversity – Kharkiv: 2023.– 80 p.

Надійшла до редакції 18 жовтня 2023 р. ; Переглянута 24 листопада 2023 р. ; Прийнята 25 грудня 2023 р.

Автори:
Катерина Кузнецова, студентка факультету комп'ютерних наук (магістр), Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків, 61022, Україна.
E-mail:kate7smith12@gmail.com
ORCID ID https://orcid.org/0000-0002-5605-9293
Антон Єжов, співзасновник Zpoken.io (https://zpoken.io/), OU, Harju maakond, м. Таллінн, Kesklinnalinnaosa, Sakala tn 7-2, 10141, Естонія.
E-mail:anton.yezhov@zpoken.io

Використання ZK-SNARK для вирішення проблеми масштабованості блокчейн.

Анотація. Роботу присвячено викладенню основних концепцій технології блокчейн та опису ключових параметрів роботи блокчейн-технології для викладення проблеми масштабованості блокчейн мереж та аналізу її особливостей, вивчення існуючих напрямів вирішення масштабованості блокчейн, аналіз та порівняння відомих протоколів. Для детального вивчення було обрано технологію доказів з нульовим знанням, на основі протоколів якої розроблено систему перевірки валідності ланцюга блоків. Наведені експериментальні дослідження обґрунтовують перспективність даного напрямку для вирішення проблем масштабованості сучасних блокчейн систем. Актуальність обраної теми зумовлена необхідністю впровадження блокчейн систем в різні галузі людського життя. Однак, із розвитком будь-якої мережі зростає об'єм інформації, що необхідно безперервно обробляти. Цей виклик змушує розробляти рішення для вдосконалення систем, роблячи їх гнучкими у роботі з мільйонами користувачів. Водночас вкрай важливим питанням є підтримка безпеки та конфіденційності даних в оновлених системах та дотримання децентралізованої організації процесу обміну даними. Отже, у сучасному світі блокчейн індустрії головним питанням є пошук моделей та методів для вирішення проблеми масштабованості мереж для подолання бар'єру повномасштабного впровадження блокчейн додатків.

Ключові слова: *блокчейн, трилемаблокчейн, проблема масштабованості блокчейн, докази з нульовим знанням, ZK-SNARK, PLONK, FRI.*

DOI: 10.26565/2519-2310-2023-2-03
УДК 004.056.5

ПОРІВНЯЛЬНИЙ АНАЛІЗ ШТУЧНОГО ІНТЕЛЕКТУ НА ОСНОВІ ІСНУЮЧИХ ЧАТ-БОТІВ

Кобилянська Олена[1], Єсіна Марина[1,2],Горбенко Юрій[2]
[1]Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків, 61022, Україна
e-mail: kobol1801@gmail.com, ORCID:https://orcid.org/0000-0003-3405-3429 ,
e-mail: m.v.yesina@karazin.ua, ORCID: https://orcid.org/0000-0002-1252-7606
[2]АТ «ІІТ», вулиця Коломенська, 15, Харків, 61166, Україна
jsciitua@gmail.com
Надійшла до редакції 1 листопада 2023 р. Переглянута 2 грудня 2023 р. Прийнята 25 грудня 2023 р.

Анотація: У даній роботі представлено комплексний аналіз двох провідних систем штучного інтелекту (ШІ) – ChatGPT-4 від OpenAI та Bard від Google AI. Також наводиться огляд розвитку штучного інтелекту в різних галузях та його впливу на повсякденне життя людини, особливо в таких сферах, як медицина, фінанси, державне управління тощо. Проводиться заглиблення в детальне порівняння різних версій ChatGPT (GPT-3 та GPT-4), шляхом обговорення та аналізу їхніх можливостей, вдосконалення та обмежень. У статті також розглядається інтеграція системи Bard із сервісами Google, її унікальні функціональні можливості та останні оновлення. Мета дослідження полягає в порівнянні можливостей систем штучного інтелекту ChatGP-4T та Bard, висвітленні їхніх сильних і слабких сторін, а також їх практичного застосування. Проведено порівняльне тестування для оцінки продуктивності кожної моделі (системи) в різних завданнях, включаючи розв'язання логічного завдання, написання есе, аналіз із подальшим внесення пропозицій щодо покращення веб-сайту та написання коду HTML/CSS для веб-сторінки. Результати підкреслюють той факт, що, незважаючи на визнані переваги цих моделей, їхні функціональні характеристики іноді можуть бути обмежені або не відповідати очікуванням при виконанні специфічних завдань, а вибір системи (моделі) буде коригуватися у залежності від потреб користувачів.
Ключові слова: *ChatGPT-4,Bard,OpenAI,GoogleAI,штучний інтелект.*

1.Вступ

На сьогоднішній день, штучний інтелект (ШІ) швидко набуває популярності у різних секторах, включаючи корпоративний світ, бізнес-кола та повсякденне життя людей. Застосування ШІ в областях, як-от медицина, банківська сфера та урядові структури, стає все частішим. ШІ полегшує обробку даних, оскільки вона відбувається без втручання людської праці та зазвичай забезпечує точність виконаних завдань. Згідно зі статистикою, у 2023 році 35% компаній використовували ШІ у своїй діяльності, а 90% організацій вважають ШІ важливою для досягнення конкурентних переваг [1].

Системи штучного інтелекту впливають і на людське повсякдення, спрощуючи наступні аспекти їх діяльності: планування та організація денних справ, використання засобів ефективності у фінансах, навчанні та здоров'ї тощо. Завдяки йому, суспільство може ефективніше використовувати свій час, отримуючи доступ до швидкої та точної інформації.

Дана стаття зосереджена на аналізі особливостей двох провідних систем штучного інтелекту –*Bard* та *ChatGPT*. Вона включає в себе практичне порівняння однакових параметрів обох систем, а також виявлення переваг та недоліків кожної з них.

2. Огляд мовної моделі ChatGPT

ChatGPT, створена OpenAI, є системою генерації тексту, яка належить до серії *GPT* (*Generative Pretrained Transformer*). Базуючись на трансформерній архітектурі, ця модель навчена на великих масивах текстових даних для генерації даних подібних за стилем написання до тексту, створеним людиною. Розроблена для реагування на запити користувачів, *ChatGPT* підходить для використання у діалогових програмах, таких як чат-боти, обслуговування клієнтів та віртуальні асистенти. Ця модель була тренована на даних з різних джерел, таких як

Інтернет-ресурси, книги та соцмережі, що дозволяє їй створювати зв'язні та контекстуальні текстові відповіді. Щоб використовувати *ChatGPT*, користувач подає підказку, таку як питання або коментар, і модель генерує відповідь, враховуючи отримані дані та своє попереднє навчання. Однією з головних переваг *ChatGPT* є її здатність до контекстуально релевантного тексту. Наприклад, при запитанні про моду, модель може надати інформацію, що включає наступні слова: стиль, вбрання, крій. *ChatGPT* також може продовжувати діалог, використовуючи попередню розмову як контекст. *ChatGPT* також застосовується для інших задач, таких як відповіді на питання, узагальнення та класифікація тексту, завдяки доопрацюванням під конкретні цілі. Ця модель є частиною більш широкої тенденції використання великих мовних моделей для застосунків, що має потенціал перетворити спосіб взаємодії з технологіями та спілкування з пристроями на більш природній і інтуїтивно зрозумілий[2].

Вище було представлено загальний огляд моделі *ChatGPT*. Далі ми зосередимося на порівнянні двох версій цієї моделі: *ChatGPT-3*, що з'явилася у 2020 році, та *ChatGPT-4*, випущеної у 2023 році. Це дозволить нам визначити, яка з цих моделей краще підходить для порівняльного аналізу з моделлю *Bard*.

ChatGPT-3 вирізняється своєю високою здатністю до розуміння та створення текстів. Він навчений на обширному спектрі Інтернет-даних, що надає йому широкі знання. Ця модель ефективно виконує багато завдань, створюючи оригінальні тексти. Однак, вона може давати неточні відповіді та має тенденцію до упередженості, особливо у складних сценаріях (тобто, умовно кажучи - *може «галюцинувати»*).

ChatGPT-4, з іншого боку, покращив здатність розрізняти та відповідати на більш складні питання завдяки удосконаленій трансформерній архітектурі. Модель отримала більше навчальних даних і зменшила частоту помилок порівняно з попередніми версіями. *ChatGPT-4* вирішує складні завдання точніше та надійніше, показуючи краще розуміння контексту. Також, до функціоналу системи був доданий наступний функціонал: обробка та генерація графічних зображень, додаткові утиліти на обробку файлів обсягом більш, ніж 50 сторінок. Однак, попри поліпшення, вона все ще схильна до деяких помилок, і її складність може потребувати більше ресурсів. У табл. 1 наведена порівняльна характеристика поданих моделей.

Таблиця 1 – Порівняльна характеристика *GPT-3*та*GPT-4*
Table 1 – Comparative characteristic*GPT-3&GPT-4*

Характеристики	**GPT-3**	**GPT-4**
Параметри	*175 млрд*	*наразі невідомо*
Модальність	*текст*	*текст і зображення*
Продуктивність	*слабка у вирішенні складних задач*	*на одному рівні із людиною*
Галюцинації	*схильність до упередженості та помилок*	*менш упереджена та більш стабільна*

Пояснимо деякі поняття із табл. 1 відносно даного дослідження:

1. У контексті мовних систем, категорія «параметри» відносяться до налаштованих внутрішніх змінних або інших налаштувань. Більша кількість параметрів вказує на те, що ця модель краще пристосована до вивчення та узагальнення закономірностей на основі даних, на яких вона «навчалася». *GPT-3* була випущена з 175 мільярдами параметрів, що робить її

однією з найбільших великих моделей (*LargeLM*). Про параметри *GPT-4* офіційно не повідомлялося, але можна з упевненістю казати, що їх кількість значно перевищує 175 млрд.

2. *GPT-3* є унімодальною, тобто може приймати лише текстові дані. Вона може обробляти і генерувати різні текстові форми, але не може обробляти зображення або інші типи даних. *GPT-4* є мультимодальною, вона може приймати і створювати текстові і графічні вхідні та вихідні дані, що робить її набагато різноманітнішою. Вона, також, може виконувати більш складні завдання, які вимагають поєднання текстової та графічної вихідної інформації, такі як підписи, підбиття підсумків або переклад зображень.

3. Продуктивність системи визначається її здатністю адекватно реагувати на вхідні запити. Це відображає, наскільки успішно модель вловлює суть мови та надає значущі відповіді. Таку ефективність зазвичай вимірюють за критеріями, як: «збентеженість», «точність» і «плавність». Завдяки збільшеній кількості параметрів та розширеним мультимодальним можливостям, *GPT-4* випереджає *GPT-3* у термінах її продуктивності.

4. Галюцинації в моделі – це «відповіді», які не мають сенсу або не мають відношення до отриманих вихідних даних. Це відбувається тому, що модель покладається на свої первинні навчальні дані або знання, щоб генерувати наступні відповіді на основі вивчених шаблонів. У роботі [3]зазначається, що ймовірність галюцинацій у *GPT-3* становить від 15% до 20%. Хоча наразі невідомо, наскільки *GPT-4* схильна до галюцинацій, генеральний директор комп. *OpenAI* Сем Альтман каже, що *«вона галюцинує значно менше...»*.

Зважаючи на усі аргументи, доходимо висновку: - *GPT-4* перевершує *GPT-3* у ефективності, що є логічним, враховуючи, що кожне нове покоління моделі покращується, виправляючи недоліки та вносячи значні удосконалення. Тому, для порівняння із *Bard*, обираємо модель *GPT-4*, оскільки вона виявляє менше помилок у відповідях, має вищу точність та підтримує мультимодальні функції.

3. Огляд мовної моделі Bard

Bard API від Google– це інструмент, який дозволяє розробникам отримувати доступ до даних з різних джерел і використовувати їх. Він використовує обробку природної мови (*NLP*) для вилучення інформації з різних типів документів, таких як веб-сайти, PDF-файли та інші текстові формати. Окрім доповнення пошуку Google, *Bard* може бути інтегрований у веб-сайти, платформи обміну повідомленнями або додатки для надання реалістичних відповідей природною мовою на запитання користувачів.

У грудні 2023 року Google Bard був оновлений за допомогою новітньої мовної моделі *Gemini*. Ця модель, разом із такими попередниками, як *Pathways Language Model 2 (PaLM 2)* та *Google's Language Model for Dialogue Applications (LaMDA)*, створена на основі архітектури *Transformers*, розробленої Google в 2017 році. Завдяки відкритому вихідному коду Transformer, ця архітектура лягла в основу численних інших генеративних інструментів штучного інтелекту, в тому числі мовної моделі *GPT-3*, яка використовується в *ChatGPT*.

Bard зосереджений на пошукових можливостях, намагаючись забезпечити більш природне використання мовних запитів замість стандартних ключових слів. Його штучний інтелект навчається на основі реальних діалогів, пропонуючи не просто відповіді, а контекстуалізовану інформацію. *Bard* розроблено також для обробки додаткових запитань, що є новинкою у сфері пошуку. Має функції для спільної роботи та подвійної перевірки результатів, допомагаючи користувачам у перевірці отриманої інформації. Він також інтегрований з різними додатками та сервісами Google, включаючи *YouTube, Maps, Hotels, Flights, Gmail, Docs* та *Drive*, дозволяючи користувачам використовувати його для роботи з особистим контентом.

GoogleBard, з його розширеними можливостями штучного інтелекту, пропонує користувачам ряд унікальних функцій. Ось деякі з ключових:

1. Інтеграція з *Google Lence* для читання зображень. Тепер став можливий аналіз зображення, розширюючи свої можливості у роботі з діалоговим текстом.

2. Генерація зображень. Розробники додали функцію створення зображень, покращуючи візуальні можливості.

3. Візуальна інформація до відповідей. *Bard* здатен доповнювати текстові відповіді візуальною інформацією для глибшого розуміння.

4. Широка інтеграція з сервісами Google. Ефективне інтегрування з Google сервісами, такими як карти, документи, таблиці та інші.

5. Підтримка плагінів. Плагіни для розширення своєї функціональності, включаючи інтеграцію з іншими веб-сайтами і компаніями.

6. Збереження чернеток. Вирішена проблема втрати відповідей при оновленні вкладки.

7. Експорт чату одним кліком. Функція експорту відповідей у *Bard* дозволяє користувачам легко зберігати свої відповіді для подальшого використання [4].

4. Порівняльний аналіз ChatGPT-4 та Bard

Розглядаємо практичне порівняння двох передових мовних моделей –*ChatGPT-4* від *OpenAI* і *Bard* від Google. Обидві ці системи базуються на найновіших досягненнях у галузі штучного інтелекту та нейронних мереж, але водночас пропонують унікальні характеристики та функціональності. Зосередимося на порівнянні їхніх особливостей, здатності до виконання різноманітних завдань, а також розглянемо їхні сильні та слабкі сторони в контексті різних сценаріїв використання. Це порівняння допоможе користувачам зрозуміти, яка з цих моделей краще підходить для їхніх конкретних потреб.

Перший тест включатиме вирішення головоломки для підлітків. Задача формулюється так: «*У 12-поверховій будівлі є лише один ліфт. На нижньому поверсі проживає 2 людини, і на кожному наступному поверсі кількість мешканців подвоюється. На якому поверсі цього будинку найчастіше використовують кнопку виклику ліфта?*». Логічно, що найбільше викликів ліфта відбувається на 1-му поверсі. Відповіді цих моделей, наведено на рис. 1.

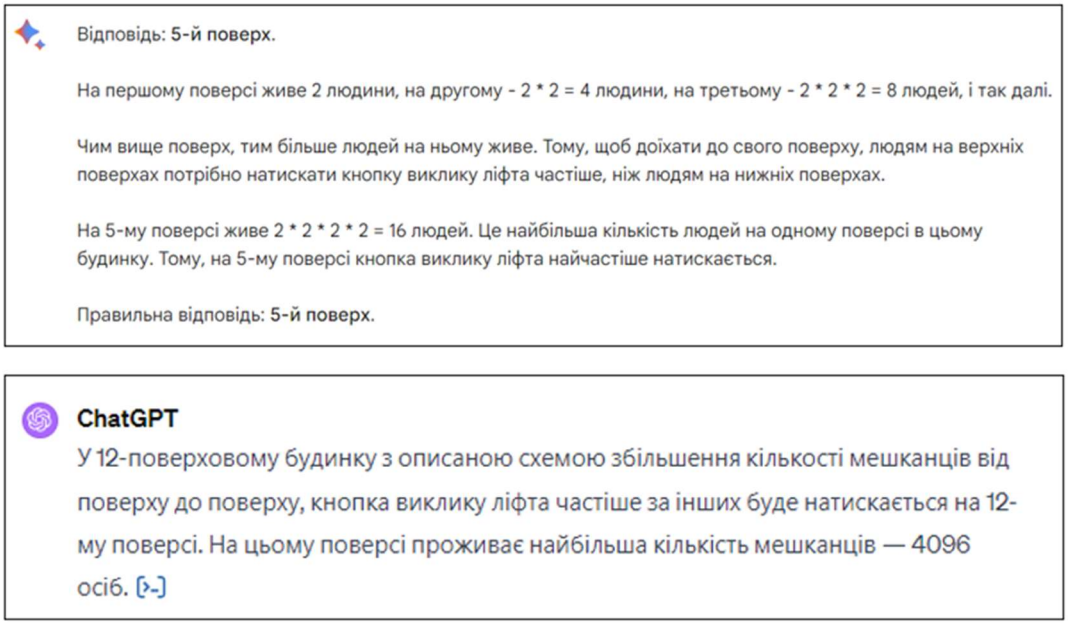


Рис. 1 – Відповіді моделей на логічну задачу
Fig.1 – Answers of models to a logical problem

З результатів тестування слід, що жодна з моделей не надала правильної відповіді. Проте, на користь *ChatGPT-4*, він правильно визначив кількість жителів на 12-му поверсі, тоді як *Bard* зупинився на обчисленнях для 4-го поверху та помилково зазначив 5-й поверх у своїй відповіді.

Другий тест полягатиме у створенні есе на тему «Захист особистих даних в інформаційному просторі» з обмеженням у 2000 символів. Результати цього завдання будуть представлені у порівняльній табл.2. Аналізуючи таблицю, слід відзначити, що *ChatGPT-4* дотримався встановленого ліміту з відхиленням у 14%, в той час як *Bard* перевищив заданий обсяг майже на 75%. Відношення кількості унікальних слів до загальної кількості складає 72% для *ChatGPT-4* і 57% для *Bard*. Середня кількість слів на речення у кожної моделі становить 10% від загального обсягу. Приблизно чверть, створеного *ChatGPT-4*, містить неважливу інформацію, в той час як у *Bard* цей показник становить п'яту частину від усього тексту.

Таблиця 2 – Порівняння створених есе *ChatGPT-4*та*Bard*
Table 2 – Comparison of generated essays *ChatGPT-4*&*Bard*

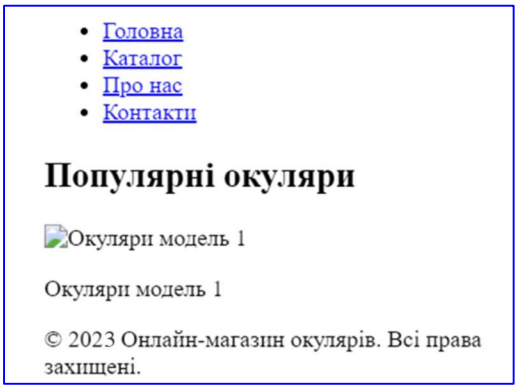
Параметр	ChatGPT-4	Bard
Кількість символів	*1726*	*3496*
Кількість слів	*198*	*364*
Унікальних слів	*142*	*206*
Кількість речень	*19*	*36*
Час на прочитання	*1 хв.*	*2 хв.*
Неважлива інформація	*26%*	*19%*

Сутність 3-го завдання полягала у вдосконаленні веб-сайту Харківського національного університету імені В. Н. Каразіна, зосереджуючись не просто на описі розділів, а на його поліпшенні для користувачів. В результаті *ChatGPT-4* запропонував такі рішення: - оновлення дизайну сайту;- розробка мобільної версії; - створення форуму чи чату для обміну інформацією між абітурієнтами й студентами;- приведення сайту у відповідність з сучасними стандартами веб-безпеки. *Bard* же висунув інші «ідеї»: - додавання розділів про історію і традиції університету, сучасне життя в університеті, відомих випускників та наукові досягнення; - покращення доступності для людей з обмеженими можливостями, включаючи збільшення шрифту і адаптацію для людей із вадами зору; - фонові звуки; - додавання розділів з відгуками, новинами та подіями й картою університету. Таким чином, моделі висунули помітно різні пріоритети, стосовно того, що важливіше для сайту: - забезпечення його безпеки або зручність та інформативність для його користувачів.

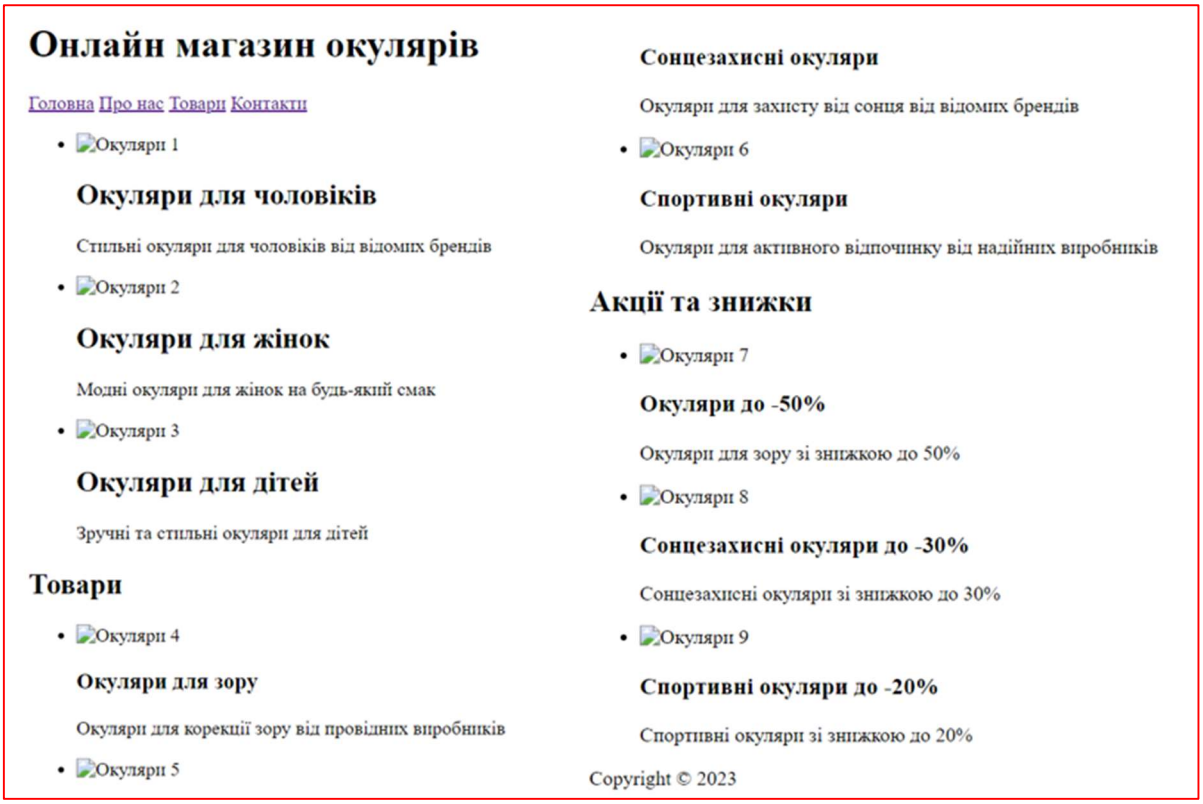
У 4-му тесті, обрані моделі займалися розробкою HTML та CSS коду для головної сторінки умовного Інтернет магазину окулярів. Оцінюючи результати, які представлені на рис. 2, можна відзначити, що *Bard* виявився більш ефективним у виконанні завдання. На головній сторінці, створеній саме *Bard*, були не тільки основні посилання на асортимент, контакти та повернення на головну сторінку, але й впорядковані категорії товарів, такі як чоловічі й жіночі окуляри, окуляри для корекції зору, а також спеціальні пропозиції і знижки.

Отже, підбиваючи підсумки всіх тестових завдань, виконаних обома моделями, слід зазначити, що вибір між ними залежатиме від специфічних потреб користувачів. Так, за однакових умов обидві моделі показали різні результати, іноді навіть відступаючи від своїх звичайних «сильних» сторін. Наприклад, хоча *ChatGPT-4* часто рекомендується для виконання

завдань програмування, у цьому порівняльному аналізі виконання тестових завдань, він показав менш значущі результати, ніж *Bard*. Водночас, *Bard* не зміг ефективно впоратися із простими завданнями на розв'язання звичайної логічної задачі.



a)



b)

Рис.2– Результати компіляції коду,створеного *ChatGPT-4 (a)* та *Bard (b)*
Fig. 2 – Results of compiling the code generated by*ChatGPT-4 (a)*&*Bard (b)*

5.Висновки

У роботі представлений порівняльний аналіз роботи двох провідних моделей штучного інтелекту – *ChatGPT-4*та*Bard*. В результаті виконання низки тестових завдань було підтверджено, що вибір між необхідною моделлю, залежить від конкретних потреб її користувачів, оскільки кожна з них демонструє помітно різні результати.

До переваг *ChatGPT-4* (за проведеними дослідженнями) слід віднести точні математичні розрахунки, виконання задач з мінімальними відхиленнями від умов, а також конкретні поради для поліпшення веб-сайту. На відміну від нього, *Bard* підтвердив більш широкий підхід до завдань, виходячи за рамки заданих умов та пропонуючи користувачам більш актуальні

(варіативні) рішення. Щодо недоліків, то обидві моделі демонструють певні «слабкості» в алгоритмах «логічного мислення». Також, тестування на генерацію зображень не проводилося через обмеження однієї з моделей, проте обидві системи продовжують безперервно розвиватися й навчатися, що скоріш за все, буде реалізовано в найближчому майбутньому.

References

[1] Webster M. (October 6, 2023) 149 AI Statictics: The Present And Future Of All At Your Fingerprints. authorityhacker.com/ai-statistics/

[2] Md Sakibul Islam Sakib (February 2023) What is ChatGPT? http://surl.li/pqywz

[3] Ayush Kudesia (March 28, 2023) GPT 3 vs. 4: Know The Difference https://fireflies.ai/blog/gpt3-vs-4.

[4] What is Bard (Google AI)? Everything you need to know https://instagantt.com/project-management/what-is-bard-google-ai.

Submitted November 1, 2023; Revised December 2, 2023; Accepted December 25, 2023

Authors:

Kobylianska Olena, CSD Student, Department of Security of Information Systems and Technologies, V.N. Karazin Kharkiv National University, Ukraine.
E-mail: kobol1801@gmail.com
ORCID: https://orcid.org/0000-0003-3405-3429

Yesina Maryna, Ph.D., Associate Professor, Department of Security of Information Systems and Technologies, V. N. Karazin Kharkiv National University, Ukraine.
E-mail:m.v.yesina@karazin.ua
ORCID: https://orcid.org/0000-0002-1252-7606

Yurii Gorbenko, Ph.D.,firstdeputychiefdesignerof JSC"IIT", Kharkiv, Ukraine.
E-mail:jsciitua@gmail.com

Comparative analysis of artificial intelligence based on existing ChatBots.

Abstract.This paper presents a comprehensive analysis of two leading artificial intelligence (*AI*) systems – *ChatGPT-4* from *OpenAI* and *Bard* from *Google AI*. It also provides an overview of the development of artificial intelligence in various fields and its impact on human daily life, especially in areas such as medicine, finance, public administration, etc. A detailed comparison of different versions of *ChatGPT* (GPT-3 and *GPT-4*) is carried out by discussing and analyzing their capabilities, improvements, and limitations. The article also discusses the integration of the *Bard* system with Google services, its unique functionality, and the latest updates. The purpose of the study is to compare the capabilities of *ChatGP-4T* and *Bard AI* systems, highlight their strengths and weaknesses, as well as their practical application. Comparative testing was conducted to evaluate the performance of each model (*system*) in various tasks, including solving a logical problem, writing an essay, analyzing followed by making suggestions for improving the website and writing *HTML/CSS* code for a web page. The results highlight the fact that, despite the recognized advantages of these models, their functional characteristics may sometimes be limited or not meet expectations when performing specific tasks, and the choice of system (*model*) will be adjusted depending on the needs of users.

Keywords: *ChatGPT-4,Bard, OpenAI, GoogleAI, Artificial Intelligence.*

DOI: 10.26565/2519-2310-2023-2-04
UDC004.056.5

METHODS FOR DETERMINING THE CATEGORIES OF CYBER INCIDENTS AND ASSESSING INFORMATION SECURITY RISKS

Kopytsia Oleksandr, Uzlov Dmytro

V. N. Karazin Kharkiv National University, 6 Svobody Sq., Kharkiv, 61022, Ukraine
oleksandr.kopytsia@student.karazin.ua, dmytro.uzlov@karazin.ua ORCID: https://orcid.org/0000-0003-3308-424X

Submitted October 25, 2023; Revised November 30, 2023; Accepted December 22, 2023

Abstract: The article is devoted to the study of categories of cyber incidents and their prioritization in the context of information security. It discusses the main sources that provide information about cyber threats and defines their role in detecting and analyzing incidents, and provides tools for collecting and analyzing data. The concepts of event, incident, and crime and the relationship between them are discussed. The author provides a classification of various types of cyber threats, how they are coded, their characteristics and impact on information systems. Examples of the use of cyber incident classification are given. The authors of the article also consider specific types of cyber incidents that may occur in various fields of activity and the threats they pose to various information systems. The necessity and methods of determining priorities in responding to cyber threats are substantiated, which allows for the effective allocation of resources and the implementation of preventive cyber security measures. The approach to assessing and classifying incidents according to their possible impact on the organization's activities, information security and ability to recover from cyber attacks is revealed. The article highlights various approaches and methodologies for identifying and managing information security risks, including the use of standards, models and assessment tools. This article is a resource for cybersecurity professionals, researchers, and executives interested in risk management and information asset protection in today's digital environment.

Keywords: *Cyber Security, Cyber Incident, Intrusion Detection System, Categories of Cyber Incidents, Prioritization of Incidents, Information Security Risks.*

1. Introduction

Cybersecurity in today's world is defined as a critical component of security as our society becomes increasingly digital. Means of protecting personal information, information systems and data of corporations and financial institutions, government agencies and critical infrastructure help to reduce the risks of cyberattacks and their consequences. Given the rapid technological development, the importance of cybersecurity is increasing as new technologies, such as the Internet of Things and artificial intelligence, create new vulnerabilities that require effective protection strategies. Thus, cybersecurity is becoming essential to ensure stability, protect personal information and national interests, requiring cooperation between government, business and civil society to develop and implement effective measures.

Prioritizing the handling of cyber incidents depending on the risks they pose to information systems is a crucial element of effective cyber defense. This allows cybersecurity professionals to optimize the use of resources, directing them to the most critical scenarios and minimizing possible losses for the organization. Rapid response to high-risk cyber incidents ensures that critical systems remain functional and helps to avoid negative consequences for business processes. Taking risks into account also helps to take preventive measures, improve security strategies, and comply with regulatory requirements. This systematic approach to cybersecurity management allows it to effectively detect, respond to, and prevent cyber threats, providing reliable protection for information systems and preserving the organization's reputation.

2. Detecting cyber security incidents

Collecting and analyzing data on cyber incidents is a task that presents a number of challenges and complexities. First, information about cyber threats can be scattered across a variety of sources, such as system logs, network data, information from antivirus systems, vulnerability reports, etc. This

requires the development of a comprehensive strategy for collecting and integrating data from various sources.

An additional challenge is that attackers are constantly improving their methods, using new technologies and tactics to evade detection. This poses a challenge for cybersecurity analysts: to constantly update their knowledge and tools to effectively detect and analyze new threats.

When it comes to tools for collecting and analyzing cyber incident data, there are a variety of software and hardware tools. Software tools include security intrusion detection systems (*SIEMs*), which provide centralized log collection and analysis, as well as intrusion detection systems (*IDSs*) and vulnerability detection systems (*VDSs*). Some platforms, such as Splunk, ELK Stack, or IBM QRadar, allow you to aggregate data from different sources and provide event correlation capabilities to identify potential threats.

There are also advanced tools for analyzing network traffic, such as Wireshark, or for detecting anomalies in systems, such as Darktrace. It is also important to use intelligent data analysis systems based on artificial intelligence (*AI*) to automatically detect anomalies and patterns that may indicate cyber threats.

The following event logs can be used by an organization to assist with detecting and investigating cyber security incidents [1]:

- Cross Domain Solutions: May assist in identifying anomalous or malicious network traffic indicating an exploitation attempt or successful compromise.
- Databases: May assist in identifying anomalous or malicious application or user behavior indicating an exploitation attempt or successful compromise.
- Domain Name System services: May assist in identifying attempts to resolve malicious domain names or Internet Protocol (IP) addresses indicating an exploitation attempt or successful compromise.
- Email servers: May assist in identifying users targeted with phishing emails thereby helping to identify the initial vector of a compromise.
- Gateways: May assist in identifying anomalous or malicious network traffic indicating an exploitation attempt or successful compromise.
- Multifunction devices: May assist in identifying anomalous or malicious user behavior indicating a cyber security incident or malicious insider activity.
- Operating systems: May assist in identifying anomalous or malicious activity indicating an exploitation attempt or successful compromise.
- Remote access services: May assist in identifying unusual locations of access or times of access indicating an exploitation attempt or successful compromise.
- Security services: May assist in identifying anomalous or malicious application or network traffic indicating an exploitation attempt or successful compromise.
- Server applications: May assist in identifying anomalous or malicious application behavior indicating an exploitation attempt or successful compromise.
- System access: May assist in identifying anomalous or malicious user behavior indicating an exploitation attempt or successful compromise.
- User applications: May assist in identifying anomalous or malicious application or user behavior indicating an exploitation attempt or successful compromise.
- Web applications: May assist in identifying anomalous or malicious application or user behavior indicating an exploitation attempt or successful compromise.
- Web proxies: May assist in identifying anomalous or malicious network traffic indicating an exploitation attempt or successful compromise.

3. Categories of cyber incidents

Not all events recorded in logs are directly indicative of cyber incidents, and this is due to several factors. First, log files include a wide range of information that can be the result of normal system or network operation. Many events can be related to normal operations, system updates, or even erroneous questions from users. Secondly, not every unusual or anomalous event is a cyber incident. Some anomalies can be the result of temporary system malfunctions, misconfigurations, or random events. Without the proper context and analysis, it is difficult to determine whether an event poses a real cybersecurity threat.

For the purpose of defining categories of incidents it is important to have a clear concept of the different scopes of an event, an incident and a crime.

An event can be defined as any observable occurrence that happened at a point in time in a system or network, especially one of importance. Thus, an event does not necessarily imply an adverse situation or a malicious activity [2].

For instance, «*to send an email*» or «*to make a phone call*» are events with no malicious implication.

On the other hand, a security incident necessarily implies a human-caused adverse event, usually with a malicious nature, which is oriented to cause a disruption of any system or network.

It is important to underline that incidents arising from negligence, as well as attempts that fail, also fall under the concept of a security incident. Examples of security incidents are «*SQL injection*» or «*Cross-Site Scripting*» attacks.

As can be observed below in Fig. 1, any security incident is considered an event but not any event is considered a security incident.

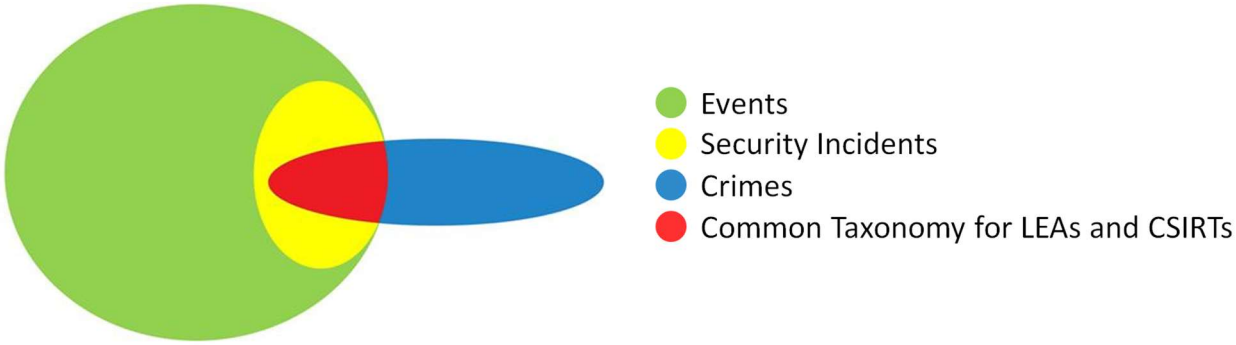


Fig 1 - Events, Security incidents, Crimes and Common Taxonomy ecosystem

Also not every security incident has a crime penalty, therefore only the security incidents able to be criminally prosecuted will be the ones falling under the scope of Common Taxonomy for *LEAs* and *CSIRTs* To clarify this, see the Fig. 1 below.

Different categories of cyber incidents manifest themselves in different ways in information systems and have different impacts on them. The threat level of a cyber incident may depend on its category.There are many different lists of cyber incident categories that take into account different aspects and characteristics of digital threats such as the type of attack, privacy impact, attack targets, and methods used by attackers. For example, the State Service for Special Communications and Information Protection of Ukraine provides the following list, which is developed using and complies with the recommendations of the *European Cyber Security Agency* (*ENISA Reference Incident Classification Taxonomy*), as well as the joint document of ENISA and the *European Cybercrime Centre Europol* (*Common Taxonomy for Law Enforcement and The National Network of CSIRTs*) [3].

According to the Table 1, a cyber incident can be described using the incident category code and the incident type code:

Example 1:Incident code: 01.01; Incident type: Spam.

Example 2: Incident code: 02.04; Incident type: Malicious connection.

Table 1 - Categories of cyber incidents

Code xx	Incident category	Code xx	Type of incident	Description of the type of incident
1	*2*	*3*	*4*	*5*
01.	**Abusive content**	01	Spam	*Sending unwanted messages or a large number of messages (flooding)*
02.	**Malicious Code**	01	Malware infection	*Spyware detected in the system*
		02	Malware distribution	*Distributing spyware, for example, by sending out emails containing malware attachments or links to download it.*
		03	Command & Control (C2)	*A system that is used as a command and control point for a botnet and/or serves as a collection point for information stolen by botnets.*
		04	Malicious connection	*Connection attempts from/to IP/URL - an address associated with a known spyware, such as C2C, or a distribution resource for components associated with a particular botnet activity.*
03.	**Information Gathering**	01	Scanning	*Collecting information about systems or networks.*
		02	Sniffing	*Unauthorized interception (logical or physical) and analysis of network traffic. Unauthorized monitoring and reading of network traffic.*
		03	Phishing	*An attempt to collect information about a user or system using social engineering techniques (mass emails aimed at collecting data, may contain links to phishing sites)*
04.	**Intrusion Attempts**	01	Vulnerability exploitationattempt	*Attempted intrusion by exploiting a vulnerability in a system, component, or network*
		02	Login attempts	*An attempt to log in to services or authentication/access mechanisms. An unsuccessful attempt to match authentication credentials or use previously compromised credentials that are no longer relevant.*

Continuation of the Table 1

1	*2*	*3*	*4*	*5*
05.	**Intrusion**	01	Account compromise	*Actual intrusion into a system, component or network by compromising a user or administrator account*
		02	System compromise	*An actual intrusion into a system or its component, service, or application through the exploitation of a vulnerability in a component or network. Unauthorized access to a system or component bypassing the access control system.*
06.	**Availability**	01	DoS/DDoS	*An impact on the normal functioning of a system or service that is achieved by sending requests from one or more sources to the target resource to overload the bandwidth or system resources.*
		02	Sabotage	*Actions (intentional or unintentional) aimed at damaging the system, interrupting processes, changing or deleting information, etc.*
		03	Outage, no malice	*Failure of a system or its components without malicious interference.*
07.	**Information Content Security**	01	Unauthorized access to information	*Unauthorized access to information. Unauthorized sharing of a specific set of information.*
		02	Unauthorized modification of info	*Unauthorized modification or deletion of a certain set of information.*
08.	**Fraud**	01	Fraudulent site	*Creating phishing sites to collect authentication or other user data. Using the institution's resources for purposes other than those intended.*
09.	**Vulnerable**	01	Vulnerability	*The presence of known vulnerabilities in the system or its components that are open to exploitation.*
		02	Misconfiguration	*Flaws in the settings that can be exploited by an attacker (default settings, etc.).*
10.	**Other**	01	Undetermined incident	*Insufficient data to process the incident.*

4. Incident prioritization

Prioritizing the handling of the incident is perhaps the most critical decision point in the incident handling process. Incidents should not be handled on a first-come, first-served basis as a result of resource limitations. Instead, handling should be prioritized based on the relevant factors, such as the following [4]:

- Functional Impact of the Incident. Incidents targeting IT systems typically impact the business functionality that those systems provide, resulting in some type of negative impact to the users of those systems. Incident handlers should consider how the incident will impact the existing functionality of the affected systems. Incident handlers should consider not only the current functional impact of the incident, but also the likely future functional impact of the incident if it is not immediately contained.
- Information Impact of the Incident. Incidents may affect the confidentiality, integrity, and availability of the organization's information. For example, a malicious agent may exfiltrate sensitive information. Incident handlers should consider how this information exfiltration will impact the organization's overall mission. An incident that results in the exfiltration of sensitive information may also affect other organizations if any of the data pertained to a partner organization.
- Recoverability from the Incident. The size of the incident and the type of resources it affects will determine the amount of time and resources that must be spent on recovering from that incident. In some instances it is not possible to recover from an incident (*e.g., if the confidentiality of sensitive information has been compromised*) and it would not make sense to spend limited resources on an elongated incident handling cycle, unless that effort was directed at ensuring that a similar incident did not occur in the future. In other cases, an incident may require far more resources to handle than what an organization has available. Incident handlers should consider the effort necessary to actually recover from an incident and carefully weigh that against the value the recovery effort will create and any requirements related to incident handling.

Combining the functional impact to the organization's systems and the impact to the organization's information determines the business impact of the incident—for example, a distributed denial of service attack against a public web server may temporarily reduce the functionality for users attempting to access the server, whereas unauthorized root-level access to a public web server may result in the exfiltration of *personally identifiable information* (*PII*), which could have a long-lasting impact on the organization's reputation.

The recoverability from the incident determines the possible responses that the team may take when handling the incident. An incident with a high functional impact and low effort to recover from is an ideal candidate for immediate action from the team. However, some incidents may not have smooth recovery paths and may need to be queued for a more strategic-level response—for example, an incident that results in an attacker exfiltrating and publicly posting gigabytes of sensitive data has no easy recovery path since the data is already exposed; in this case the team may transfer part of the responsibility for handling the data exfiltration incident to a more strategic-level team that develops strategy for preventing future breaches and creates an outreach plan for alerting those individuals or organizations whose data was exfiltrated. The team should prioritize the response to each incident based on its estimate of the business impact caused by the incident and the estimated efforts required to recover from the incident [5].An organization can best quantify the effect of its own incidents because of its situational awareness.

Table 2 provides examples of functional impact categories that an organization might use for rating its own incidents. Rating incidents can be helpful in prioritizing limited resources.

Table 3 provides examples of possible information impact categories that describe the extent of information compromise that occurred during the incident. In this table, with the exception of the "None" value, the categories are not mutually exclusive and the organization could choose more than one.

Table 2 –Functional Impact Categories

Category	Definition
None	*No effect to the organization's ability to provide all services to all users*
Low	*Minimal effect; the organization can still provide all critical services to all users but has lost efficiency*
Medium	*Organization has lost the ability to provide a critical service to a subset of system users*
High	*Organization is no longer able to provide some critical services to any users*

Table 3 –Information Impact Categories

Category	Definition
None	*No information was exfiltrated, changed, deleted, or otherwise compromised*
Privacy Breach	*Sensitive personally identifiable information (PII) of taxpayers, employees, beneficiaries, etc. was accessed or exfiltrated*
Proprietary Breach	*Unclassified proprietary information, such as protected critical infrastructure information (PCII),was accessed or exfiltrated*
Integrity Loss	*Sensitive or proprietary information was changed or deleted*

Table 4 shows examples of recoverability effort categories that reflect the level of and type of resources required to recover from the incident.

Table 4 - Recoverability Effort Categories

Category	Definition
Regular	*Time to recovery is predictable with existing resources*
Supplemented	*Time to recovery is predictable with additional resources*
Extended	*Time to recovery is unpredictable; additional resources and outside help are needed*
Not Recoverable	*Recovery from the incident is not possible (e.g., sensitive data exfiltrated and posted publicly); launch investigation*

5. Criticality levels of cyber incidents

Taking into account the above, the following consider a list of criticality levels of cyber incidents developed by the State Service for Special Communications and Information Protection of Ukraine [6]:

- *level 0, non-critical (white)* - a cyber incident/cyber attack does not threaten the sustainable, reliable and normal operation of information, electronic communication, information and communication systems, technological systems;
- *level 1, low (green)* - a cyber incident/cyber attack directly threatens the sustainable, reliable and normal operation of information, electronic communication, information and communication systems, technological systems, but does not threaten the security (*confidentiality, integrity and availability*) of information and data processed by them;
- *level 2, medium (yellow)* - a cyber incident/cyber attack directly threatens the sustainable, reliable and normal operation of information, electronic communication, information and communication systems, technological systems, which creates prerequisites for violating the security (*confidentiality, integrity and availability*) of information and data processed by

them, and creates prerequisites for the termination of functions and/or provision of services by critical infrastructure;

- *level 3, high (orange)* - a cyber incident/cyber attack directly threatens the stable, reliable and normal operation of information, electronic communication, information and communication systems, technological systems, violates the security *(confidentiality, integrity and availability)* of information and data processed by them, resulting in potential threats to national security and defense, the state of the environment, the social sphere, the national economy and its individual sectors, and the termination of business. Response at this level may require the involvement of forces and means of more than one main actor of the national cybersecurity system;
- *level 4, critical (red)* - a cyber incident/cyber attack directly threatens the stable, reliable and normal operation of several information, electronic communication, information and communication systems, technological systems, violates the security (confidentiality, integrity and availability) of information and data processed by them, resulting in real threats to national security and defense, the state of the environment, the social sphere, the national economy and its individual sectors, and the cessation of A cyber incident/cyber attack may have a cross-border impact. Response at this level requires the involvement of forces and means of the main actors of the national cybersecurity system;
- *level 5, emergency (black)* - a cyber incident/cyber attack directly threatens the sustainable, reliable and normal operation of a significant number of information, electronic communication, information and communication systems, technological systems, violates the security *(confidentiality, integrity and availability)* of information and data processed by them, resulting in imminent threats to the full functioning of the state or threats to the lives of Ukrainian citizens. A cyber incident/cyber attack may have a cross-border impact. Response at this level requires maximum involvement of the forces and means of the main actors of the national cybersecurity system and other cybersecurity actors.

6.Information security risk assessment methods

The development of Information Security Risk Assessment methods is a key element of effective cybersecurity and risk management in the modern information environment. This is important due to the complexity of cyber threats that are constantly changing and evolving. Today's information environment faces diverse and ever-changing cyber threats, and creating risk assessment methods helps identify, analyze, and manage these threats. Attackers are constantly developing new methods and techniques, so it is important to have effective methods to identify, assess, and manage these threats. Information is one of the most valuable assets for many organizations, so risk assessment methods help determine which data is most valuable and vulnerable. This makes it possible to develop strategies to protect it effectively.

Most organizations have limited resources, so it's important to allocate those resources effectively to maximize security. Risk assessment methods help to prioritize and cost cybersecurity measures. The risk assessment also takes into account compliance and regulatory requirements, helping to determine how well existing standards are met and where improvements can be made. The idea that risk assessment is a tool for proactively identifying potential problems and solving them before they lead to cyber incidents is important. Creating risk assessment methods is a strategically important task for any organization seeking to ensure reliable cybersecurity and reduce the impact of information threats.

There are a significant number of Information Security Risk Assessment (*ISRA*) methods that have been developed by various organizations. These methods help to identify, analyze and manage risks to ensure effective cyber defense:

- CIRA is a risk assessment method developed primarily by Rajbhandari and Snekkenes [7]. CIRA frames risk regarding conflicting incentives between stakeholders, such as information asymmetry situations and moral hazard situations. It focuses on the stakeholders, their actions and perceived outcomes of these actions.
- CORAS is a UML (*Unified Modeling Language*) model-based security risk analysis method developed for InfoSec. CORAS defines a UML-language for security concepts such as threat, asset, vulnerability, and scenario, which is applied to model incidents.
- The CCTA Risk Analysis and Management Method (*CRAMM v.5*) is a qualitative *ISRA* method. *CRAMM* is specifically built around the supporting tool with the same name and refers to descriptions provided in the repositories and databases present in the tool.
- FAIR (*Factor Analysis of Information Risks*) is a risk assessment method and one of the few primarily quantitative *ISRA* approaches. FAIR breaks risks down into twelve specific factors, which contains four well-defined factors for the loss and probability calculations. FAIR includes ways to measure the factors and to derive quantitative analysis results.
- The Norwegian National Security Authority Risk and Vulnerability Assessment (*NSM ROS*) [8] approach was designed for aiding organizations in their effort to become compliant with the Norwegian Security Act.
- OCTAVE (*Operationally Critical Threat, Asset, and Vulnerability Evaluation*) Allegro methodology is the most recent method of the *OCTAVE*-family, aimed at being less extensive than the previous installments of *OCTAVE*. It is a lightweight version of the original *OCTAVE* and was designed as a streamlined process to facilitate risk assessments without the need for InfoSec experts and still produce robust results.
- ISO/IEC 27005:2011 - Information technology, Security techniques, Information Security Risk Management details the complete process of *ISRM/RA*, with activities with each task. Centers on assets, threats, controls, vulnerabilities, consequences and likelihood.
- The current installment of the NIST SP 800-30 - Guide for Conducting Risk Assessments is at revision one, and was developed to further statutory responsibilities under the Federal Information Security Management Act. NIST SP 800-30 rev. one was designed for larger and complex organizations. The purpose of the publication was to produce a unified information security framework for the U.S. federal government, and the framework shows signs of being created to manage complexity.
- The ISACA (*Information Systems Audit and Control Association*) Risk IT Framework and Practitioner Guide is an *ISRM/RA* approach where the Practitioner Guide complements the Risk IT Framework. The former provides examples of how the concepts from the framework can be realized. It is an established approach developed by *ISACA*, based on ValIT and CobIT, and, therefore, has a business view on risks, defining several risk factors.
- Privacy impact assessments are methods that are supposed to address risks to privacy in a system or a project. The Norwegian Data Protection Authority's (*Datatilsynet*) Risk Assessment of Information Systems (*RAIS*) are *ISRA* guidelines that primarily are designed for aiding data handlers in their effort to become compliant with the Norwegian Data Protection and Privacy Act and corresponding regulations.
- Outsourcing services to the cloud brings new risks to the organization. Microsoft's Cloud Risk Decision Framework is a method for risk assessing cloud environments [9].

7. Conclusions

Detecting and analyzing cyber incidents is a task that requires significant resources and a wide range of data from a variety of sources. Analysts and cybersecurity professionals need to quickly collect, process, and analyze information to effectively detect and respond to cyber threats. Determining the categories of cyber incidents and their criticality levels is a complex process that also requires significant resources. This is an important component of properly classifying and prioritizing incidents to ensure a fast and effective response to the most critical events. Improving the cyber incident response process is driven by a large number of developed information security risk assessment methods. These methods allow organizations to effectively identify, assess and manage risks, as well as improve their security strategies. The application of these methods contributes to a more accurate and systematic approach to cybersecurity management and ensures reliable protection of information assets.

References

[1] ASD's ACSC - Guidelines for Cyber Security Incidents. Access mode: http://surl.li/pslnn
[2] ENISA, EUROPOL - Common Taxonomy for Law Enforcement and The National Network of CSIRTs - Access mode: https://www.europol.europa.eu/sites/default/files/documents/common_taxonomy_for_law_enforcement_and_csirts_v1.3.pdf
[3] CERT-UA - List of categories of cyber incidents. Access mode: https://cert.gov.ua/recommendation/16904
[4] ISO/IEC 27002:2022 Information security, cybersecurity and privacy protection – Information security controls. Access mode: http://www.itref.ir/uploads/editor/d3d149.pdf
[5] NIST Special Publication 800-61 rev.2 Computer Security Incident Handling Guide. Access mode: https://csrc.nist.gov/pubs/sp/800/61/r2/final, DOI: 10.6028/NIST.SP.800-61r2
[6] Resolution of the Cabinet of Ministers of Ukraine dated 04.04.2023 No. 299, Some issues of response by cybersecurity entities to various types of events in cyberspace. Access mode: https://zakon.rada.gov.ua/laws/show/299-2023-п
[7] Einar Snekkenes. Position paper: Privacy risk analysis is about understanding conflicting incentives. In Simone Fischer-Haubner, Elisabeth Leeuw, and Chris Mitchell, editors, *Policies and Research in Identity Management, volume 396 of IFIP Advances in Information and Communication Technology*, pages 100–103. Springer Berlin Heidelberg, 2013. 113DOI 10.1007/978-3-642-37282-7
[8] NSM. Veiledningirisiko- ogsï£¡rbarhetsanalyse (guidelines for risk and vulnerability assessments). Technical report, Nasjonal-Sikkerhetsmyndighet (Norwegian National Security Authority), 2006. 12, 32, 33, 43, 113, 119, 128, 131, 133, 135
[9] Doctoral theses at NTNU, 2017:153. Gaute Bjørklund Wangen. Cyber Security Risk Assessment Practices. Core Unified Risk Framework, pages 111-131. Access mode: http://surl.li/pslmiISBN 978-82-326-2378-5

Надійшла до редакції 25 жовтня 2023 р. Переглянута 30 листопада 2023 р. Прийнята 22 грудня 2023 р.

Автори:
Копиця Олександр, аспірант кафедри безпеки інформаційних систем і технологій, Харківський національний університет (ХНУ) імені В. Н. Каразіна, Харків, Україна.
E-mail: oleksandr.kopytsia@student.karazin.ua

Узлов Дмитро, к.т.н., доцент, в.о. декана факультету комп'ютерних наук, ХНУ ім. В. Н.Каразіна, Харків, Україна.
E-mail: dmytro.uzlov@karazin.ua
ORCID: https://orcid.org/0000-0003-3308-424X

Методи визначення категорій кіберінцидентів та оцінки ризиків інформаційної безпеки.

Анотація. Стаття присвячена вивченню категорій кіберінцидентів та їх пріоритезації в контексті інформаційної безпеки. Розглядаються основні джерела, що надають інформацію про кіберзагрозий визначається їх роль у виявленні та аналізі інцидентів, наводяться інструменти для збору, та аналізу даних. Розглядаються поняття події, інциденту і злочину та співвідношення між ними. Наводиться класифікація різноманітних типів кіберзагроз, спосіб їх систематизації, характеристики та вплив на інформаційні системи. Представлені приклади використання класифікації кіберінцидентів. Автори розглядають, також, специфічні види кіберінцидентів, що можуть виникнути в різних сферах діяльності та небезпеки для інформаційних систем які вони становлять. Обґрунтовується необхідність та методи визначення пріоритетів у реагуванні на кіберзагрози, що дозволяє ефективно розподіляти ресурси та здійснювати попереджувальні заходи з кібербезпеки. Розкривається підхід до оцінки та класифікації інцидентів за їх можливим впливом на діяльність організації, захист інформації та здатність відновлюватися після кібератак. Висвітлюються різноманітні підходи та методології для визначення та управління ризиками в сфері інформаційної безпеки, що включають в себе використання стандартів, моделей та інструментів оцінки. Матеріали статті є додатковим ресурсом відомостей для фахівців з кібербезпеки, дослідників та керівників, які цікавляться питаннями управлінням ризиками та захистом інформаційних активів у сучасному цифровому середовищі.

Ключові слова: *кібербезпека, кіберінцидент, IDS, категорії кіберінцидентів, пріоритизація інцидентів, ризики безпеки.*

DOI : 10.26565/2519-2310-2023-2-05
УДК 004.056.5

ДОСЛІДЖЕННЯ МОЖЛИВОСТЕЙ ЗАСТОСУВАННЯ СТЕГАНОГРАФІЧНИХ ТА КРИПТОГРАФІЧНИХ АЛГОРИТМІВ ДЛЯ ПРИХОВУВАННЯ ІНФОРМАЦІЇ

Микита Бодня[1], Марина Єсіна[1,2], Володимир Пономар[1,2]

[1]Харківський національний університет імені В. Н. Каразіна, майдан Свободи, 4, Харків, 61022, Україна
bodnia2020kb12@student.karazin.ua, m.v.yesina@karazin.ua ORCID: https://orcid.org/0000-0002-1252-7606
[2]АТ «ІІТ», вулиця Коломенська, 15, Харків, 61166, Україна
Laedaa@gmail.com ORCID: https://orcid.org/0000-0001-5271-2251

Надійшла до редакції 17 листопада 2023 р. Переглянута 18 грудня 2023 р. Прийнята 25 грудня 2023 р.

Анотація: Організація захисту інформації завжди було актуальною задачею особливо після появи інформаційно-комунікаційних систем. Базисними направленнями в області захисту інформації, які прийшли зі стародавніх часів є криптографія та стеганографія. Криптографія реалізує захист інформації шляхом перетворення інформації у нечитабельний вигляд. Стеганографія дозволяє приховати інформацію в різних контейнерах, при цьому факт наявності інформації залишається непоміченим для випадкових спостерігачів. У статті розглядаються підходи до криптографії та стеганографії, концепція гібридного застосування криптографічних та стеганографічних методів для забезпечення подвійного рівня захисту даних, загальна архітектура криптографічних та стеганографічних систем. Традиційними криптографічними системами, які застосовуються в сучасних системах захисту інформації є симетричні та асиметричні криптосистеми. Хоча симетричні системи еволюціонували з появою нових математичних перетворень, але вони мають суттєвий недолік. Він полягає в потребі додаткової передачі секретного ключа отримувачу. Така стратегія вимагає використання захищеного каналу зв'язку, оснащеного технічними системами захисту. При цьому існує ризик несанкціонованого доступу, який може спричинити компрометацію секретного ключа. Виходячи з вищенаведених проблем симетричних криптосистем, при розробці механізмів захисту, перевагу віддають асиметричним алгоритмам. Проведено аналіз криптосистеми RSA, яка ґрунтується на асиметричному підході шифрування. Ця система використовується в сучасних протоколах автентифікації та забезпечення конфіденційності в інформаційних системах та Інтернеті. Проведено дослідження швидкодії програмних модулів генерації ключової пари, шифрування та розшифрування для системи RSA, шляхом зміни загальних параметрів алгоритму (модуля перетворень, розміру вихідних даних). Результати часових вимірювань наведені в таблиці, на базі яких побудовані залежності часу від модифікації конкретних параметрів. Досліджено стеганографічний алгоритм модифікації найменш значущого біту (НЗБ), який застосовується для приховування даних в зображеннях. Нині існує широкий спектр стеганоалгоритмів, які розробляються на базі особливостей сенсорних систем людини (системи зору або слуху). Розглядаються властивості зорової системи людини, які використовуються в стеганографії.

Ключові слова: *криптографія, стеганографія, ключ, інформаційне повідомлення, асиметрична криптосистема, симетрична криптосистема, криптограма, стеганограма..*

1.Вступ

Інформація завжди займала провідне місце в житті людини. Поняття «інформація» [1] можна інтерпретувати як сукупність публічно оголошених або документованих відомостей, які охоплюють явища природи, навколишнього середовища та різноманітні області діяльності соціуму й держави. Вагомість і класифікація інформації визначається її вмістом. Поява інформаційно-комунікаційних систем і глобальних мереж спрощує доступність й обмін інформацією. Стрімкий технологічний прогрес призвів до появи загроз несанкціонованого доступу, порушення конфіденційності, цілісності інформації, фальсифікації даних тощо. Поряд з цим питання забезпечення інформаційної безпеки (ІБ) завжди було актуальним, починаючи зі стародавніх часів і до теперішнього моменту. Основними напрямами, що впроваджують надійні механізми забезпечення ІБ є криптографія і стеганографія [2].

Для розв'язання проблем ІБ широко використовуються відповідні алгоритми криптографії і стеганографії. Сучасні системи ІБ розробляються з реалізацією перспективних криптографічних і стеганографічних методів захисту. Система інформаційної безпеки (СІБ) [1]

призначена для забезпечення безпеки інформації, яка циркулює у інформаційно-телекомунікаційній системі (ІТС) від неавторизованих сторін. Сучасні СІБ оснащені відповідними апаратними модулями безпеки, котрі спрямовані на протидію фізичним загрозам. Ці модулі містять інтегровані мікропроцесори, що здатні виконувати потрібні математичні обчислення для реалізації відповідних криптографічних та стеганоалгоритмів.

Криптографія – наука про методи захисту інформації від несанкціонованого доступу чи модифікації. Метою криптографії є реалізація захисту інформації шляхом спеціального її перетворення (шифрування). Загальною ідеєю криптографії є конвертування вмісту даних в нерозбірливий вигляд. Повернення зашифрованого тексту у вихідний стан здійснюється за допомогою спеціального ключа, яким володіє лише власник інформації або довірена сторона. Зловмисник гіпотетично може перехопити шифртекст в момент передачі по каналу зв'язку (КЗ), але не матиме можливість ознайомитися зі вмістом вихідного повідомлення, оскільки у нього не має секретного ключа необхідного для виконання процедури дешифрування (криптоаналіз). Криптографія забезпечує конфіденційність, цілісність та автентичність інформації, використовуючи математичні методи та алгоритми.

Стеганографія – наука про методи і способи зберігання та передачі інформації де сам факт передачі чи зберігання корисної – прихованої інформації, залишається в таємниці. Приховування інформаційних даних здійснюється в так звані контейнери (*зображення, аудіо файли, файлові системи тощо*). При вбудовуванні прихованих даних, різні стеганографічні методи використовують різні властивості природних сенсорних систем людини (насамперед зорових та звукових).Для вбудовування корисних повідомлень в стеганографії використовуються надмірності, якими характеризуються контейнери– переносники даних. Ці надмірності можуть бути природними чи штучними, в залежності від структури контейнерів. Наприклад, у кластерних файлових системах надмірність реалізується штучно, використовуючи для цього характеристики і структуру файлової системи[3-4].

З розвитком систем, які володіють великими обчислювальними потужностями почався стрімкий розвиток комп'ютерної криптографії та стеганографії. Сучасні обчислювальні системи здатні оперативно обробляти та перетворювати великі масиви даних, що в свою чергу, спонукає створення нових стеганографічних методів, які ускладнюють процес детектування повідомлень, а криптографічні ключі генеруються таким чином, щоб виключити ймовірність їх вгадування. Таким чином, проявляється тенденція комплексування застосування криптографічних і стеганографічних методів захисту інформації задля підвищення загального рівня безпеки [5]. Симбіоз криптографії і стеганографії є критично необхідним при обміні чутливої інформації між абонентами сучасних ІТС на фоні постійного ускладнення спектру загроз безпеки та зростання можливостей апаратного оснащення [6].

Криптографічні методи широко використовуються для побудови систем автентифікації, шифрування даних для захисту конфіденційної інформації в мережах, підтвердження цілісності та автентичності даних. Криптографія застосовується в банківському секторі для забезпечення безпеки персональних даних клієнтів та інформації щодо банківських операцій. Поряд з цим, стеганографія використовується для захисту авторських прав, забезпечення безпеки конфіденційної інформації при її передачі через мережі, публікації анонімних матеріалів або звітів, проведення потайної комунікації в умовах, при яких неможливо застосувати класичні (*з різних причин*) криптографічні методи тощо.

Метою цієї статі є аналіз структурних схем стеганографічноїі криптографічної систем захисту інформації та дослідження можливостей застосування відразу обох векторів захисту у їх комбінації.

2. Структурна схема стеганографічної системи

Узагальнена структурна схема стеганографічної системи представлена на рис.1, де вона розглядається, як специфічна реалізація системи зв'язку [2].

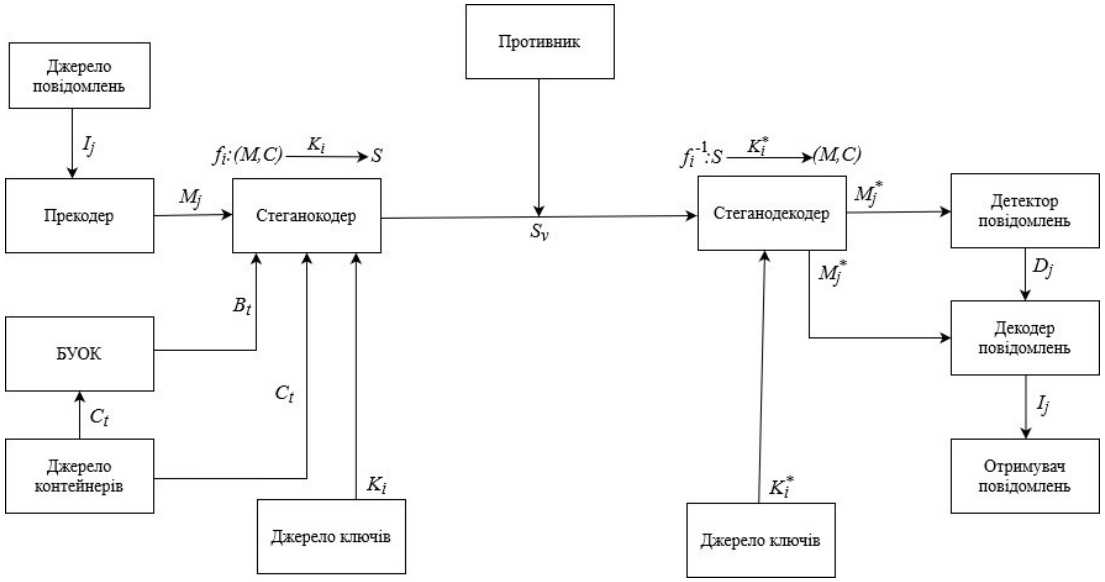


Рис. 1 –Узагальнена структурна схема стеганографічної системи
Fig. 1 - Generalized structural scheme of the steganographic system

В рамках даної схеми, джерело повідомлень генерує масив інформаційних повідомлень, яке представлено множиною $I=\{I_1,I_2,\ldots,I_n\}$. Повідомлення I_j є одним з повідомлень множини I, яке перетворюється прекодером. Результатом перетворення є сформоване повідомлення $M_j \in M$,де M_j –потайне інформаційне повідомлення, яке необхідно приховати в контейнері, $M=\{M_1,M_2,\ldots,M_n\}$– множина можливих секретних повідомлень [2]. Процес генерації інформаційних повідомлень джерелом повідомлень можна уявити випадковим процесом. Розподіл ймовірностей випадкового процесу визначається сукупним розподілом ймовірностей випадкових величин в рамках даного процесу. Тоді можна представити випадковий процес у вигляді множини $P_M=\{P(M_1),P(M_2),\ldots,P(M_n)\}$,складові якої є ймовірностями випадкових величин випадкового процесу. Джерело контейнерів формує спектр пустих контейнерів, який представлений множиною $C=\{C_1,C_2,\ldots,C_l\}$. Результатом роботи джерела контейнерів є випадковий контейнер C_t, який входить до складу множини C. Саме функціонування пристрою генерування контейнерів може бути охарактеризоване, як випадковий процес. Оскільки поява будь-якого контейнера з множини C є випадковою, кожному елементу множини C може бути присвоєно відповідні ймовірності. Випадковий процес генерування контейнерів може бути описаний множиною ймовірностей $P_C=\{P(C_1),P(C_2),\ldots,P(C_l)\}$, елементи якої є розподілені ймовірності між випадковими величинами цього процесу. Після створення контейнера, блок урахування особливостей контейнера (БУОК), аналізує контейнер C_t для виділення особливостей, які будуть враховуватися при вбудовуванні приховуваного інформаційного повідомлення M_j. Контейнер $C_t \in C$, з визначеними БУОК властивостями B_t, поступає на стеганокодер, де здійснюються спеціальні операції з вбудовування (*або інкапсуляції*) стеганографічних даних (*контенту*). Результатом здійснення інкапсуляції секретних повідомлень до

контейнерів є стеганограми (*тобто, заповнені контейнери)*, де $S=\{S_1,S_2,\ldots,S_m\}$– множина утворених стеганограм.

Тривіальне подання стеганографічного вбудовування інформації [7] можна подати у вигляді множини відображень $f:\{f_1,f_2,\ldots,f_k\}$, де $f_i:(M,C)\to S$, $i=1,2,\ldots,k$.В аналогії з виразом інкапсуляції контенту, можна відобразити процедуру вилучення інформаційних даних у вигляді множини обернених відображень $f^{-1}:\{f_1^{-1},f_2^{-1},\ldots,f_k^{-1}\}$, де $f_i^{-1}:S\to(M,C)$, $i=1,2,\ldots,k$. У відображенні $f_i\in f$ кожному елементу множини S ставиться у відповідність елемент множин «*M*» та«*C*».

У стеганосистемах для здійснення процесів вставки (інкапсуляції) та вилучення контенту використовуються відповідні секретні ключі. Такий підхід застосовується для підвищення стійкості до детектування повідомлень зловмисником, забезпечення стійкості стеганографічного алгоритму проти можливих атак та зниження ймовірності несанкціонованого вилучення повідомлення зловмисником. Ці ключі породжуються джерелом ключів, звідки вони надходять до стеганокодеру. Управління стеганокодером здійснюється за допомогою секретних ключів. Тож визначимо множину ключів $K=\{K_1,K_2,\ldots,K_k\}$ таким чином, що кожне відображення $f_i\in f$ задається секретним ключем K_i, де $i=1,2,\ldots,k$:

$$f_i:(M,C)\xrightarrow{K_i}S. \tag{1}$$

Кожному відображенню f_i відповідає метод вбудування інформаційного повідомлення $M_i\in M$ в контейнер $C_i\in C$ за допомогою секретного ключа K_i. Аналогічним чином визначимо множину секретних ключів $K^*=\{K_1^*,K_2^*,\ldots,K_k^*\}$ для обернених відображень $f_i^{-1}\in f^{-1}$, які позначають процедуру вилучення інформаційних даних з контейнеру:

$$f^{-1}{}_i:S\xrightarrow{K_i^*}(M,C). \tag{2}$$

Кожному оберненому відображенню $f_i^{-1}\in f^{-1}$ відповідає спосіб вилучення інформаційних даних з контейнера за допомогою секретного ключа K_i^*. Важливо підкреслити, що в основному в стеганографічних перетвореннях використовується один ключ $(K_i=K_i^*)$ для забезпечення узгодженості між процесами вбудовування та вилучення даних. Випадковий процес генерування секретних ключів можна подати у вигляді множини ймовірностей:

$$\begin{aligned} P_K&=\{P(K_1),P(K_2),\ldots,P(K_k)\},\\ P_{K^*}&=\{P(K_1^*),P(K_2^*),\ldots,P(K_k^*)\}. \end{aligned} \tag{3}$$

У (3) кожному ключу $K_i\in K=\{K_1,K_2,\ldots,K_k\}$ відповідає певна ймовірність $P(K_i)$, а ключу $K_i^*\in K^*=\{K_1^*,K_2^*,...,K_k^*\}$ відповідає ймовірність $P(K_i^*)$. Кожному відображенню $f_i\in f$ відповідає секретний ключ K_i. Формування стеганограми (заповненого контейнера) здійснюється за допомогою відображення f_i, яке однозначно задається ключем K_i за повідомленням M_j та контейнером C_t з урахуванням особливостей даного контейнеру B_t. Сформована стеганограма задається наступним співвідношенням:

$$\begin{gathered} S_v=f_i(K_i,M_j,C_t),\\ j\in[1,2,\ldots,n],\ t\in[1,2,\ldots,l],\ i\in[1,2,\ldots,k],\ v\in[1,2,\ldots,m],\ m\ge n \end{gathered} \tag{4}$$

Створена стенограма S_v передається каналом зв'язку на приймальну сторону, під час передачі вона може бути перехоплена противником. Після отримання стеганограми отримувачем, стеганодекодер реалізує зворотне відображення $f_i^{-1} \in f^{-1}$ з множини стеганограм S до множин повідомлень «*M*» і порожніх контейнерів «*C*» під управлінням ключа K_i^*:

$$\left(M_j, C_t\right) = f_i^{-1}\left(K_i^*, S_v\right). \quad (5)$$

Слід підкреслити, що при передачі стеганограми через мережу під впливом завад або противника можливе спотворення заповненого контейнера. На приймальній стороні маємо поєднання стеганограми і результатів «впливу» на неї в процесі передачі по КЗ. Отриману комбінацію можна подати у вигляді $S_v + \partial$, де ∂ – величина, що визначає степінь спотворення стеганограми під випливом зовнішніх факторів. В результаті виконання процедури вилучення стеганодекодером, отримаємо певну оцінку можливого інформаційного повідомлення та порожньому контейнеру:

$$\left(M^*{}_j, C^*{}_t\right) = f_i^{-1}\left(K_i^*, S_v + \partial\right). \quad (6)$$

Для робастних стеганографічних систем[2] незначне спотворення стеганограми $\left(\partial \neq 0\right)$ не призведе до повного руйнування вбудованого повідомлення M_j, в ідеальному випадку оцінка повідомлення M_j^* співпадатиме з вихідним повідомленням M_j. Тому для робастних стеганосистем справедливе наступне співвідношення:

$$\left(M_j, C_t\right) = f_i^{-1}(K_i^*, S_v + \partial). \quad (7)$$

Крихкі стеганографічні системи [2,7]нестійкі до впливу на заповнений контейнер, тому будь-яке спотворення стеганограми $\left(\partial \neq 0\right)$ призводить до руйнування вбудованого повідомлення $\left(M_j^* \neq M_j\right)$, тобто для крихких систем виконується наступна нерівність:

$$\left(M_j, C_t\right) \neq f_i^{-1}(K_i^*, S_v + \partial). \quad (8)$$

На базі отриманої оцінки M_j^* спеціальна функція детектування «приймає рішення» про наявність чи відсутність прихованого повідомлення в переданому контейнері S_v. Завадостійкий декодер використовує рішення апарату детектування повідомленьD_j для винесення бінарного рішення (так/ні) про присутність чи відсутність невиправної помилки в отриманому повідомленні. Операція декодування здійснюється в декодері, де базисними функціями пристрою декодування є зіставлення вилученої оцінки з одним із можливих повідомлень M_jй перетворення їх у вихідний формат повідомленняI_j, що надається отримувачу.

3. Структурна схема криптографічної системи

Криптографічна система – комплекс взаємопов'язаних криптографічних алгоритмів, засобів захисту інформації, нормативної, експлуатаційної документації, необхідних для реалізації захищеності інформації, що зберігається, обробляється або передається [8]. Методологія криптографічного захисту інформації повинна забезпечувати високий (заданий) рівень захисту даних при передачі або зберіганні в інформаційному просторі. Узагальнену структурну схему криптографічної системи проілюстровано на рис. 2, де спектр повідомлень, представлено множиною *«M»*, яка формується джерелом повідомлень.

Інформаційне повідомлення M_i є одним з можливих повідомлень множини M. Кожному інформаційному повідомленню $M_i \in M = \{M_1, M_2, \ldots M_n\}$ відповідає певна ймовірність $P(M_i)$, оскільки кожне повідомлення є реалізацією випадкового процесу.

Розподіл ймовірностей випадкового процесу можна подати у вигляді множини ймовірностей $P(M)=\{P(M_1),P(M_2),\ldots,P(M_n)\}$.

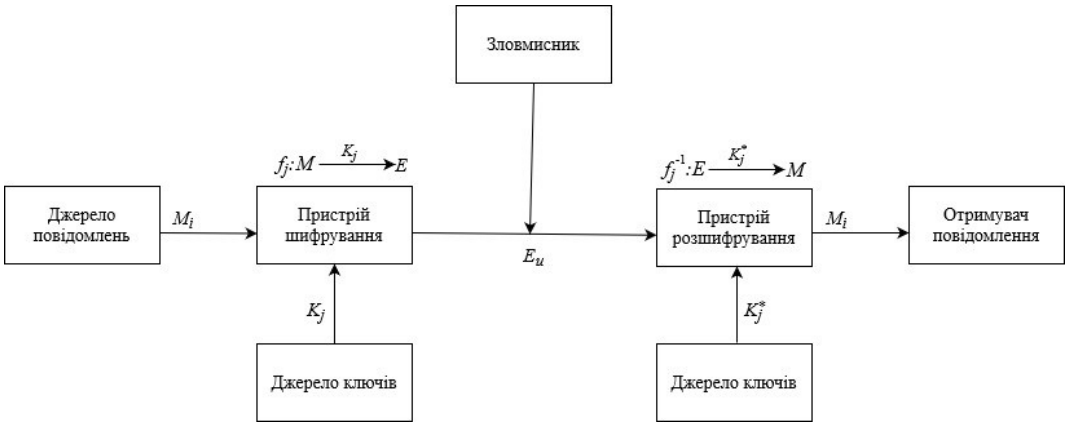


Рис. 2 –Узагальнена структурна схема криптографічної системи

Fig. 2–A generalized structural diagram of a cryptographic system

Множина $E=\{E_1,E_2,\ldots E_v\}$ позначає криптограми шифрованих повідомлень. Криптограма E_u представляє собою шифртекст вихідного повідомлення M_i. Процедуру шифрування здійснює пристрій шифрування, на вхід якого надходить повідомлення M_i. Процес шифрування можна представити у вигляді відображення $f_j \in f$ множини вихідних повідомлень M, у множину криптограм E. Оскільки відображення $f_j \in f$ сюр'єктивне та ін'єктивне (рис. 3), а множини M та E рівнопотужні, то існує обернене відображення $f_j^{-1} \in f^{-1}$, яке позначає процедуру розшифрування повідомлення [8].

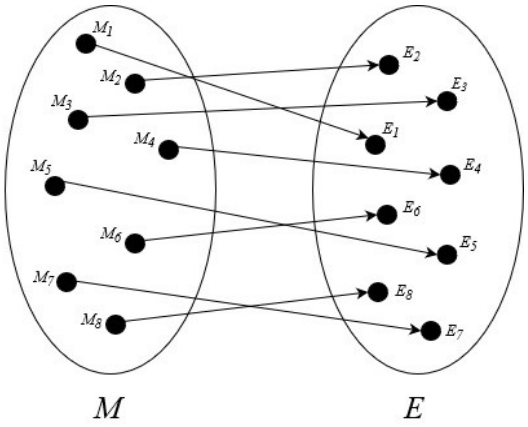


Рис. 3 –Сюр'єктивність та ін'єктивність відображення f_i

Fig. 3–Surjectivity and injectivity of reflection f_i

Джерело ключів створює потік ключів $K=\{K_1,K_2,\ldots,K_k\}$ чи $K^*=\{K_1^*,K_2^*,\ldots,K_k^*\}$, в загальному випадку $K_j \neq K_j^*$. При цьому, якщо $K_j = K_j^*$, то система симетрична, і навпаки, якщо $K_j \neq K_j^*$ –асиметрична [8]. Оскільки породження ключів джерелом ключів є випадковим процесом, то кожному ключу $K_j \in K$ можна присвоїти певну ймовірність $P(K_j)$, а ключам $K_j^* \in K^*$ – ймовірність $P(K_j^*)$. Даний випадковий процес можна представити у вигляді розподілу ймовірностей $P(K)=\{P(K_1),P(K_2),\ldots,P(K_k)\}$ для ключів $K_j \in K$ і $P(K^*)=\{P(K_1^*),P(K_2^*),\ldots,P(K_k^*)\}$, та ключів

$K_j^* \in K^*$. Управління пристроєм шифрування здійснюється за допомогою ключа K_j, а пристроєм розшифрування – ключем K_j^*. Для всіх $j = 1,2,\ldots,k$ відображення $f_j \in f$ задається ключем K_j:

$$f_j : M \xrightarrow{K_j} E . \tag{9}$$

Кожне відображення $f_j \in f$ визначає спосіб шифрування повідомлення $M_i \in M$ ключем K_j (*рис. 4*). Відповідно, ключем K_j^* задається обернене відображення $f_j^{-1} \in f$, яке позначає спосіб розшифрування повідомлення за допомогою ключа K_j^* (*див. рис. 5*):

$$f_j^{-1} : E \xrightarrow{K_j^*} M . \tag{10}$$

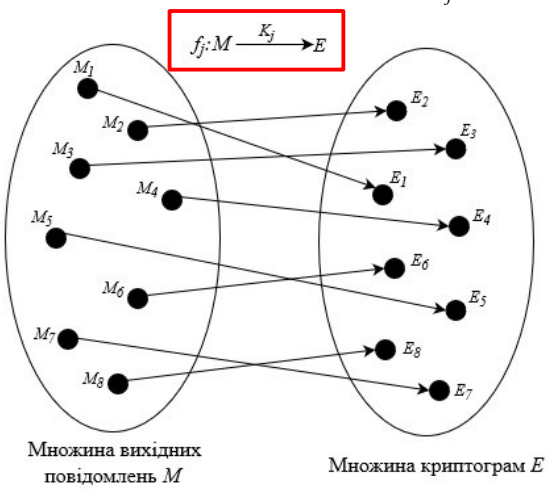


Рис. 4 –Відображення (9) множини вихідних повідомлень в множину криптограм

Fig. 4 - Mapping (9) of a set of output messages into a set of cryptograms

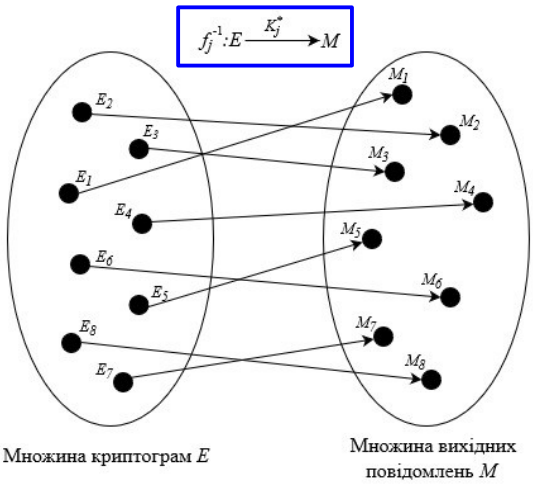


Рис. 5 –Обернене відображення (10) множини криптограм в множину вихідних повідомлень

Fig. 5 - Inverse mapping (10) of a set of cryptograms into a set/multiple of outgoing messages

Ключ K_j дозволяє зашифрувати один елемент з множини M, навпаки ключем K_j^* можливо отримати лише один елемент з криптограми E_u. Криптограма E_u формується за допомогою відображення $f_j \in f$, яка співвідноситься з ключем K_j за повідомленням M_i:

$$E_u = f_j\left(K_j, M_i\right). \tag{11}$$

Сформована криптограма E_u передається каналом зв'язку на приймаючу сторону. В момент передачі шифртекст E_u може бути перехоплений зловмисником. Пристрій розшифрування здійснює перетворення криптограми E_u у вихідне повідомлення M_i. Відновлення вихідного повідомлення здійснюється за допомогою оберненого відображення f_j^{-1}, яке пов'язане з ключем K_j^*:

$$M_i = f_j^{-1}\left(K_j^*, E_u\right). \tag{12}$$

Вилучене з криптограми повідомлення M_i надходить отримувачу.

4. Сутність традиційних криптосистем для реалізації захищеності інформації

Шифрування– це процедура направлена на забезпечення захисту інформації шляхом перетворення її у нечитабельний вигляд. Доступ до вмісту конфіденційних даних можливо отримати лише після виконання процедури розшифрування за допомогою секретного ключа. Цим

ключем володіє лише власник ключа чи довірена сторона. Секретні ключі повинні зберігатися в секреті, в захищеному середовищі, оскільки компрометація ключа призведе до несанкціонованого доступу до вмісту конфіденційних даних. Алгоритми шифрування використовують різноманітні математичні операції: - арифметику в полях Галуа $GF(q)$, математичні операції в групах точок еліптичних кривих, в групах простих чисел, модульну арифметику, перетворення Фур'є тощо. Управління процесами шифрування й розшифрування здійснюється за допомогою секретного ключа, тому криптосистеми класифікуються за способом використання ключів на дві групи: - симетричні і асиметричні. В асиметричних криптосистемах (рис. 6) генеруються 2 ключі: - відкритий та секретний. Відкритий ключ використовується для реалізації процедури шифрування, а секретний ключ застосовується для виконання розшифрування повідомлення. В симетричних криптосистемах (рис. 7) для процедур шифрування та розшифрування використовується лише один секретний ключ [9].

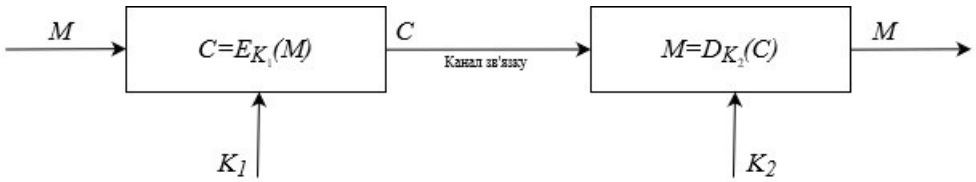


Рис. 6 – Спрощена модель асиметричної криптосистеми
Fig. 6 – A simplified model of an asymmetric cryptosystem

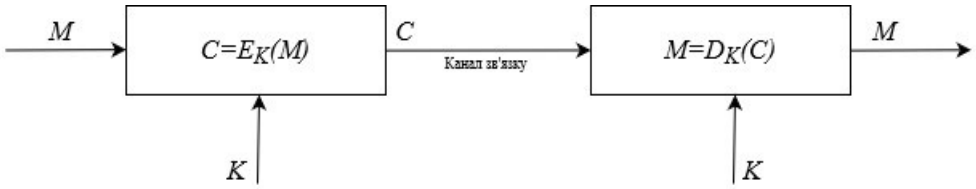


Рис. 7 – Спрощена модель симетричної криптосистеми
Fig. 7 – A simplified model of a symmetric cryptosystem

Сучасні криптографічні протоколи базуються на криптографії з відкритим ключем. В електронних комунікаціях для шифрування інформації використовуються *асиметричні* криптосистеми при передачі інформації по відкритих КЗ. В *симетричних* криптосистемах присутня проблема розподілу ключів [8], незалежно від структури криптографічного алгоритму. Перед початком сеансу обміну даними між двома сторонами, одна з них повинна згенерувати секретний ключ та передати іншої. При цьому для передачі ключа потрібно використовувати захищений КЗ, при цьому існує ризик компрометації секретного ключа, оскільки немає гарантії, що зловмисник не зможе обійти системи захисту. У табл.1 наведено спрощений опис характеристик симетричних та асиметричних криптосистем.

Розглянемо алгоритм RSA, який належить до криптографії з відкритим ключем. Шифр RSA, названий на честь його винахідників Ріверса (RonRivers), Шаміра (AdiShamir) і Адлемана (LeonardAdleman) [9-11]. Криптосистема RSA базується на застосуванні односторонньої функції [9] утворення добутку двох великих чисел, що є простішою задачею порівняно з розкладанням великого числа на прості множники [11]. Безпека криптосистеми RSA ґрунтується на факторизації великих чисел. Основною ідеєю алгоритму є генерування простих чисел для обчислення їх добутку, що визначає модуль n, який буде використовуватися в процедурах шифрування та розшифрування. Метою криптоаналізу в парадигмі RSA є знаходження секретного ключа d ключової пари (d,e), де e – відкритий ключ.

Таблиця 1 – Стислий опис класичних криптосистем

Table 1 – Brief description of classical cryptosystems

Тип системи	Характеристика
Асиметрична	*Для процедури шифрування і розшифрування використовуються 2 різні ключі. Асиметричні алгоритми потребують значно більше обчислювальних ресурсів порівняно із симетричними. Алгоритми асиметричної криптографії програють симетричним за швидкодією. Ключовою перевагою асиметричних криптосистем є використання ключів великої довжини (512 – 4096 біт), що позначається на швидкодії алгоритму. Асиметрична криптографія використовується у протоколах SSL (Secure Sockets Layer) та TSL (Transport Layer Security) для забезпечення безпеки обміну даними в мережі Інтернет.*
Симетрична	*Шифрування і розшифрування реалізується за допомогою 1 ключа, попередньо узгодженого між суб'єктами комунікації. Алгоритми симетричної криптографії за швидкодією перевершують асиметричні. Довжина ключа в симетричних системах помітно менша (40 – 256 біт). Область застосування симетричних криптосистем охоплює захист конфіденційної інформації фінансових установ, комерційних компаній та державних установ.*

Алгоритм RSA складається з трьох етапів:

1. Генерація ключової пари.
2. Шифрування інформаційного повідомлення M .
3. Розшифрування зашифрованого повідомлення C.

Генерація загальносистемних параметрів і ключів системи RSA має наступні кроки:

1. Обираються два простих числа p та q, які тримаються в секреті.
2. Обчислюється модуль n, що визначається співвідношенням $n = pq$.
3. Обчислюється функція Ейлера для модуля n, $\varphi(n) = (p-1)(q-1)$.
4. Вибирається таке значення відкритого ключа e, щоб воно було взаємно простим стосовно $\varphi(n)$, а саме $(\varphi(n), e) = 1$.
5. Визначається таке значення секретного ключа d, щоб $de \equiv 1 \bmod \varphi(n), \quad d < \varphi(n)$.

Результат:

1. Загальносистемні параметри p, q, n, $\varphi(n)$.
2. Секретний ключ $\{d, n\}$.
3. Відкритий ключ $\{e, n\}$.

Криптограма в системі RSA обчислюється за наступним правилом:

$$C \equiv M^e \bmod n. \tag{13}$$

Розшифрування зашифрованого повідомлення обчислюється за допомогою формули:

$$M \equiv C^d \bmod n. \tag{14}$$

У табл. 2 наведені результати продуктивності програмного модуля генерування ключової пари криптосистеми RSA, а рис. 8 відображає залежність часу генерування ключів від розміру модуля (*пакет моделювання MATLAB*).

Таблиця 2 – Результати генерування ключової пари для системи RSA

Table 2 – Key pair generation results for RSA system

№	Довжина модуля, біт	Час створення, секунди
1	*512*	*0,025*
2	*768*	*0,045*
3	*1024*	*0,141*
4	*2048*	*1,636*

У табл. 3 наведено результативність швидкодії виконання програмних модулів шифрування й розшифрування для криптографічної системи RSA в залежності від розміру файлу та довжини модуля перетворення.

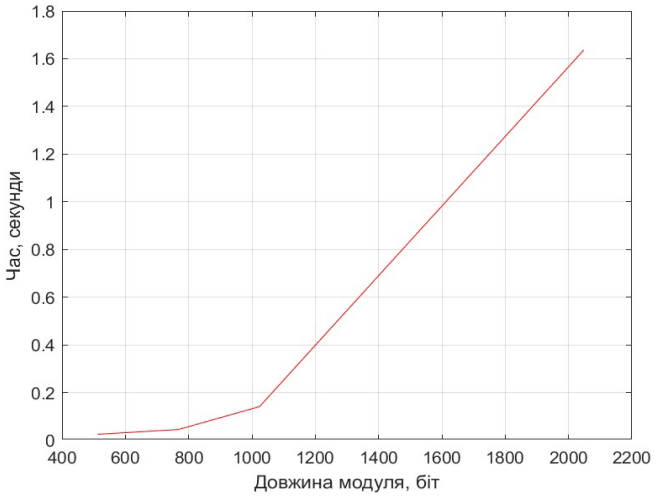


Рис. 8 – Залежність часу генерування ключової пари від розміру модуля
Fig. 8 – Dependence of key pair generation time on module size

Таблиця 3 – Оцінка швидкодії шифрування та розшифрування для системи RSA
Table 3 – Evaluation of encryption and decryption performance for the RSA system

Розмір файлу	Розмір модуля перетворення, біт	Час шифрування, секунди	Час розшифрування, секунди
219 КБ	*512*	*0,068*	*0,875*
	768	*0,089*	*1,520*
	1024	*0,127*	*2,347*
	2048	*0,184*	*7,988*
4,43 МБ	*512*	*1,391*	*18,151*
	768	*1,731*	*31,173*
	1024	*2,497*	*48,524*
	2048	*3,775*	*165,003*
8,65 МБ	*512*	*2,700*	*35,256*
	768	*3,401*	*61,187*
	1024	*4,889*	*94,335*
	2048	*7,388*	*325,451*

На рис. 9 проілюстровано залежність часу виконання програмного модуля шифрування і розшифрування для шифру RSA в залежності від розміру файлу та довжини модуля перетворення.

5. Приховування інформації в просторовій області нерухомих зображень

Приховування криптограм в контейнерах, наприклад, графічних зображеннях, дозволяє підвищити рівень безпеки захисту інформації. Основна ідея стеганографічного захисту полягає в тому, що приховування даних здійснюється таким чином, щоб це не було помітно для не проінформованого спостерігача. Методи приховування даних в зображеннях використовують властивості зорової системи людини (ЗСЛ) та класифікуються на 2 групи: низькорівневі (*фізіологічні*) та високорівневі (*психофізіологічні*) [2,7]. До низькорівневих властивостей слід віднести наступні:

1. Слабка чутливість до незначної зміни яскравості.

2. Слабка чутливість до незначної зміни контрасту.
3. Частотна чутливість.
4. Ефект маскування.
5. Слабка чутливість до незначної зміни яскравості синього кольору.

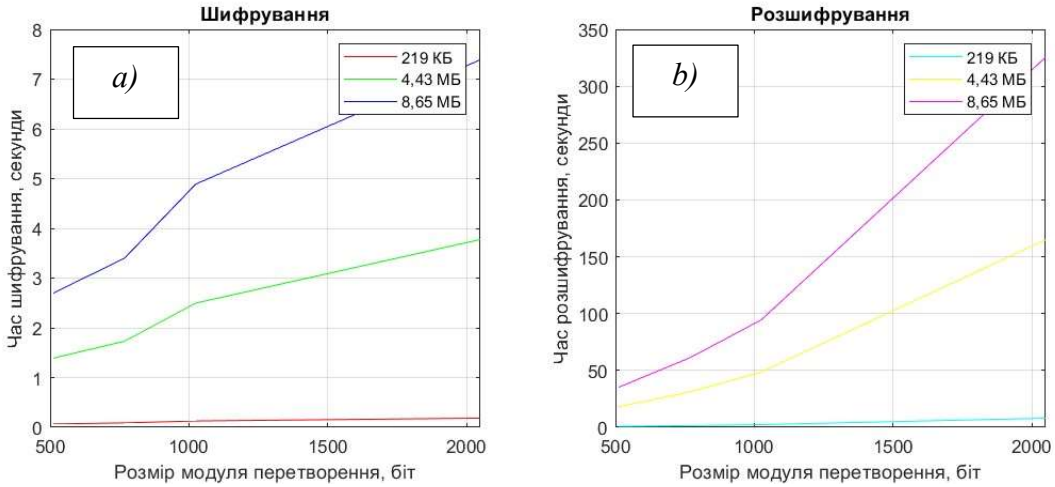


Рис. 9 – Час шифрування (*a*) і розшифрування (*b*) для різних довжини модуля перетвореннята розміру файлу

Fig. 9 – Encryption (*a*) and decryption (*b*) times for different lengths of the transform module & file size

До високорівневих властивостей слід віднести наступні [12]:

1. Чутливість до кольору – деякі кольори привертають більше уваги людини порівняно з іншими кольорами. Ефект помітності підвищується, коли відтінок заднього тла суттєво відрізняється від кольорів об'єктів розташованих на ньому [7].
2. Чутливість до розміщення – передусім, людині властиво розглядати центральну ділянку зображення, а вже потім звертати увагу на його околиці.
3. Чутливість до зовнішніх подразників – рух очей людини залежить від таких факторів як конкретна ситуація або наявність додаткової інформації, інструкцій щодо способу перегляду.
4. Чутливість до контрасту – різкі контрастні області зображення та значні перепади яскравості викликають до себе більше уваги.
5. Чутливість до розміру – великі за розміром області зображення більш помітні порівняно з меншими. При цьому існує поріг насичення, коли подальше збільшення розміру не має істотного значення.
6. Чутливість до форми – у людини значно більше уваги викликають довгі та тонкі об'єкти, у порівнянні з однорідними та округлими.

З урахуванням вказаних властивостей побудовані відомі методи стеганографічної вставки інформації в нерухомі зображення, наприклад такі, як:

1. Метод вбудовування на основі зміни найменш значущих біт (*методи псевдовипадкової перестановки, блокового вбудовування та ін.*.).
2. Метод квантування.
3. Метод Куттера-Джордона-Боссона.
4. Метод вбудовування в частотній області на основі кодування різниць абсолютних значень дискретного косинусного перетворення (метод Коха-Жао).
5. Метод Бенгама–Мемона–Ео–Юнг.
6. Метод прямого розширення спектра.

На практиці, не всі стеганографічні методи приховування інформації в нерухоме зображення гарантують безпомилкове вилучення повідомлення. Для того, щоб правильно розшифрувати криптограму потрібно використовувати методи, які дозволяють вилучити повідомлення без спотворень. Альтернативним рішенням може бути застосування методів завадостійкого кодування (наприклад, коди Хеммінга, БЧХ коди чи коди Ріда-Соломона, які дозволяють виправити можливі помилки. Ці методи дозволяють корегувати помилки, тобто виявляти й вилучати їх, що збільшує надійність вилучення прихованої інформації.

6. Вставка даних в нерухомі зображення на основі модифікації найменш значущого біта

Метод заміни найменш значущого біту (НЗБ, *LSB – Least Significant Bit*) є найпростішим способом стегановставки інформації в нерухоме зображення без видимих спотворень контейнеру. Метод *LSB* ґрунтується на експлуатації 1-їнизькорівневої властивості ЗСЛ [12]. Суть методу полягає в заміні менш значущих бітів пікселів зображення на біти прихованого інформаційного повідомлення. При цьому людина не спроможна виявити ці зміни. Колір кожного пікселю представлений комбінацією трьох кольорових компонентів, т.з. RGB кольорова модель. Рівень інтенсивності кожної з RGB складових (рис. 10) може приймати значення $0\ldots255$ (всього $L_m = 256$ рівнів квантування) та кодується 8 бітами [7,12].

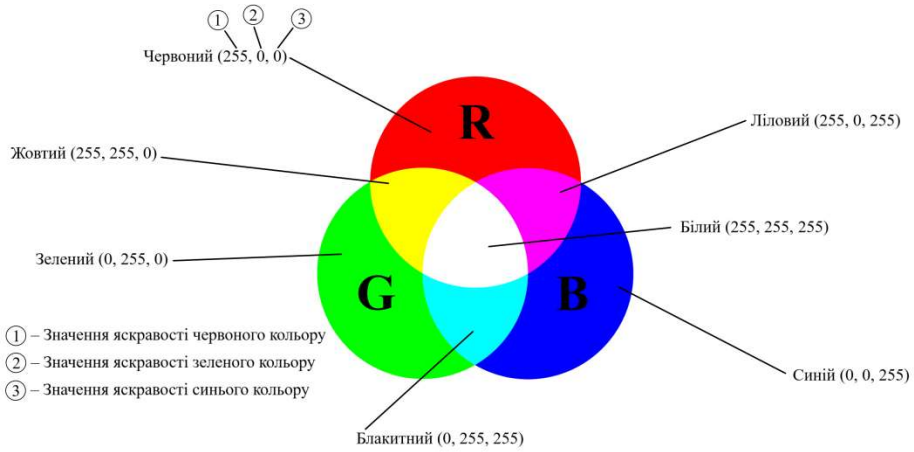


Рис. 10 – Модель RGB
Fig. 10 – RGB model

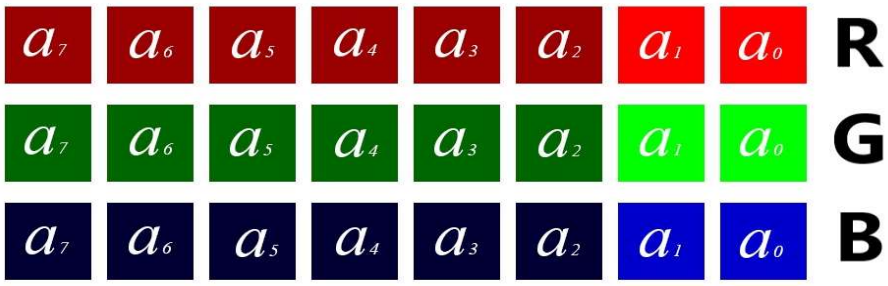


Рис. 11 –Бітове представлення кольорових компонентів
Fig. 11 – Bit representation of color components

Піксель являє собою одну точку зображення, яка містить рівні інтенсивності кожного з 3-х кольорових RGB каналів. Відповідні рівні інтенсивності кодуються 3 байтами (рис. 11), які в сукупності визначають потрібний відтінок кольору для конкретного пікселю. Отже, інформація про колір для певного пікселю представлена 24 бітами (3 байтами).

Відомо, що поріг чутливості ЗСЛ до зміни яскравості складає 2-3% [12]. Це означає, що, якщо модифікація рівня яскравості пікселів знаходиться нижче порога чутливості ЗСЛ, то він не виявить візуального викривлення зображення-контейнеру. При цьому менш «вагомі» (*тобто, молодші*) бітові розряди у цифровому зображенні, мають менший вплив на його візуальні характеристики в порівнянні зі старшими бітами [12], тому ці біти кольорових каналів (або градації сірого (визначається одним каналом)) можуть використовуватись для інкапсуляції бітів прихованого повідомлення. Перспективним вектором розробок є гібридне застосування криптографії та стеганографії: - приховування криптограм в контейнерах. В цьому разі для вбудовування методом НЗБ шифрованої інформації за допомогою криптосистеми RSA застосовується пустий контейнер (рис. 12). Кожен символ зашифрованого повідомлення, прочитаний в кодуванні ASCII, використовуючи команди середовища *MathCAD*, а вставка бітів криптограми зроблено в нульові біти (рис.11) послідовних байтів, масиву растрових даних червоного каналу (R).

В загальному випадку важливо розуміти, що для вставки прихованої інформації можна використовувати растрові дані будь-якого кольорового каналу, при цьому ЗСЛ не помічає спотворень зображення-контейнеру, оскільки його найменш значущий біт асоціюється з «шумом» та, за замовчуванням, не є важливим для візуальної оцінки вихідного зображення. Заповнений контейнер, який містить криптограму представлений на рис.13. З порівняння порожнього та заповненого контейнерів (*рис.12- 13*) можна стверджувати, що помітні ЗСЛ викривлення контейнеру, відсутні.

"1.bmp"

Рис. 12 –Порожне зображення - контейнер
Fig. 12 – An empty image - container

"Stego.bmp"

Рис. 13 – Заповнений контейнер
Fig. 13 – Filled container

В наведеному прикладі, причина малої помітності видимих викривлень контейнеру обумовлена тим, що *max*спотворення, які вносяться до окремих пікселів в наслідок зміни їх яскравості (для випадку вставки даних повідомлення в нульові біти контейнеру), не перевищують величину 2^0, що лежить нижче порогу чутливості ЗСЛ до незначної зміни яскравості контейнеру [12]. Специфікацією формату **bmp24* загальна кількість рівнів квантування яскравості кожного окремого пікселю дорівнює $2^8=256$. Тоді оцінити поріг чутливості (ПЧ)ЗСЛ до незначної зміни яскравості зображення, можна як:

$$ПЧ = \frac{\Delta}{256} * 100\%, \quad (15)$$

де, Δ – величина внесених спотворень яскравості (*число рівнів квантування*) окремих пікселів при використанні методу НЗБ для приховування інформаційного повідомлення.

Високий рівень популярності методу НЗБ (*LSB)* зумовлений, тим, що він достатньо простий в реалізації та ефективний для приховування значних обсягів інформації в невеликих

файлових об'єктах [7]. Метод *LSB* може бути вразливим до різних видів атак, існуючих як у пасивних, так і в активних сценаріях атак. Основний недолік *LSB* полягає у його високій чутливості до найменших спотворень контейнера[12], наприклад: - компресія нерухомих зображень та/чи геометричні атаки, можуть призвести до втрати прихованої інформації чи її хибного відображення прихованого контенту. Щоб нівелювати можливі спотворення вилученого повідомлення, внаслідок зовнішнього впливу на заповнений контейнер, слід додатково використовувати методи завадостійкого кодування (Хеммінга, БЧХ та ін.).

7.Висновки

1. Розглянуто узагальнені структурні схеми стеганографічної і криптографічної систем, їх специфікація та компоненти, виконано огляд традиційних криптосистем, що використовуються в сучасних комплексних системах захисту інформації. Запропоновано бліц-огляд основних особливостей стеганоалгоритму НЗБ, що використовується для приховування даних в зображеннях та властивості ЗСЛ, котрі враховуються відомими стеганографічними методами при інкапсуляції (вставці) стеганографічного контенту в структуру контейнерів.

2. За результатами аналізу, можна стверджувати, що стеганографічну та криптографічну системи можна розглядати, як специфічний варіант системи зв'язку і передачі даних. Абстрактне визначення стеганографічної системи включає наступні множини: - множина вихідних інформаційних повідомлень; - множина контейнерів; - множина стеганограм; - множини прямих та обернених відображень; - множини ключів-екстракторів даних, які відповідають цим відображенням. Абстрактне визначення криптографічної системи охоплює такі множини, як: - множина вихідних повідомлень; - множина криптограм; - множини прямих та обернених відображень і відповідні їм ключі.

3. Асиметричні системи криптографічного захисту, більш стійкі до атак компрометації секретних ключів, оскільки використовуються різні ключі для процедур шифрування/розшифрування. Вагомою перевагою асиметричних криптосистем є застосування ключів великої довжини в криптографічних алгоритмах, що дозволяє підвищити обчислювальну складність. Складні математичні операції та ключі великих розмірів уповільнюють виконання асиметричних алгоритмів у порівнянні з симетричними. Асиметричні алгоритми вимагають значно більше обчислювальних ресурсів для здійснення високорівневих обчислень.

4. Алгоритм RSA використовує асиметричний підхід й широко використовується в сучасних протоколах автентифікації та забезпечення конфіденційності інформації в глобальних мережах. Збільшення довжини модуля призводить до зростання часу генерації ключової пари. Графіки залежностей для процедур шифрування та розшифрування різняться, оскільки на розшифрування даних витрачається більше часу, ніж на їх шифрування. Причиною цього може бути істотна різниця відритого, секретного ключа. Збільшення розміру файлу та модулю перетворення, призводять до збільшення часу, який витрачається на виконання процедур шифрування і розшифрування для системи RSA.

5. У стеганографії існує широкий спектр методів для реалізації приховування даних в нерухомі зображення. Вилучення вихідного повідомлення з криптограми вимагає не лише відповідний секретний ключ, але й збереження цілісності криптограми. Виходячи з цього, можна стверджувати, що не всі методи придатні для приховування криптограм в контейнерах, оскільки вони не можуть гарантувати безпомилкового вилучення інформації. При виборі методу потрібно звертати увагу на ймовірність правильного вилучення. Додатковим рішенням з протидії можливим помилкам є методи завадостійкого кодування, які дозволяють виявляти та виправляти можливі помилки стеганографічного повідомлення.

6. Комплексне застосування методів криптографії та стеганографії дозволяє забезпечити високий рівень захисту інформації від потенційних атак.

References

[1] Zamula, O. A., Horbenko, Y. I., &Shumov, O. I. (2010). The Regulatory and Legal Framework of Information Security. Integrated Information Protection Systems. Kharkiv: KhNURE.[In Ukrainian]

[2] Kuznetsov, O. O., Yevseyev, S. P., & Korol, O. G. (2011). Steganography. Kharkiv: KhNEU.[In Ukrainian]

[3] Shekhanin, K., Gorbachova, L., & Kuznetsova, K. (2021). Comparative analysis and study of information carrier properties for steganographic data hidden in cluster filesystems. Computer Science and Cybersecurity, (1), 37-49.[In Ukrainian] https://periodicals.karazin.ua/cscs/article/view/17266/15910 DOI: 10.26565/2519-2310-2021-1-03

[4] Kuznetsov, A., Shekhanin, K., Kolgatin, A., Kuznetsova, K., & Demenko, Y. (2018). Hiding data in file structure. Computer Science and Cybersecurity, 9(1), 43-52.[In Ukrainian]https://periodicals.karazin.ua/cscs/article/view/12013

[5] Lesnaya, Y., Goncharov, M., & Malakhov, S. The results of modeling attempts of unauthorized extraction of stego-content for various combinations of an attack on the experimental steganographic algorithm. Scientific Collection *«Inter Conf»*, (141), 338–345. Retrieved from https://archive.interconf.center/index.php/conference-proceeding/article/view/2319/2348

[6] Yesina, M., & Shahov, B. (2021). Research of implementation of candidates of the second round of NIST PQC competition focused on FPGA Xilinx family. *Computer Science and Cybersecurity*, (1), 16-36. https://periodicals.karazin.ua/cscs/article/view/17265/15909 DOI: 10.26565/2519-2310-2021-1-02

[7] Konakhovych, G. F., Progonov, D. O., & Puzirenko, O. Yu. (2018). Computer steganographic processing and analysis of multimedia data. Kyiv: "Centerfor Educational Literature".[In Ukrainian]

[8] The Decree of the President on the Regulations on the Procedure for Cryptographic Protection of Information in Ukraine" dated May 22, 1998, No. 505/98. [In Ukrainian] https://ips.ligazakon.net/document/u505_98?an=1&ed=1999_09_27

[9] Lakhno, V. A. (2016). Lecture Notes on the Discipline 'Fundamentals of Cryptographic Information Protection.' Kyiv. [In Ukrainian]

[10] CCNA Cyber Ops (Version 1.1) – Chapter 9: Cryptography and the Public Key Infrastructure. (2019). Вилучено з https://itexamanswers.net/ccna-cyber-ops-version-1-1-chapter-9-cryptography-and-the-public-key-infrastructure.html

[11] Tarnavsky, Yu. A. (2018). Information Security Technologies (pp. 107-108). Kyiv: Igor Sikorsky Kyiv Polytechnic Institute. [In Ukrainian]

[12] Kuznetsov, O. O., Poluyanenko, M. O., & Kuznetsova, T. Yu. (2019). Data hiding in the spatial domain of still images by modifying the least significant bit. Kharkiv: V. N. Karazin Kharkiv National University.[In Ukrainian]

Submitted November 17, 2023; Revised December 18, 2023; Accepted December 25, 2023

Authors:
Bodnia Mykyta, CSD Student, V.N. Karazin Kharkiv National University, Ukraine.
E-mail:bodnia2020kb12@student.karazin.ua
Yesina Maryna, Ph.D., Associate Professor, Department of Security of Information Systems and Technologies, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine; research associate-consultantof JSC"IIT", Kharkiv, Ukraine.
E-mail:m.v.yesina@karazin.ua
ORCID: https://orcid.org/0000-0002-1252-7606
Ponomar Volodymyr, Ph.D., researcherof Security of Information Systems and Technologies, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine; design engineerof JSC"IIT", Kharkiv, Ukraine.
E-mail:Laedaa@gmail.com
ORCID:https://orcid.org/0000-0001-5271-2251

Researching the possibilities of using steganographic and cryptographic algorithms for information hiding.

Abstract. The organization of information security has always been a relevant task, especially after the emergence of information and communication systems. The fundamental directions in the field of information security, dating back to ancient times, include cryptography and steganography. Cryptography implements information protection by transforming it into an unreadable form. Steganography allows the concealment of information in various containers (such as images, texts, audio recordings), keeping the presence of information unnoticed by casual observers. The article discusses approaches to cryptography and steganography, the concept of hybrid application of cryptographic and steganographic methods to provide a dual-layer data protection, and the overall architecture of cryptographic and steganographic systems. Traditional cryptographic systems applied in modern information security systems include symmetric and asymmetric cryptosystems. Although symmetric systems have evolved with the appearance of new mathematical transformations, they have a significant drawback. It consists of the need for an additional transfer of the secret key to the recipient. Such a strategy requires the use of a protected communication channel equipped with technical protection systems. At the same time, there is a risk of unauthorized access, which can cause the secret key to be compromised. Based on the above problems of symmetric cryptosystems, preference is given to asymmetric algorithms when developing protection mechanisms. An analysis of the RSA cryptosystem, based on an asymmetric encryption approach, has been conducted. This system is used in contemporary authentication protocols and ensures confidentiality in information systems and the Internet. The performance of software modules for key pair generation, encryption, and decryption for the RSA system was investigated by modifying the algorithm's general parameters (transform module, source data size). The results of time measurements are presented in a table, based on which dependencies of time on specific parameter modifications are built. The steganographic algorithm of modification of the least significant bit (LSB), which is used to hide data in images, is studied. Currently, there is a wide range of steganographic algorithms developed based on the characteristics of human sensory systems (such as vision or hearing). The article discusses the properties of the human visual system (HVS) utilized in steganography.

Keywords: *Cryptography, Steganography, Key, Information Message, Asymmetric Cryptosystem, Symmetric Cryptosystem, Ciphertext, Steganogram.*

DOI: 10.26565/2519-2310-2023-2-06
UDC621.391:004.056.5

RESULTS OF MODELING DIFFERENT SCHEMES OF THE SPATIAL ORIENTATION AND SCANNING SERIES OF BASE BLOCKS OF IMAGES TO CONFRONT AN UNAUTHORIZEDEXTRACTION OF STEGANOGRAPHIC DATA

Honcharov Mykyta, Malakhov Serhii, Kolovanova Ievgeniia
V. N. Karazin Kharkiv National University, St. Svobody Square, 4, Kharkiv, 61022, Ukraine
m.honcharov@student.karazin.ua ORCID ID: https://orcid.org/0000-0002-9790-7260 , malakhov@karazin.ua, ORCID ID: https://orcid.org/0000-0001-8826-1616 e.kolovanova@karazin.ua ORCID ID: https://orcid.org/0000-0002-0326-2394

Submitted October 5, 2023; Revised November 12, 2023; Accepted December 18, 2023

Abstract: This work presents the results of modeling attempts at unauthorized extraction of steganocontent (halftone test images) under the condition of selective compromise of each of the two active processing parameters of the source array series of base blocks (BB) of content, i.e.: - the scheme scanning of BB series and the spatial processing of BB. The current program version ensures consistent realization of the main stages of content processing with the necessary settings parameters. As part of the modeling, it is suggested that the attacker has correctly identified one of the two current content processing parameters. Several modifications of the main schemes scanning of BB series and the spatial orientation of BB (rotation and horizontal mirroring) as an additional mechanism to counteract attempts of illegitimate content extraction are considered. The modeling was conducted on the examples of three types of images: - portrait, landscape, and mnemonic scheme. Manipulations with the spatial orientation parameter of BB strengthen the opportunities to counteract attempts at unauthorized data extraction. Characteristic quantitative and time histograms for different dimensions BB of content, changes in the peak of value signal-to-noise ratio for different types of schemes scanning BB series are presented, and samples of attacked test images are presented. The analysis and generalization of the main differences in the attack results using different parameters of the spatial processing of BB and ways of scanning series of BB of image-content are performed. Attention is drawn to the fact that the use of two active processing parameters of the source array of BB series is an effective and computationally «simple» means of counteracting attempts at unauthorized data extraction. The relationship between the stage of preprocessing the source content and the parameters of the formed arrays BB is emphasized. It is concluded that the introduction into the structure of the data extractor key, the elements of «The state of scanning» and «The spatial processing of BB», strengthens the overall capabilities to counteract attacks. The used processing parameters of the source array of BB series determine the structure of visual artifacts of attacked images but do not produce a simple solution to identify the attacked image at the level of classifying the type of source images. Prospective directions for further modeling of the main protection mechanisms within the proposed algorithm concept are indicated.

Keywords: *Content, Steganography, Encoding Series Lengths, Images, Scanning, Spatial orientation, Encoding with transformation, Encapsulation, Data extraction.*

1.Introduction

One of the most effective directions to ensure the hiding of the facts of information transmission and storage is the use of various steganographic methods that make it possible to use the properties of digital content to ensure more effective solutions to the issues of hidden transmission, storage and protection against unauthorized extraction of target information. Regardless of the used steganography direction, it is necessary to ensure the minimization of unmasking anomalies of the data carriers (*containers*) applied and maintain a given level of content resistance to attempts of its unauthorized extraction, and in some cases, also resistance to attempts of deliberate container distortion.

When hiding (*encapsulating*) in digital images any other information (*in this case, images*), there are certain distortions of these objects - data carriers. Through the use of balanced settings of the data encapsulation algorithm, it is possible to ensure the level of distortion of the used container images at a level below the threshold of sensitivity of the human visual system. This ensures the actual absence of noticeable anomalies in information carriers, complicates the work of a steganoanalyst, and introduces the necessary balance between the preservation of characteristic properties for

the type of container used and the amount of permissible distortions acceptable for a given type of hidden content (*hereinafter referred to as steganocontent*) [1-2].

Undoubtedly, the number, structure, and manifestation intensity of artifacts of the image-content encapsulation process and the consequences of attempts to illegitimately extract, always depend on the processing modes chosen for them at all stages of the current prototype steganoalgorithm [3-4]. When processing container and content data, different processing modes can mostly be used, both the same type (symmetric) processing modes and modes that implement different data processing parameters (*asymmetric*) [2].

Such differences can include: the size of blocks (*fragments*) of the source images; parameters of pre-processing of data arrays of the container and content; criteria for evaluating the significant information of containers and content; differences in implementations of accelerating computing procedures, etc. Based on the totality of these differences, different effects can be obtained on the same types of source data [5] from the point of view of the visibility of image artifacts and individual parameters of the entire algorithm based on the results of the performed steganographic insertion. According to the concept of the being created algorithm [2], for authorized content extraction, information is required regarding the current data multiplexing parameters at both main security levels (*inter-block and intra-block*) [4,6], both for content and for the container. All this information is contained in the structure of the composite key of the data extractor, where each of its elements determines the current processing modes of the steganocontent and container [2]. Violation of each of its individual elements of the extractor key structure and/or the current parameters (values) leads to the impossibility of content extraction [7], or its significant distortion [4,8-9].

2. Main part

The main purpose of the paper is to summarize and compactly compile the results obtained during the cycle of modeling various scenarios of content attack that counteracts attempts at its unauthorized extraction, by using the use of different scanning schemes of base block (*BB*) series and spatial transformation schemes of BB of an array of image-content series when imitating a conditional attack of steganocontent, in the assumption that the attacker managed to determine the current parameters of content processing [10], which are implemented at 2 main levels of protection (*inter-block and intra-block, Figs. 2-4* [4]) of the investigated steganoalgorithm.

Within the scope of the conducted modeling, the most indicative (*from the point of view of the clarity of the obtained consequences*) parameters of the settings of the current algorithm [8,11-12] were used, which facilitates the general perception of the observed processes and the evaluation of the character and structure artifacts of the attacked images.

The current method of organizing the scanning of BB series [6] is determined by the corresponding element in the data extractor key structure (*element №2 in Table 1* [8]). The characteristic results of the attack (*attempts at unauthorized extraction*) of the test image-content when implementing some scan schemes are presented in works [8,13]. The work [14] presents the results of unauthorized content extraction attempts when implementing the two-pass scanning mode (*i.e., through block sampling*) of BB series for a test image of the «portrait» type.

In Fig. 2 presents the results that characterize the total number of BB and the average length of series BB for scanning schemes shown in Fig. 1. The imitation modeling of the test content attack was carried out for four image block dimensionality: *4×4, 8×8, 12×12* and *16×16 elements*. It should be noted that all the scanning ways (Fig.1) are not computationally complex, but they can significantly complicate the attacker's «work», increasing the overall protective potential of the algorithm [8,13].From Fig. 2 shows that an increase in the dimensionality of the blocks leads to a sharp decrease

(*comparison of the blue and red histograms*) in the number of BB image series [12], for all the considered scan methods (Fig.1).

The use of blocks of large dimensions (*red histogram in Fig.2*) virtually eliminates the difference in the number of BB series for different scanning methods. In other words, the indicator characterizing the number of series to be formed depends on the operating parameters (*scanning multiplicity*) and scanning scheme and decreases with increasing BB dimensionality [11].

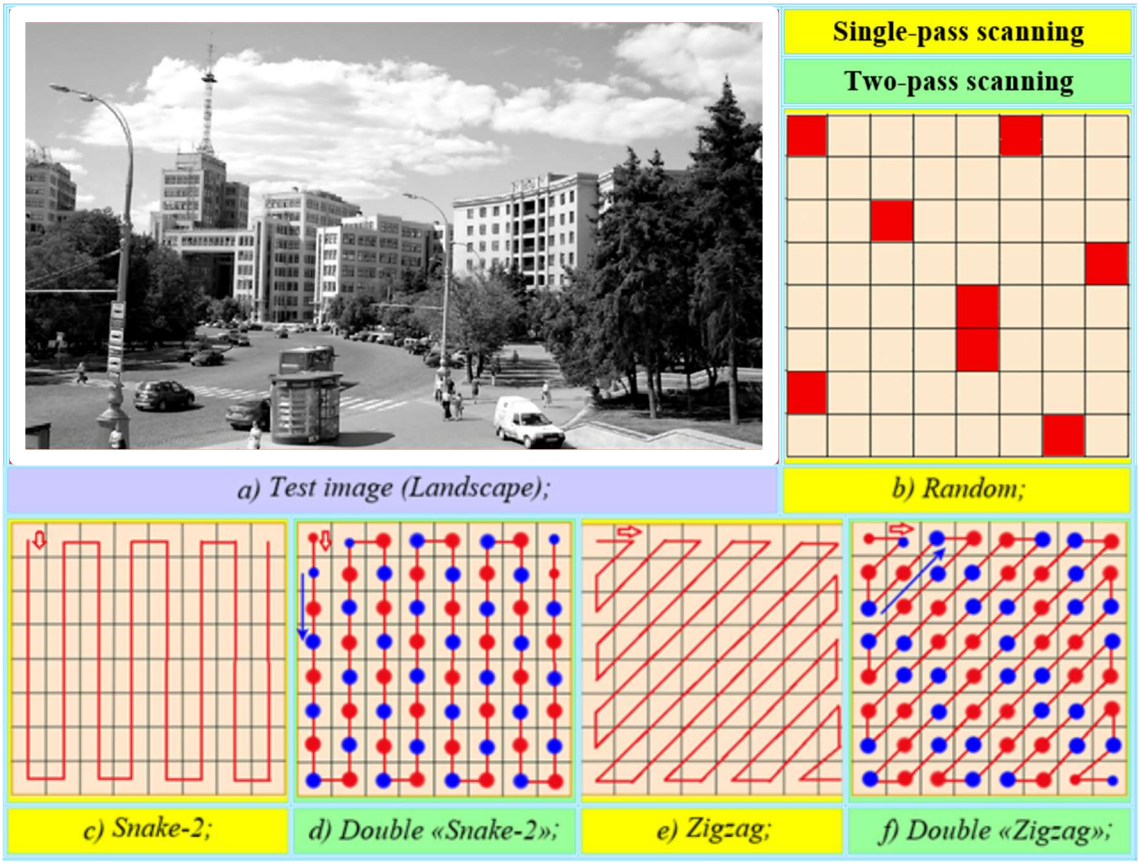


Fig. 1– The researched scanning schemes *(b-f)* and sample test image *(a)*.

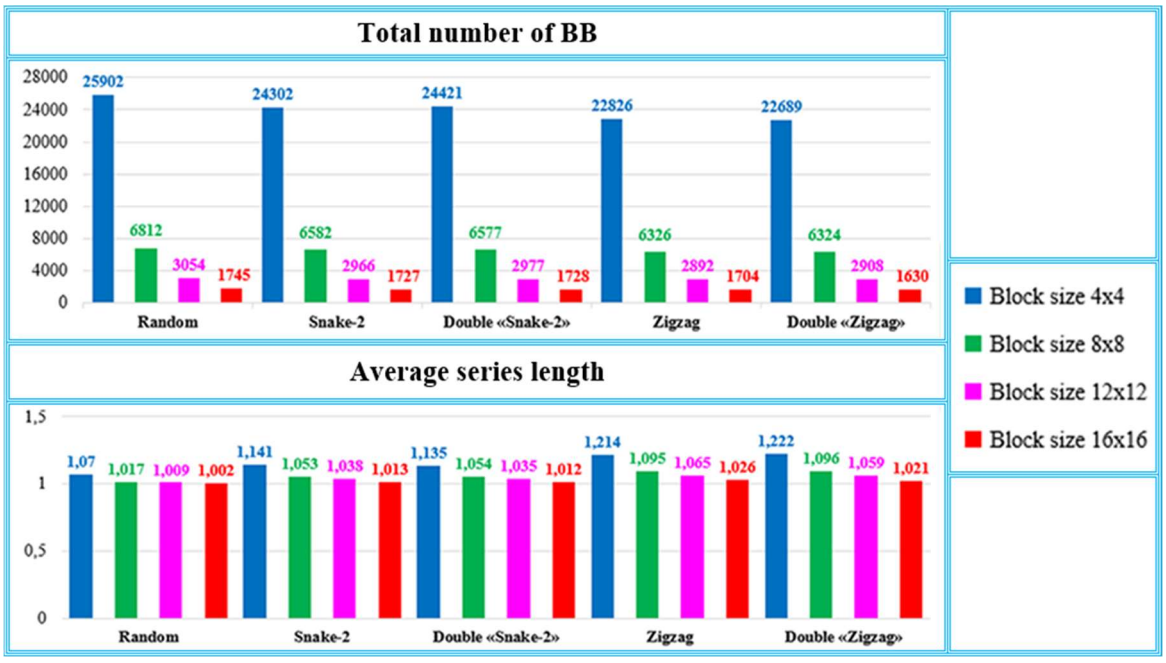


Fig. 2. The total number of obtained BB and their average series length from different scanning schemes and block sizes (*for test sample (a) in Fig. 1*)

It should be emphasized that the two-pass modification of the scanning (*var. (d,f) in Fig.1*) during the first scanning of the source content array involves the sequential sampling of all odd blocks (*red markers in samples (d,f) in Fig.1*), and during the second pass/scanning, all even blocks of content (*blue markers*).

The use of a two-pass scanning *(Fig.1, var.(d,f))*, according to the indicators of the formed series of BB and their lengths (*Fig. 2*), compared to the use of a random scanning mode (*var. (b)*), gives the closest results, and for all «practically interesting» (*8 and 12 el.*) dimensions of the blocks.

At the same time, the visual fragmentation (*destruction of the structure*) of the attacked content for the above cases is significantly different (*see Fig.4in work*[11]).

The random mode of scanning BB series (Random), in the case of a successful attack of 2 levels of protection at once [13], provides much greater fragmentation of the source content (*see Fig.3* [8]), but significantly increases the total processing time (Table 1).

With the use of «difficult» scanning ways and modes (*in this case, two-pass and/or random scanning*), the effect of visual fragmentation of the source content increases, namely, the number of formed series of BB increases. From the point of view of multiplexing combinatorics, this looks very good, but in doing so, unfortunately, it increases the computational complexity of the procedures at the 2nd level of protection, which is an undesirable effect that contradicts the general trend of reducing the computational complexity of the entire algorithm as a whole [2]. First of all, this concerns the implementation of coding procedures with transformation [5], immediately before the implementation of the procedures for multiplexing the average brightness parameter of the BB at the 2nd level of protection (*step №6 in Fig.1 in work* [13]) of the experimental algorithm.

Table 1 - Execution time for different scanning schemes and dimensionality of BB.

Execution time in the second [*sec*]					
Dimensionality of blocks	**Random**	**Snake-2**	**Double «Snake-2»**	**Zigzag**	**Double «Zigzag»**
Block 4×4	***36***	***0,08***	***0,2***	***0,08***	***0,13***
Block 8×8	***2,22***	***0,01***	***0,04***	***0,02***	***0,03***
Block 12×12	***0,53***	***0,001***	***0,012***	***0,008***	***0,013***
Block 16×16	***0,16***	***0,006***	***0,007***	***0,005***	***0,012***

The characteristic values of the execution time of different schemes scanning of BB are presented in Table 1. Based on the obtained results, the following can be stated:

- the execution time of the scanning schemes of BB series depends on the dimension of the blocks, the type of image, and, accordingly, the number of BB series to be formed;
- the total time of data processing procedures decreases when increasing the dimension of BB;
- application of the «*Random*» scanning scheme requires more time for all dimensions of blocks, and, compared to other scanning methods, this difference is very significant;
- the use of the multiplicity mechanism in the scanning schemes increases the number of logical procedures that are implemented within the corresponding instructions, which leads to an increase in the total processing time (*for example, comparing «Snake-2» and «Double Snake-2»*);
- the time of implementation of «simple» scanning schemes (*rows, columns or spiral, etc., see Fig.3(a-e) in* [8]) is much shorter than for «complex» schemes *(var. (b,d,f))*. However, the latter significantly complicates the attacker's ability to localize the vector of potential searches relative to the implemented scanning scheme.

Table 2 presents the *PSNR* values (*Peak Signal-To-Noise Ratio - PSNR*) which correspond to some samples of attacked content presented below in Fig. 4.

The analysis of the structure and intensity of the manifestations of artifacts of the attacked images (Fig. 4) shows that even the existence of acceptable *PSNR* values ($PSNR \geq 28 \div 30$ *dB* [7]) does not completely guarantee the successful identification of objects scene on all used in the course of modeling scanning schemes.

It should be emphasized that usually the value of *PSNR* ranges from 20 to 50 dB, i.e. the higher the value, the closer the restored image is to the original.

In Fig. 3 presents a visualization of the obtained difference between the original and recovered (*i.e., illegally extracted*) images at different dimensions of BB and ways of the series scanning for both types of test images. In this case, the more brightly the point and/or fragment of the image *(samples (a-d))*, the bigger the difference between the «hacked» content and its original.

Accordingly, than the indicated darker the element/fragment, the nearer its recovery parameters are to the original values. It should be emphasized that all the images shown in Fig.3 show attempts to falsely restore content by using a «by row» scanning scheme (see *var.(a) in Fig. 1 in work* [8]). Characteristic examples of unsuccessful selection of the current parameters of series scanning under the condition of simultaneous compromise of the other two levels of protection of the experimental algorithm are presented in works [8,11,13-14].

However, two important circumstances should be taken into consideration: - the type of content being processed and the degree of complexity of the reverse compilation of the source content during the attacker's attempts to «work» with the compromised data array [8,13]. From the point of view of the complexity of reverse compilation of the source content, it is worth highlighting the scheme of scanning that implements the «*Zigzag*» principle *(var. (e,f))* this scheme provides the greatest visual fragmentation of the content and makes it impossible for the attacker to obtain indirect instructions regarding the implemented method scanning of BB.

Compared to the «*Snake-2*» principle, which is characterized by pronounced visual transparency, the «*Zigzag*» scheme is an effective solution to complicate the reverse compilation, provides the greatest visual fragmentation of content, and deprives the attacker of indirect clues in the part of the implemented method of scanning of BB.

The variant of the random scanning scheme is not considered as a priority (from the point of view of the degree of visual fragmentation of the content), due to the decrease in the average length of the formed series (Fig. 2) in the most balanced, from the practical point of view, range of block sizes (*from 8×8 to 12×12 elements, Fig. 2*) and significant time losses (*Random in Table 1*), which are the result of the features of this scheme, this variant of the scanning scheme will be considered in the next part of the research.

Summarizing all of the above, it can be argued that the scanning schemes that implement the «*Zigzag*» scheme combine in the best way the structural features that are inherent for images of the city type and provide the best conditions for maximizing the difficulty of attempts to unauthorized reverse compilation of the source content. In addition, the possibility of implementing different «Zigzag» schemes (*for example, «start» at different points and/or through block scanning*) additionally increases the combinatorics of the corresponding element in the integrated structure of the extractor key [8]. Thus, even in the case of compromise of the main protective mechanisms at two multiplexing levels at once [4,13], the use of various variations of the «*Zigzag*» scanning scheme allows to successfully counteract attempts to unauthorized content extraction.

The next stage of the research was focused on modeling possible attack scenarios (*step №3 in Fig.1,* [4]), relative to two constituent components of the integrated data extractor key structure (*Fig.1 in* [8]): -an element that defines the current scanning scheme of BB [15]; -an element that

defines the current implementation of the spatial processing of BB of the content series length array (this is a new element).

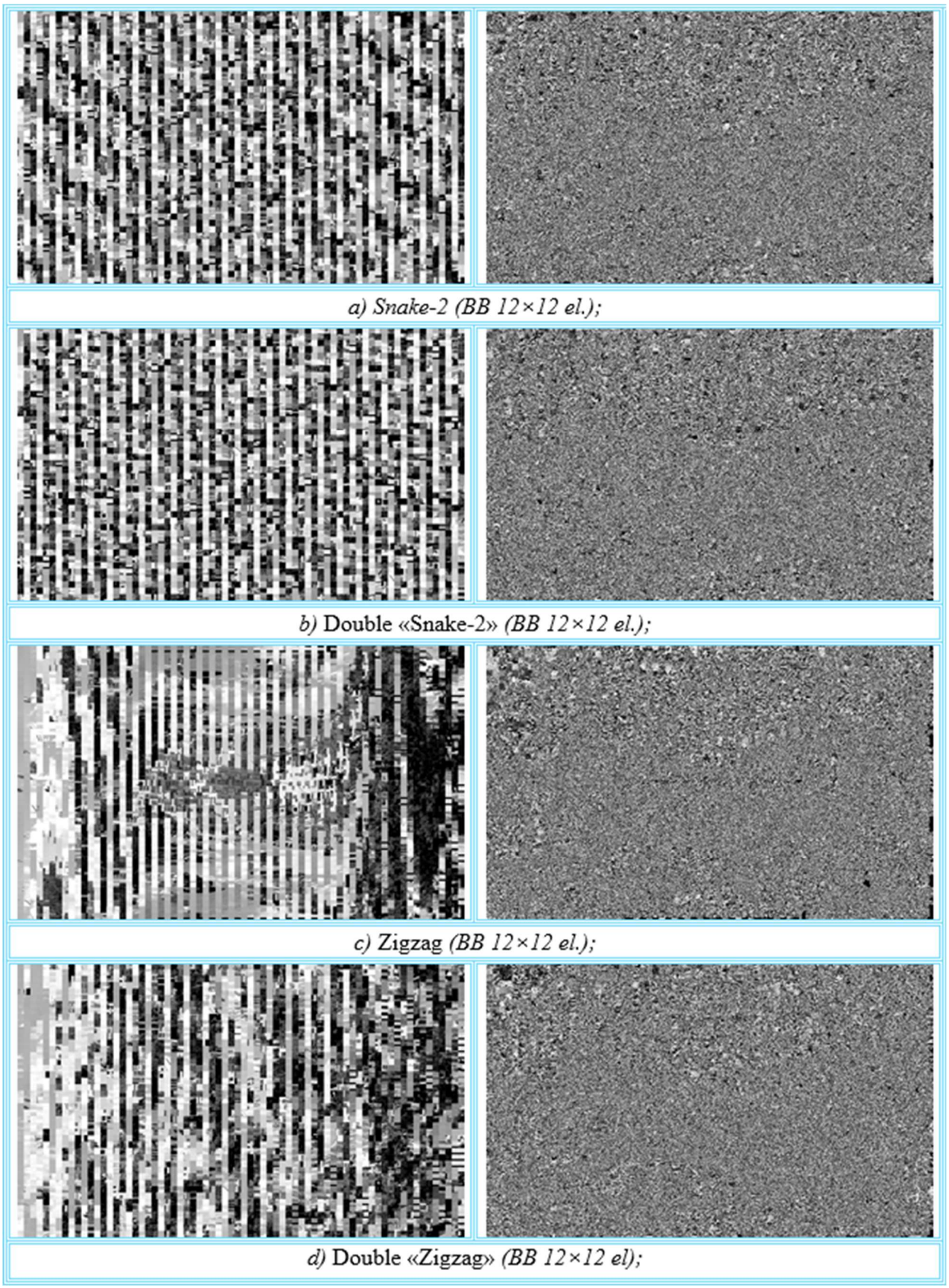


Fig. 3 - Visualization of the difference between the original and attacked content (*Landscape*) for different scanning ways

In Fig. 4 presents test samples of halftone images that are characteristic of 2 different types of content (*portrait and mnemonic scheme*) and the essence of the manipulations used with the spatial orientation parameter for all formed series (*see step 3 in Fig. 1 in [4]*), which were used during the modeling cycle.

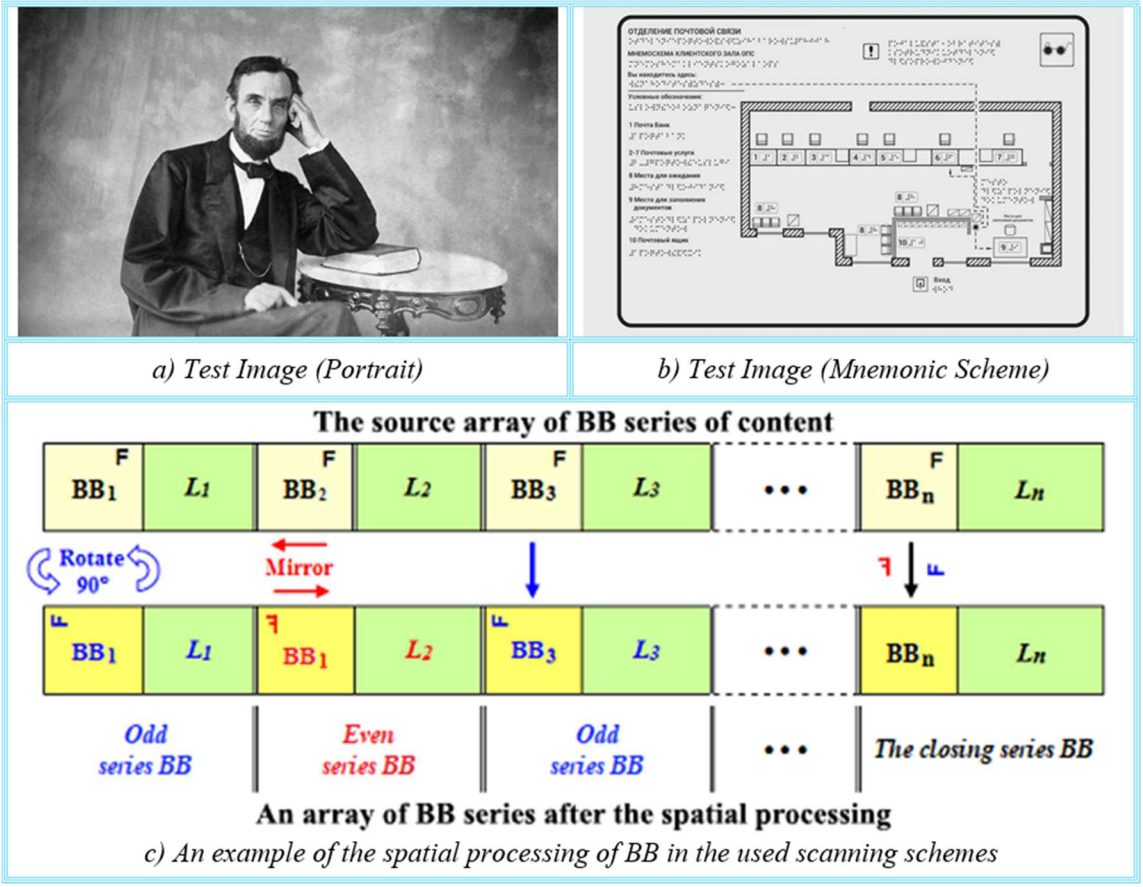


Fig. 4 - Samples of test images *(a, b)* and the used scheme for the spatial processing of BB *(c)*

Before starting the modeling, it was suggested that the parameter use of the spatial orientation of BB could introduce its own contribution to the distortion structure of the attacked image and determine the further course of events, regarding the «*success*» of attempts to unauthorized extract and identification of the target content. That is, the introduction of different schemes of the spatial processing of BB of the source images can strengthen the general combinatorics of the integrated structure key of the data extractor [2]. Therefore, even in the case of compromise of the main protective mechanisms at both levels of multiplexing [2,4], the application of different schemes of the spatial orientation of BB will allow us to successfully counteract attempts of illegitimate content extraction. In accordance with the idea, when encoding content, for all odd base blocks, the blocks are rotated to the left by 90° (*marked as Rotate 90°*), and for all even base blocks, their horizontal mirroring (*Mirror*) is performed. Thus, after the formation of the array of BB series, when using the appropriate scanning scheme (*in Fig. 4, marked as «The source array of BB series …»*), there is a change in the source orientation of BB in such a way that all neighboring blocks of the resulting array (*in Fig. 4, marked as «An array of BB series after…»*) have a different spatial position. At the same time, as follows from Fig.4, the simplest test encoding scheme was used, which is repeated without changes for each subsequent pair of BB.

Taking into consideration the possible combinations/consequences of the attack which are implemented in relation to the two elements indicated above (*scanning and spatial orientation of BB*), the general modeling scheme is as follows:

1 - Successful selection of the scanning scheme, but an error in the spatial orientation of BB;
2 - Successful selection of the spatial orientation of BB, but an error in the scanning scheme.

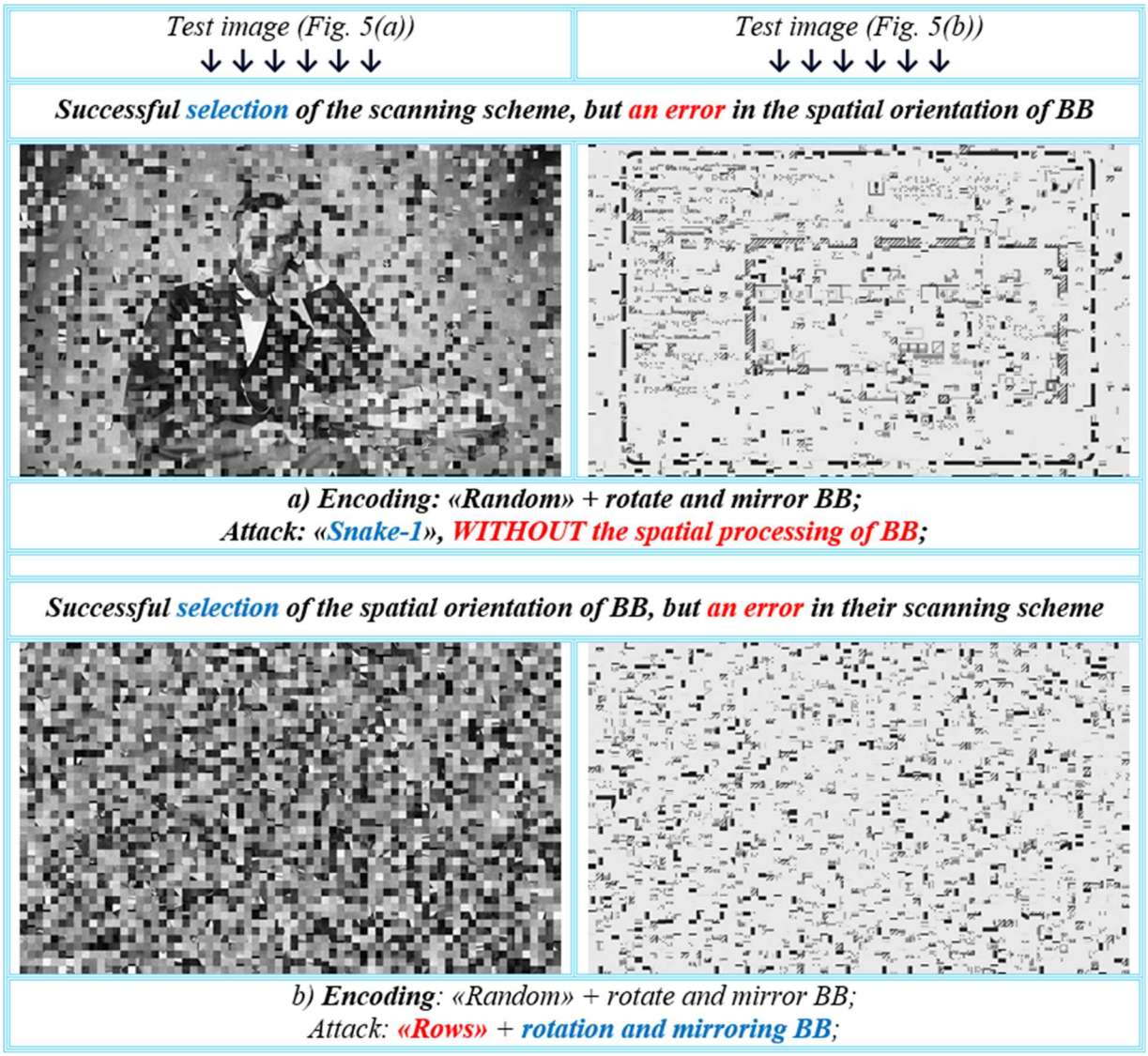


Fig. 5. The results of content recovery when different combinations of attacks *(BB 12×12 el.)*.

The corresponding results of modeling attempts at unauthorized content extraction, presented in Fig. 5, were obtained when the same parameters of the algorithm [2]: - the dimension of image blocks; - the dimension of the smoothing matrix; - the smoothing method; - the parameter value of P_Z. In addition, it is important to emphasize that the same P_Z values were used at the stages of content pre-processing and the formation of the array of BB series.

In Fig. 5(*a*) presents the results of attempts to unauthorized extract test images in the assumption of correct selection of the series scanning active scheme and an error in the parameters of the spatial orientation of BB when using the average values of the dimensionality of BB and the value of P_Z for the smoothing mask *3×3 el.* [2]. From the sample in Fig.5(*a*), it is clearly visible that the errors of spatial positioning of the image blocks for both scannings are practically invisible in lengthy and low-information image fragments *(see the attacked image for the test image type «Mnemonic scheme»)*. This feature of processing is by no means a «*weak*» side of the algorithm used, since in highly detailed image areas, the process of fragmenting a series of similar blocks demonstrates all the necessary qualities. That is, in these areas, the necessary decompilation of content is supported, which makes further identification of image objects impossible.

In Fig. 5(*b*) presents the results of attempts to unauthorized extract test images, assuming that the attacker successfully selected the current parameters of the spatial orientation of the formed BB (Fig. 4(*c*)), but made a mistake when restoring the current scheme scanning series of BB. That is, in this case, the attacker correctly determined the dimensionality of BB and the current spatial orientation scheme of the available BB, but made a mistake in the part of the implemented scanning scheme, namely: - the attacker used the «*Rows*» scheme for the initial «*Random*» scanning. In other words, the samples presented in Fig. 5(*b*) reflect the situation opposite to the one presented earlier in Fig. 5(*a*). The analysis of the samples presented in Fig. 5(*b*) allows us to state that the distortion structure and fragmentation intensity of the attacked images of unauthorized extracted content differs significantly from the results obtained when imitating the conditions of successful selection of the current parameters of spatial orientation of BB (Fig. 5(*a*)).

As revealed by the analysis of the presented results in both attack scenarios, the uncompromised part of the extractor key elements (*highlighted in blue letters in Fig. 4*) plays a critical role in preventing further identification of objects in the unauthorized extracted content. This testifies to the effectiveness of the protective measures that are applied to the elements of the extractor key and indicates their importance for preserving the integrity and confidentiality of data. However, it is important to note that the used scheme scanning «*Random*» may allow an attacker to partially identify the content in case the attacker is able to pick up the current series scanning scheme but makes a mistake in the current scheme spatial processing BB.

This indicates the need for further measures to improve security, in particular, the choice of the most complex and secure scanning schemes (*«Zigzag» or «Double Zigzag»*) and/or spatial orientation of BB, it is recommended to set more secure algorithm parameters, use images of a different type and different pre-processing options to create optimal starting conditions for improving both the performance of the algorithm and providing more effective protection to complicate the process of identifying confidential information (*for example, in the «Mnemonic» type image in Fig. 5, it actually deprives the attacker of the ability to classify the type of content extracted*).

The solution to this problem always requires careful analysis and selection of optimal algorithm parameters (*in this case, scanning schemes*) which will ensure a high level of security and make it impossible for attackers to gain access to confidential information.

In Fig. 6-7 presents a visualization of the existing difference between the original and the restored *(i.e., illegally extracted)* images for the above smoothing parameters[2], but in conditions of simultaneous error in determining the current scanning parameters and spatial orientation of the available BB *(i.e., false rotation and mirroring of the BB (see Fig. 5(c))*.

In this case, the more brightly a point or any image fragment (*samples (b) and (d) in Fig. 6*), then the greater the difference between the attacked content and its original. Accordingly, the darker the specified element or fragment, then the closer its recovery parameters are to the original image values *(the brightness level of the original elements)*. Characteristic examples of unsuccessful selection of the current parameters of the scanning series without any manipulation of the spatial orientation of the available BB, in the conditions of simultaneous compromise at once of 2 main levels of protection *(inter-block and intra-block)*, presented in works [8,11,14]. Comparison of the image samples in Fig. 6*(b,d)* with their originals demonstrates that even the presence of a large number of dark elements in the «hacked» content does not contribute to the successful visual identification of scene objects (*although the obtained PSNR values do not exclude this possibility* [15]).

When analyzing the structure and intensity of distortions (*the difference in brightness between the original and restored elements*) of the attacked sample of the «*Portrait*» test image in Fig.6*(b,d)*, the following conclusions can be drawn.

When using the «*Random*» scanning scheme, the interconnections between neighboring blocks are strongly destroyed, which is characterized by a large number of small details (*high-frequency components in spectral analysis* [5]). This area with the structure of existing distortions (*fine grain*) differs from the rest of the peripheral part of the image. It is noticeable that in this case, the structure of the «grain» of recovery errors is proportional to the dimensionality of the blocks used. This confirms the presence of a large number of BB with a single or very small length of the formed series of BB.

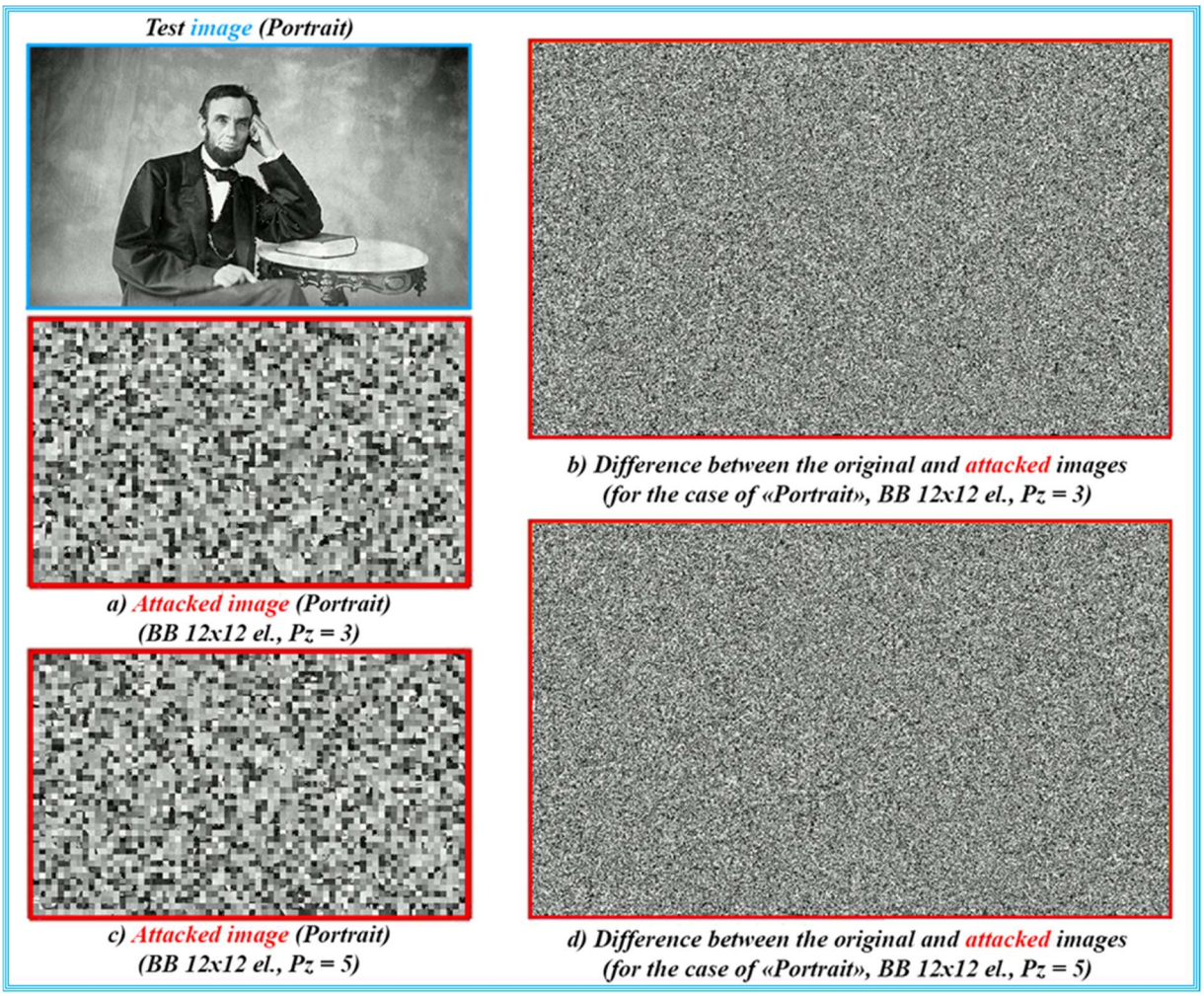


Fig. 6. Distortion structure of a test image of the «*Portrait*» type with a simultaneous error in determining the scanning and the spatial orientation of the BB

In Fig. 7(*b*), special attention should be paid to the area of the image, which is highlighted with a yellow marker. In this example, the yellow marker outlines the image fragments with the most noticeable distortions, which correspond to the least informative content areas: 1 - the background of the image; 2 - the area with captions and schemes on the mnemonic scheme. That is, in these image fragments, the «work» of the algorithm to form «*long*» series of BB is most noticeable. In other words, within the limits of the «yellow area» in Fig. 7(*b*), mostly there are blocks that are very close to their source content, which is not the case with the results of the visual evaluation of the attacked samples (*in Fig. 7(a) and (c), highlighted in red*).

Thus, the used processing algorithm ensures the lowest level of distortions in the structure of BB, which form highly detailed fragments of images, realizing content protection in these areas due to the implementation of appropriate scanning schemes and changes in the spatial orientation of existing BB. Moreover, the attacker's mistake in the two specified parameters at once only increases the overall effect (*the lower thumbnails marked with a red frame in Fig. 7*(*a,c*)).

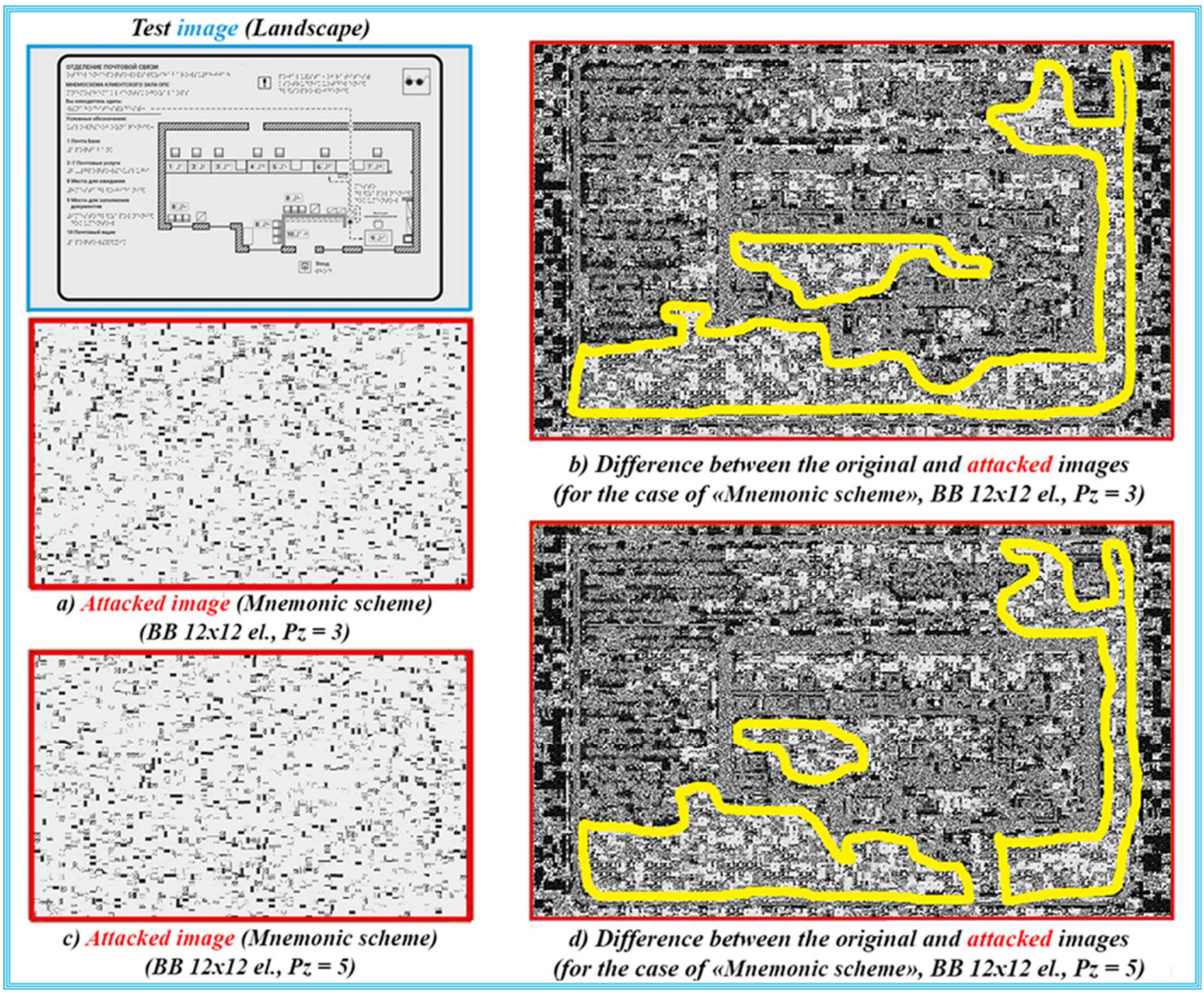


Fig. 7. Distortion structure of a test image of the «*Mnemonic scheme*» type with a simultaneous error in determining the scanning and the spatial orientation of the BB

3.Conclusions

1. The conducted modeling is of a demonstration nature and should confirm the main assumptions regarding the selected data processing modes at each stage [4, 11], as part of implementing the general concept of creating a low-resource hybrid steganoalgorithm [2].

2. The conducted simulation allows us to visualize the consequences of using different attack schemes (*attempts of unauthorized extraction*) of steganocontent under the condition of selective compromise of each of the two current processing parameters of the output array of BB content series, i.e.: - *the scheme scanning and the variant of spatial positioning of existing BB*.

3. The use of different scanning schemes highlights the importance of supporting the necessary compromise between: - the complexity of implementing one or another scanning method and its capabilities, in relation to countering unauthorized content extraction attempts and reducing the total number of series, as a pledge of the process of reducing the computational complexity of the entire algorithm [2].

4. The structure of the artifacts of the attacked images does not allow identifying the obtained content samples, at least at the level of classifying the type of source images (Figs. 6-7).

5. The conducted modeling confirms that changes in the spatial orientation of the formed BB are an effective tool to counteract attempts of unauthorized content extraction, even if successful selection of the current BB scanning scheme.

6. The introduction in the structure of the extractor key [8] of a new element that is responsible for the spatial processing of BB (Fig. 5(*c*)) allows us to reduce the requirements for the complexity of the used schemes of scanning, which is an important component within the chosen concept of implementing a low-resource hybrid steganoalgorithm.

7. Errors in the spatial orientation of content blocks in low-information image fragments are practically imperceptible, but this is not a «weak» side of the used algorithm, since in highly detailed image areas, the necessary decomposition of the source content is maintained, which makes its further identification impossible.

8. The used data processing modes provide a low level of distortion in the BB structure, which forms highly detailed image fragments, implementing content protection in such areas due to the use of appropriate scanning schemes and spatial orientation of the available BB. An attacker's mistake in both of the mentioned parameters increases the overall destructive effect (*i.e., content fragmentation*).

9. The used method of scanning BB content series determines the nature and structure of the distortions of the attacked images and determines the further course of events regarding the success of unauthorized extraction attempts and the identification of target content.

10. From the obtained results, it can be seen that even a successful selection of current data processing parameters at two main levels of protection does not guarantee successful reverse compilation of the source content, as proven by samples of «attacked» images.

References

[1] Hrybunin, V. G. Digital Steganography / Hrybunin, V. G., Okov, I. N., Turintsev, I. V. – M: Solon-Press, 2002. – 272 p. [In Russian]

[2] Lesnaya, Y., Goncharov, N., & Malakhov, S. (2021). Elaboration of the concept of multi-level data multiplexing for a hybrid steganographic algorithm. Collection of scientific papers SCIENTIA. (Vol. 2),48-55. [In Ukrainian] https://ojs.ukrlogos.in.ua/index.php/scientia/article/view/17666

[3] Goncharov, M., Lesnaya, Y., &Malakhov, S. (2021). Investigation of properties of the prototype hybrid steganographic algorithm. Computer Science and Cybersecurity, (2)., 45-56.[In Ukrainian] https://doi.org/10.26565/2519-2310-2021-2-05

[4] Lesnaya, Y., Goncharov, M., & Malakhov, S. (2023). Results of modeling attempts of unauthorized extraction of stego-content for various combinations of attacks of the experimental stegoalgorithm. Scientific Collection «Inter Conf», (141), 338–345. [In Ukrainian] https://archive.interconf.center/index.php/conference-proceeding/article/view/2319

[5] Pratt, W. (1985). Digital Image Processing (Translated from English by D. S. Lebedeva). Vols. 1, 2. M: Mir. [In Russian]

[6] Goncharov, N., Lesnaya, Y., & Malakhov, S. (2022). Adaptation of the Run-Length Encoding Principle to Counter Unauthorized Extraction Attempts of Steganographic Content. Grail of Science, (17), 241-247. [In Ukrainian] https://doi.org/10.36074/grail-of-science.22.07.2022.042

[7] Kuznetsov, O.O., Yevseyev, S.P., Korol, O.G. (2011). Steganography: a textbook. Kharkiv: Publishing House of Kharkiv National Economic University. [In Ukrainian] http://repository.hneu.edu.ua/handle/123456789/2289

[8] Lesnaya, Y., Goncharov, M., Azarov, S., & Malakhov, S. (2023). Visualization of unauthorized extraction attempts of steganographic content with misidentification of active series unfolding methods. Grail of Science, (24), 335–340. [In Ukrainian] https://doi.org/10.36074/grail-of-science.17.02.2023.061

[9] Lesnaya Y., Goncharov M., & Malakhov S. (2023). Results of intrablock multiplexing of the average brightness parameter of reference blocks of steganographic contenton a rearrangement-based basis. Scientific Debates and Prospective Directions of Scientific Development: Collection of Scientific Papers "ΛΌΓΟΣ" from the Materials of the IV International Scientific and Practical Conference (pp. 78-81). November 11, 2022. Paris, France.«ΛΌΓΟΣ». https://doi.org/10.36074/logos-11.11.2022.21

[10] Lesnaya, Y., Goncharov, M., & Malakhov, S. (2023). Methods of unfolding serial parameters of reference image blocks as an element of the composite key of the data extractor of a steganographic algorithm. Grail of Science, (23), 254–258. [In Ukrainian] https://doi.org/10.36074/grail-of-science.23.12.2022.37

[11] Lesnaya, Y., Goncharov, M., Semenov, A., & Malakhov, S. (2023). Modeling the unfolding of series of reference image blocks as a tool to counter attempts of unauthorized extraction of steganographic content. Grundlagendermodernenwissenschaftlichenforschung: Collection of scientific papers "ΛΌΓΟΣ" based on the materials of the IV International Scientific and Practical Conference (pp. 109–116). March 31, 2023. Zurich, Switzerland: ΛΌΓΟΣ. https://archive.logos-science.com/index.php/conference-proceedings/issue/view/9

[12] Goncharov O., Lesnaya Y., Pohorila K., Bohdanova Y., Malakhov S. Study of the parameter "series of reference blocks" as an element of the composite key of the data extractor of the steganographic algorithm.// Problems of science and practice, tasks and ways to solve them. Proceeding softhe XX International Scientific and Practical Conference. Warsaw, Poland. 2022. Pp. 779-785. https://doi.org/10.46299/ISG.2022.1.20

[13] Lesnaya, Y., Goncharov, M., & Malakhov, S. (2023b). Methods of unfolding the parameters of series of reference blocks of images as a component of the composite key of the data extractor of the steganographic algorithm. Grail of Science, (23), 254–258. [In Ukrainian] https://doi.org/10.36074/grail-of-science.23.12.2022.37

[14] Lesnaya, Y., Goncharov, M., Malakhov, S., & Melkozyorova, O. (2023). Results of unauthorized extraction of steganographic content in the implementation of two-pass unfolding of series of output blocks. Ricerche scientific he e metodidellalororealizzazione: esperienzamondiale e realtàdomestiche: Collection of scientific papers "LOGOS" based on the materials of the III International Scientific and Practical Conference (pp. 65-67). March 3, 2023. Bologna, Italy.«ΛΌГОΣ». https://doi.org/10.36074/logos-03.03.2023

[15] Honcharov, M., & Malakhov, S. (2023). Investigation of methods for unfolding output blocks of image steganography as a mechanism to counter unauthorized dataextraction. Scienceand Technology Today, 4(18). [In Ukrainian] https://doi.org/10.52058/2786-6025-2023-4(18)-293-308

Надійшла до редакції 5 жовтня 2023 р. Переглянута 12 листопада 2023 р. Прийнята 18 грудня 2023 р

Автори:
Гончаров Микита, аспірант кафедри безпеки інформаційних систем та технологій, Харківського національного університету імені В. Н. Каразіна, Україна.
ORCID ID: https://orcid.org/0000-0002-9790-7260
E-mail:m.honcharov@student.karazin.ua

Малахов Сергій, к.т.н., с.н.с., доцент кафедри безпеки інформаційних систем та технологій, Харківського національного університету імені В. Н. Каразіна, Україна.
ORCID ID: https://orcid.org/0000-0001-8826-1616
E-mail:malakhov@karazin.ua

КоловановаЄвгенія, к.т.н., доцент кафедри безпеки інформаційних систем та технологій, Харківського національного університету імені В. Н. Каразіна, Україна.
ORCID ID: https://orcid.org/0000-0002-0326-2394
E-mail:e.kolovanova@karazin.ua

Результати моделювання різних схем просторової орієнтації та розгортки серій опорних блоків зображень для протидії несанкціонованої екстракції стеганографічних даних.

Анотація. В роботі представлені результати моделювання спроб несанкціонованого вилучення стеганоконтенту (напівтонових тестових зображень) при умові вибіркової компрометації кожного з двох діючих параметрів обробки вихідного масиву серій опорних блоків (ОБ) контенту, тобто: - схеми розгортки серій ОБ та просторової обробки ОБ. Діюча програмна версія забезпечує послідовну реалізацію основних етапів обробки контенту з потрібними параметрами налаштувань. В рамках моделювання зроблено припущення, що атакуючий вірно визначив один із двох діючих параметрів обробки контенту. Розглянуто декілька модифікацій основних схем розгорток серій ОБ та просторової орієнтації ОБ (*обертання та горизонтальне віддзеркалення*), як додаткового механізму з протидії спробам нелегітимної екстракції контенту. Моделювання проводилося на прикладах трьох типів зображень: - портрет, пейзаж та мнемосхема. Маніпуляції з параметром просторової орієнтації ОБ, посилюють можливості з протидії спробам неавторизованого вилучення даних. Представлено характерні кількісні та часові гістограми для різних розмірностейОБ контенту, зміни пікового значення сигнал/шум для різних різновидів схем розгортки серій ОБ та наведено зразки атакованих тестових зображень. Виконано аналіз і узагальнення основних відмінностей результатів атаки при використані різних параметрів «Просторової обробки»ОБ та «Способів розгортки» серій ОБ зображення - контенту. Звернено увагу, що використання двох діючих параметрів обробки вихідного масиву серій ОБ є ефективним та обчислювально «простим» засобом з протидії спробам неавторизованої екстракції даних. Підкреслено взаємозв'язок між етапом предобробки вихідного контенту та параметрами формованих масивів ОБ. Зроблено висновок, що введення до структури ключа екстрактору даних, елементів «Стану розгорток» та «Просторової обробки ОБ», посилює загальні можливості з протидії атакам. Використовувані параметри обробки вихідного масиву серій ОБ, визначають структуру візуальних спотворень атакованих зображень, але не дають простого рішення, щодо наступної ідентифікації атакованого зображення на рівні класифікації типу вихідних зображень. Зазначені перспективні напрями для подальшого моделювання основних механізмів захисту, в межах запропонованого концепту алгоритму.

Ключові слова: *контент, стеганографія, кодування довжин серій, зображення; розгортка, просторова орієнтація, кодування з перетворенням, інкапсуляція, екстракція даних.*

DOI: 10.26565/2519-2310-2023-2-07
УДК 004.056.5

ВПЛИВ РІЗНИХ ФОРМ КІБЕРЗАГРОЗ НА СТІЙКІСТЬ ІНФОРМАЦІЙНИХ СИСТЕМ: АНАЛІЗ ТА СТРАТЕГІЇ ЗАХИСТУ

Євгеній Осадчий[1], Марина Єсіна[1,2], Віктор Онопрієнко[2]

[1]Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків, 61022, Україна
xa12850357@student.karazin.ua, m.v.yesina@karazin.ua
[2]АТ «Інститут Інформаційних технологій», вул. Коломенська 15, Харків, 61166, Україна
v25258@gmail.com

Надійшла до редакції 9 листопада 2023 р. Переглянута 16 грудня 2023 р. Прийнята 24 грудня 2023 р

Анотація: Дана робота присвячена дослідженню проблематикикібербезпеки в контексті сталого розвитку сучасного інформаційного суспільства. Починаючи з огляду різноманітних форм кіберзагроз, у статті запропоновано аналіз їхнього впливу на конфіденційність, цілісність та доступність інформації. Критична залежність сучасного суспільства від інформаційних технологій, робить тематику захисту від кіберзагроз надзвичайно актуальною. В межах роботи запропоновано аналіз зростання кількості та складності кіберзагроз, що вимагає постійного удосконалення та оновлення стратегій захисту від них. Важливим етапом висвітлення теми є аналіз впливу різних форм кіберзагроз на сучасні інформаційні системи. Розглянуто основні різновиди фішингу та соціальної інженерії, а також наслідки впливу вірусів, троянських програм та інших шкідливих програм. Детальний огляд цих аспектів дозволяє визначити ключові питання та небезпеки, які виникають в контексті проблематики кіберзагроз. Також, стаття містить матеріали, присвячені різним стратегіям захисту. Вона розглядає існуючи стратегії для захисту інформаційних систем, включаючи виявлення вразливостей, використання багатофакторної автентифікації та заходи для забезпечення стійкості. Загальні висновки даної роботи підсумовують необхідність постійного оновлення та адаптації стратегій захисту, щодо зростаючої складності кіберзагроз у світі швидкого технологічного розвитку. В цілому, дана робота е ще одним кроком у розумінні сутності викликів, які пов'язані із проблематикою забезпечення кібербезпеки в сучасному інформаційному суспільстві.
Ключові слова:*кіберзагроза, аналіз та захист, стійкість інформаційних систем, стратегії захисту.*

1. Вступ

У сучасному інформаційному суспільстві питання кібербезпеки стають надзвичайно актуальними, оскільки з кожним днем зростає обсяг цифрової активності та залежність від інформаційних технологій. Разом із швидким розвитком технологій зростає і рівень кіберзагроз, які стають важливим аспектом забезпечення безпеки в Інтернет просторі. Ці загрози викликають серйозні проблеми, а також стають причиною порушення конфіденційності, цілісності та доступності інформації. У межах цієїроботи в стислому вигляді розглянуті різноманітні форми кіберзагрозта їхній вплив на сучасні інформаційні системи (ІС), а також можливі заходи для захисту від них в умовах постійно зростаючого цифрового середовища.

Актуальність теми зумовлена впливом відразу кількох ключових аспектів. По-перше, інформаційні технології стали не тільки необхідною частиною повсякденного життя, але і критично важливим ресурсом для функціонування великої кількості суспільних, комерційних та господарських процесів. По-друге, зростання залежності від цих технологій відкриває нові можливості для кіберзлочинців, які використовують різноманітні та вдосконалені методи для атак на ІС.Забезпечення надійності і безпеки ІС стає надзвичайно важливим завданням, оскільки кіберзагрози, такі як атаки на мережеві структури, витоки конфіденційної інформації та шкідливі програми, можуть мати серйозні наслідки для економіки, політики та суспільної безпеки. У цьому контексті розуміння різних форм кіберзагроз та їхнього впливу стає стратегічно важливим для розробки ефективних заходів захисту, що відповідають викликам сучасного інформаційного простору.

Зростання кількості та складності кіберзагроз стає серйозним викликом для сфери кібербезпеки. Різноманітність атак, включаючи витончені техніки фішингу, атаки з використанням шкідливого програмного забезпечення (ПЗ) та атаки на інфраструктуру, свідчать, що кіберзлочинці постійно вдосконалюють свої методи, адаптуючись до новітніх технологій та змін

у сфері кібербезпеки [1].Неперервний розвиток кіберзагроз вимагає не лише реактивних, але й проактивних стратегій захисту. Організації та індивіди, що прагнуть залишатися попереду, повинні не лише оновлювати свої системи та ПЗ, але й розвивати нові методи виявлення та запобігання кібератак. Важливість цього завдання зумовлена тим, що відповідальність за захист ІС стає не тільки завданням технічних спеціалістів, але й ключовою складовою стратегічного управління будь-якою організацією чи державною установою. У цьому контексті огляд різних форм кіберзагроз та їхнього впливу стає невід'ємною частиною ефективного й безпечного управління, спрямованого на забезпечення стабільності та надійності функціонування сучасних ІС.

Проведення аналізу впливу кіберзагроз на стійкість ІС є невід'ємною частиною вдосконалення стратегій кібербезпеки в умовах постійного еволюційного середовища [2]. З урахуванням стрімкого розвитку технологій та збільшення кількості цифрових аспектів нашого життя, зростає і сфера кіберзагроз, що накладає серйозний вплив на ІС.

Ця стаття має на меті розглянути проблематику кібербезпеки в контексті сучасних реалій й пропонує оглядовий аналіз різноманітних різновидів кіберзагроз, який охоплює їх вплив на конфіденційність, цілісність та доступність інформації. Зокрема, надаючи огляд найновіших тенденцій у сфері кібербезпеки, автори роботи мають на меті виокремити ключові аспекти безпеки, що піддаються ризику внаслідок сучасних кібератак.

Важливим етапом у запропонованому аналізі є визначення різних стратегій захисту, які спроможні ефективно відповідати викликам безпеки сучасного кіберпростору. Слід під креслити, що розглянуті стратегії враховують, як технічні, так і стратегічні аспекти, котрі спрямовані на удосконалення стійкості сучасних ІС та забезпечення їхньої функціональності в умовах постійного впливу широкого спектру загрози інформаційної безпеки (ІБ).

2. Різновиди кіберзагроз та їх вплив на ІС

2.1 Фішинг та соціальна інженерія

Фішинг та соціальна інженерія стали неодмінною частиною сучасного цифрового простору, ставши важливими елементами кібербезпеки. Ці методи атак, спрямовані на отримання конфіденційної інформації через маніпулювання психологією користувачів, стали більш витонченими та поширеними, викликаючи серйозні загрози для особистої та корпоративної безпеки. Цей розділ розглядає методи фішингу та соціальної інженерії, їх вплив на користувачів та пропонує практичні підходи до захисту від цих загроз [1].

2.1.1 Фішинг: відомі методи та їх варіації

Фішинг – це один із найбільш поширених методів атак в сфері кібербезпеки, який використовує соціальні інженерні техніки для отримання конфіденційної інформації, такої як паролі, номери банківських карт або особисті дані, від користувачів. Нижче наведено узагальнений перелік найбільш поширених методів і варіацій фішингу, наслідків їх впливу на користувачів та можливих стратегій захисту.

Основні методи фішингу:

- *Електронна пошта (E-mail): підступні листи, що виглядають як від відомих вам компаній чи сервісів, які закликають вас ввести конфіденційні дані на фіктивних веб-сайтах.*
- *Соціальні мережі та месенджери: фішингові атаки через популярні соціальні мережі та месенджери, де атакуючі видають себе за знайомих чи колег.*
- *Веб-сайти: створення фішингових веб-сайтів, які імітують офіційні ресурси для отримання особистої інформації.*

Варіації фішингу:

- *Розмовний фішинг – вішинг (Vishing): фішинг через телефонні дзвінки, де атакуючий намагається отримати конфіденційну інформацію від потенційної жертви.*
- *Смішинг (Smishing): атаки через SMS-повідомлення, де у користувачів намагаються виманити особисті дані через текстові повідомлення.*
- *SpearPhishing або таргетований фішинг: атаки, де зловмисники висококваліфіковано атакують конкретні цілі - фізичні особи та/чи організації.*
- *Соціальна інженерія: використовує психологічні аспекти впливу на свідомість персоналу сучасних ІС. Як метод атаки, не лише спрямований на експлуатацію технічних організаційних вразливостей діючої системи захисту, але й ефективно використовує психологічні прийоми для масштабування наслідків атаки. Розгляд впливу соціальної інженерії на психологічний стан користувачів ІС та їх вразливість, є ключовим аспектом безпеки в онлайн середовищі. Зловмисники використовують такі методи, як створення терміновості, виклик емоцій, та швидкі перевтілення в авторитетність, щоб викликати потрібну реакцію у потенційної жертви.*

2.1.2 Ефективні стратегії захисту від загроз фішингу.

- *Навчання та професійна відповідальність:* завчасне передбачення фішингових атак розуміє під собою безперервне навчання користувачів сучасних ІС, розпізнавати характерні ознаки шахрайства. Тому, регулярні тренінги та інструкції персоналу, можуть значно підвищити рівень їх профільних компетенцій.
- *Використання антивірусних програм:* встановлення та регулярне оновлення антивірусних програм є ефективним заходом захисту від фішингу. Вони виявляють та блокують шкідливі віруси та веб-сайти.
- *Багатофакторна автентифікація:* використання багатофакторної автентифікації додає додатковий шар захисту, оскільки для входу необхідні два чи більше види автентифікації. Багатофакторна автентифікація дедалі стає необхідністю в умовах постійного зростання фішингових атак. Аналіз відомих інцидентів безпеки показує, що використання не лише паролів, але й інших методів ідентифікації, таких як біометричні дані чи одноразові коди, робить процес автентифікації значно більш надійним. Це зменшує ймовірність «успіху» атак та робить доступ до особистих облікових записів складнішим для зловмисників.
- *Управління паролями:* широке застосування бездротових мереж підвищує ризик несанкціонованого доступу до особистої (приватної) та/чи корпоративної інформації. В цьому сенсі одним із найважливіших заходів безпеки є коректне адміністрування паролями. Користувачі мають створювати складні та унікальні паролі для кожного облікового запису і регулярно їх змінювати.
- *Впровадження механізмів багатофакторної автентифікації:* додатково зміцнює захист доступу до особистих/корпоративних даних [3].
- *Оновлення ПЗ та операційних систем:* є ключовим аспектом безпеки. Своєчасне встановлення оновлень та патчів безпеки дозволяє виправляти виявлені уразливості та запобігати можливим атакам зловмисників.
- *Використання шифрування* даних на пристроях та під час обміну інформацією через бездротові мережі є також невід'ємною частиною захисту чутливих даних від фішингу. Шифрування забезпечує конфіденційність та цілісність інформації під час передачі її через мережі, в т.ч. незахищені.
- *Обмеження доступу до інформації.* є додатковим кроком з протидії фішингу, тому користувачі ІС повинні ретельно контролювати, кому та за яких умов

надають(делегують) доступ до своїх особистих даних та/чи службових повноважень при роботі з ПЗ та/чи мережевим устаткуванням корпоративної ІС.

Інтеграція зазначених стратегій сприятиме підвищенню поточного рівню захисту інформаційних ресурсів від загроз фішингових атак та покращити безпеку особистих та/чи конфіденційних корпоративних даних при їх зберіганні й циркуляції між користувачів в онлайнсередовищі.

2.2 Віруси та шкідливе ПЗ

Шкідливе ПЗ є важливою складовою у загальному спектрі загроз ІБ. В загальному випадку її основною метою є завдання шкоди інформаційним і апаратним ресурсам сучасних ІС. Нижче наведено перелік найбільш характерних (часто використовуваних) різновидів шкідливих програм та їх способи їх поширення:

- *Комп'ютерні черв'яки (або трояни): самостійні програми, які розповсюджуються через носії даних та/чи мережеву взаємодію без необхідності подальшого мануального втручання (супроводження) з боку їх розробника.*
- *Рекламні віруси: програми, що намагаються розповсюджувати рекламу або навіть змінюють (підмінюють) сторінки веб-сайтів.*
- *Шпигунське ПЗ: програми, які збирають конфіденційну, в тому числі, технологічну інформацію, без відома її користувачів.*

Вплив на інформаційні системи: наслідки використання шкідливого ПЗ для інформаційних систем можуть передбачати втрату конфіденційності, порушення цілісності даних та обмеження доступу до важливих ресурсів.

Протидія шкідливим програмам: використання антивірусного ПЗ, засобів міжмережевого екранування, систем виявлення вторгнень тощо. Також треба акцентувати увагу на важливості своєчасного оновлення ПЗ та вдосконалення кіберграмотності користувачів ІС.

2.3 Відмова в обслуговуванні та DDoS атаки

2.3.1 DDoS атаки

DDoS (*Distributed Denial of Service- розподілені атаки з відмовою в обслуговуванні*) є серйозною загрозою для сучасних ІС, що здатна призводити до великих збоїв у роботі веб-серверів та мережевої інфраструктури. Цей розділ присвячений аналізу різних типів DDoS атак, їх впливу та ефективним заходам для запобігання відмові в обслуговуванні. На сьогоднішній день ці атаки досі залишаються однією з найбільш поширених та руйнівних форм реалізації деструктивного впливу на функціонування сучасних ІС. Вони спрямовані на перевантаження ресурсів цільового сервера, мережі чи програми (програмного додатку), шляхом навмисного відправлення надмірного злочинного трафіку. Розгляд цієї теми є надзвичайно важливий, оскільки *DDoS* атаки можуть призвести до відмови в обслуговуванні та серйозно зашкодити бізнес та чи промисловим/технологічним процесам. За останні роки збільшилась частота та підвищилась складність реалізації *DDoS*.

Так, наприклад, зловмисники додатково використовують атаки підсилення (як своєрідний різновид забезпечення, або каталізатор цих атак) та синтез ботнетів, для максимізації впливу й наслідків основної атаки. Зокрема, атаки підсилення, такі як *DNS Amplification* та *NTP Amplification*, дозволяють помітно збільшити обсяг надлишкового (постановочного) трафіку і тим самим відчутно перевантажити мережеві з'єднання цільового об'єкту-жертви.

- *Відмова в обслуговуванні та її вплив на систему*

DDoS атаки можуть призвести до відмови в обслуговуванні, зробивши ресурси недоступними для легітимних користувачів. Це може викликати серйозні фінансові втрати, погіршення репутації компанії та втрату клієнтів.

- *Захист від DDoS атак*

Захист від *DDoS* атак вимагає комплексного підходу. В цьому сенсі важливо мати системи моніторингу трафіку, які виявлятимуть аномальні патерни, що можуть бути характерними для *DDoS* атак. Використання *CDN* (*Content Delivery Network*) може розподіляти трафік та мінімізувати вплив атак. Також, вкрай важливо мати системи фільтрації та обробки трафіку, які можуть відокремити легітимний трафік від атак.

2.3.2 Особливості реалізації атак підсилення

DNS Amplification

- *Збільшення обсягу відповідей:* атакуючі використовують *DNS*-сервери як посередників для збільшення обсягу трафіку. Вони відправляють запити до DNS-серверів з підробленими адресами цільової жертви. *DNS Amplification* базується на тому, що DNS-запити можуть бути короткими, але відповіді можуть бути значно більшими. Атакуючі використовують це, щоб збільшити обсяг зловмисного/паразитного трафіку, використовуючи ресурси легальних *DNS*-серверів.
- *Відсилання запитів у великому масштабі:* атакуючі відправляють велику кількість підроблених DNS-запитів відразу до великої кількості DNS-серверів, збільшуючи тим самим відповіді, які спрямовані на жертву атаки.

NTP Amplification

- *Використання NTP-серверів:* ці атаки використовують *Network Time Protocol (NTP)* для збільшення обсягу паразитного трафіку. Атакуючі відправляють підроблені запити відразу до великої кількості діючих *NTP*-серверів.

Провокування значної сукупності *DNS* та *NTP* серверів до одночасного формування ними відповідей на масштабні короткі злочинні запити, котрі формуються в межах *атак підсилення,* ґрунтується на тому, що такі відповіді можуть бути значно більшими чим запити, що й дозволяє атакуючим збільшити паразитний трафік.

2.3.3 Синтез та використання ботнетів (Botnet-based DDoS)

Синтез та наступне використання ботнетів у *DDoS* атаках є ефективним та небезпечним методом перевантаження мережевих ресурсів цільового об’єкта. Розглянемо деякі основні особливості цього процесу.

Синтез ботнетів DDoS:

- *Створення ботнетів:* зловмисники використовують різноманітні методи для зараження тисяч або навіть мільйонів пристроїв, перетворюючи їх на боти.
- *Координовані атаки:* дозволяє атакуючим синхронізувати дії ботів, направляючи трафік на цільовий сервер одночасно, збільшуючи таким чином вплив атаки.
- *Розподілена відмова в обслуговуванні:* ботнет *DDoS* атаки призводять до розподіленої відмови в обслуговуванні, внаслідок чого цільовий об’єкт стає недоступним для легітимних користувачів.

Особливості DDoS ботнетів:

- *Інфікування:* зараження пристроїв шляхом використання шкідливого ПЗ, експлуатації вразливостей та/або методів соціальної інженерії.
- *Приховане управління:* зловмисники використовують різноманітні методи для прихованого управління ботами, уникнення їх виявлення та блокування.
- *Збільшення ресурсів атаки:* використання ботнетів для збільшення (посилення) обсягу нелегітимного злочинного трафіку та «силового» впливу на цільовий сервер.

Заходи захисту:

- *Мережевий моніторинг: п*остійний моніторинг мережі для виявлення аномалій та надмірного трафіку, які можуть вказувати на DDoS атаку.

- *Виявлення та блокування ботів:* використання систем виявлення ботів для ідентифікації та блокування зламаних пристроїв у ботнеті.
- *Системи фільтрації трафіку:* впровадження швидкодіючих (хмарних) систем фільтрації трафіку, які блокують надмірний трафік та «відсікають» шкідливий.
- *Захист Інтернету речей (IoT):* збільшення поточного рівня безпеки пристроїв IoT, щоб унеможливити їх використання в якості ботів.

2.4 Інші типи кіберзагроз та їхні методи впливу на системи

У світі кібербезпеки існує розмаїття кіберзагроз, які відображаються у різних формах та методах впливу на ІС. Тому приділимо увагу й іншим типам загроз ІБ, зокрема атакам на безпеку мережі та застосунків, уточнюючи їх методи впливу та заходи із захисту.

Атаки на безпеку мережі:

- *Перехоплення трафіку (Man-in-the-Middle):* тип атаки, при якому зловмисники здійснюють перехоплення трафіку між взаємодіючими сторонами, що може призвести до доступу до конфіденційної інформації.
- *Атаки на DNS (Domain Name System):* методи атак на інфраструктуру *DNS* з метою спрямовування трафіку на злочинний ресурс та / чи посилення атак (*див. вище*).

Атаки на застосунки:

- *Хакерські атаки на вразливості коду (SQL Injection, XSS):* техніки використання вразливостей коду для впровадження зловмисного коду або отримання несанкціонованого доступу.
- *Атаки на автентифікацію та онлайн сесії*: методи обходу механізмів автентифікації та зловживання сесій для несанкціонованого доступу.

Вплив на інформаційні системи:

- *Втрата конфіденційності та цілісності даних:* можливі наслідки атак на безпеку мережі і додатків, зокрема, втрати конфіденційності та порушення цілісності даних.
- *Втрата доступності сервісів:* ці атаки можуть впливати на доступність інформаційних систем та відповідних онлайн послуг/сервісів.

Заходи для захисту:

- *Шифрування та/чи приховування (стеганографія) трафіку.*
- *Постійний моніторинг мережі, виявлення вторгнень та/чи припинення недекларованої мережевої активності.*

3. Аналіз вразливостей і можливих стратегії захисту

3.1 Виявлення вразливостей ІС

Виявлення вразливостей ІС, це ключовий етап в забезпеченні їхньої стійкості та захисту від кібератак. У світі, де загрози зростають щодня, ефективні методи виявлення вразливостей стають їх обов'язковою необхідністю [3]. Виявлення цих вразливостей у власних інформаційних системах – перший крок до їхнього ефективного захисту. Тому коротко розглянемо деякі з нових підходів, щодо виявлення вразливостей і розробці стратегій для їхнього негайного парирування.

Методи виявлення вразливостей:

- *Сканування портів та аналіз вразливостей:* автоматизовані засоби можуть проаналізувати «відкриті порти» та визначити потенційні точки входу для атак.
- *Статичний та динамічний аналіз коду:* виявлення вразливостей використовуваного ПЗ, через аналіз вихідного коду, дозволяє виявити вразливості, які можуть бути використані для вторгнень/атаки/витоку даних.

- *Системи виявлення Інтранет-загроз:* моніторинг внутрішнього сегменту мережі для виявлення аномальної мережевої активності та потенційних загроз безпеки.
- *Ethicalhacking та Penetrationtesting:* етичнийхакінг (пентестінг) для виявлення існуючих вразливостей з метою їх подальшого усунення (в межах аудиту ІБ).

3.2 Стратегії захисту від різних типів кіберзагроз

Багатошаровий підхід до захисту:

- *Системи виявлення та захисту:* встановлення інтегрованих систем безпеки для виявлення та блокування атак в реальному часі (*IDS/IPS/DLP/Honey Net* тощо).
- *Фільтрація трафіку:* використання систем фільтрації для блокування небезпечних пакетів трафіку на різних рівнях/сегментах мережі (*proxy, firewall, IPS* тощо).

Сегментація та ізоляція ресурсів:

- *Делегування прав доступу:* обмеження доступу до важливих ресурсів за допомогою делегування прав (*firewalls, біометричні системи, системи захисту від несанкціонованих дій, впровадження алгоритмів виконання сумісних дій та ін.*).

3.3 Заходи із забезпечення стійкості ІС до кібератак

Забезпечення стійкості ІС до кібератак – це необхідна передумова для збереження конфіденційності, цілісності та доступності даних. Сучасний кіберпростір вимагає від організацій постійно вдосконалювати свої стратегії захисту та вживати комплексних заходів для завчасного парирування існуючих загроз ІБ.

- *Захист від внутрішніх загроз.* Розробка та впровадження ефективної корпоративної політики ІБ (ПІБ), включаючи обмеження доступу та моніторинг внутрішніх користувачів, стає важливим етапом у попередженні можливих загроз зсередини.
- *Використання сучасних систем виявлення загроз.* Системи виявлення вторгнень та аномалій дозволяють вчасно виявляти та реагувати на небезпечні активності. Використання штучного інтелекту та машинного навчання (AI/LM) дозволяє автоматизувати процес виявлення навіть найскладніших кіберзагроз.
- *Захист мережі та захист умовного «периметру» безпеки.* Міжмережеві бар'єри та захист «зовнішнього» периметру безпеки, визначають умовну першу лінію захисту. Вони включають в себе використання *firewalls*, систем виявлення вторгнень (*IDS*) та засоби фільтрації трафіку (корпоративні *proxy*та*DLP*тощо) для блокування можливих атак/витоку даних на рівні мережі.
- *Управління доступом та автентифікація.* Забезпечення стійкості включає в себе також впровадження ефективної системи управління доступом та багатофакторної автентифікації. Це дозволяє обмежити доступ до чутливої інформації та ускладнює можливі атаки на облікові записи користувачів.
- *Регулярні аудити стану безпеки та оновлення ПЗ та ПІБ.* Проведення систематичних аудитів поточного стану ІБ для виявлення існуючих вразливостей та невикористаних можливостей є ключовим аспектом безпеки. Регулярне оновлення програмного та апаратного забезпечення дозволяє усувати виявлені вразливості та підтримувати систему в актуальному стані.

3.4Впливрівню технологічного розвитку у забезпеченні ІБ

Технологічний розвиток є неодмінним фактором впливу на рівень безпеки сучасних ІС. Постійний прогрес у галузі інформаційних технологій (ІТ) створює нові можливості для забезпечення ІБ та вимагає від організацій постійного адаптування до змін у кіберпросторі. Тож коротко розглянемо основні з найбільш перспективних технологій:

- *Штучний інтелект та машинне навчання (AIта LM)*

Застосування технологій *AI* та *LM* в галузі кібербезпеки дозволяє автоматизувати виявлення та аналіз аномалій в мережах. Алгоритми машинного навчання можуть швидко адаптуватися до нових типів загроз, забезпечуючи більш ефективний захист.

- *Блокчейн для забезпечення «імунітету» до змін*

Технологія блокчейн визначається своєю децентралізацією та непроникністю до змін. У сфері кібербезпеки, вона може служити основою для безпечного зберігання та обміну конфіденційної інформації, запобігаючи атакам на централізовані системи.

- *Квантова обчислення та технології віртуалізації процесів (VR)*

З розвитком квантових технологій виникає можливість у нових методах аналізу мережевого трафіку, віртуалізації процесів, каскадування обчислювальних можливостей та криптографічного захисту даних. Квантові комп'ютери можуть «зламувати» традиційні криптографічні алгоритми, тому створення нових квантово-стійких захисних методів, дедалі стає все більш актуальним завданням ІБ.

- *Інтернет речей (IoT) та кіберфізичні системи*

Розширення Інтернету речей передбачає додавання та й маніпулювання величезними обсягами додаткових даних, що потребують їх ефективного захисту. Розвиток кіберфізичних систем дозволяє об'єднати в собі фізичний та кібер- світи, вимагаючи при цьому впровадження інноваційних технологій й методів безпеки.

- *Спрощення управління безпекою через Cloud Security*

Використання сукупності хмарних та*VR*технологій дозволяє компаніям зосередитися на вдосконаленні стратегій безпеки, адже адміністрування та оновлення захисних систем може бути здійснене централізовано [4].

4. Висновки

1. На сьогоднішній день кіберзагрози створюють загрози для конфіденційності, цілісності та доступності інформації. Зростання залежності суспільства від поточного рівня розвитку й впровадження ІТ підкреслює актуальність питання захисту ІС від кіберзагроз.

2.Проведено аналіз впливу різних типів кіберзагроз на ІС та розглянуті основні стратегії щодо їх захисту. Запропоновано огляд нових тенденцій у кібербезпеці і деяких інноваційних підходів з питань захисту від нових загроз. Підкреслено наявність нерозривного взаємозв'язку питань технологічного розвитку й фактичного стану можливостей із ІБ.

3. Підкреслено необхідність постійного оновлення й адаптації діючих стратегій захисту до зростаючої складності кіберзагроз. Звернено увагу, що захист ІС вимагає одночасного поєднання впровадження інноваційних технологій, глибокого розуміння сучасних тенденцій кібербезпеки та глобальної співпраці для ефективної протидії актуальним кіберзагрозам.

References

[1] Jon Erickson (2010). "Hacking: The Art of Exploitation" ISBN-13: 978-1-59327-144-2

[2] Edward Amoroso. (2010). "Cybersecurity: Protecting Critical Infrastructures from Cyber Attack and Cyber Warfare" ISBN-13: 978-1-4822-3923-2

[3] P.W. Singer та Allan Friedman (2014). "Cybersecurity and Cyberwar: What Everyone Needs to Know" ISBN: 978-0-19-991811-9

[4] International Journal of Computer Science and Information Technologies "Cybersecurity: A Journal of Technology, Society and Policy". ISSN:0975-9646

Submitted November 9, 2023; Revised December 16, 2023; Accepted December 24, 2023

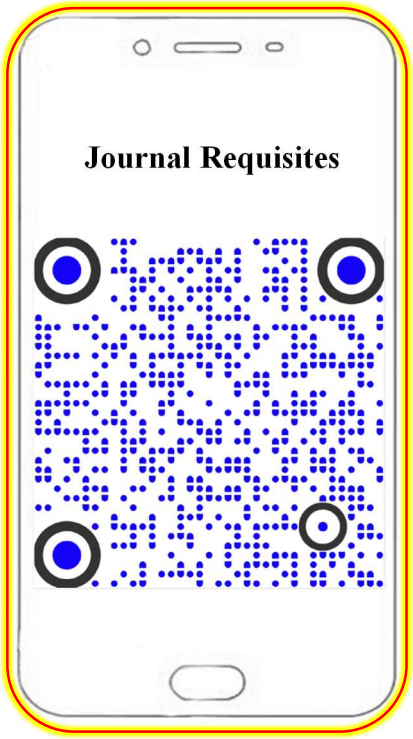
Authors:
Osadchyi Yevhenii, CSD Student, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine.
E-mail: xa12850357@student.karazin.ua

Yesina Maryna, Ph.D., Associate Professor, department of security of information systems and technologies, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine; research associate-consultant of JSC "IIT", Kharkiv, Ukraine.
E-mail: m.v.yesina@karazin.ua
ORCID: https://orcid.org/0000-0002-1252-7606
Victor Onoprienko, Ph.D., CEO of JSC "IIT", Kharkiv, Ukraine.
E-mail: v25258@gmail.com

The influence of different forms of cyber threats on the stability of information systems: analysis and protection strategies

Abstract. This work is dedicated to the further investigation of cybersecurity issues in the context of the ongoing development of the current information industry. Starting with an overview of various forms of cyber threats, the article examines the analysis of their impact on the privacy, integrity and availability of information. The critical dependence of modern society on information technology makes the topic of protection against cyber threats extremely relevant. This work offers an in-depth analysis of the growth in the number and complexity of cyber threats, which requires constant improvement and updating of protection strategies against them. An important stage of coverage of the topic is the analysis of the impact of various forms of cyber threats on information systems. The main types of phishing and social engineering are considered, as well as the consequences of exposure to viruses, Trojans and other malicious programs. A detailed review of these aspects allows us to highlight the key issues and dangers that arise in the context of cyber threats. Also, the article contains materials devoted to various protection strategies. It examines effective strategies for protecting information systems, including identifying vulnerabilities, using multi-factor authentication, and measures to ensure resilience. The general conclusions of this work summarize the need for constant updating and adaptation of protection strategies in relation to the growing complexity of cyber threats in the world of rapid technological development. In general, this work is another step in understanding the essence of the challenges associated with the issue of ensuring cyber security in the modern information society.

Keywords: *Impact, Cyber Threat, Analysis And Protection, Resilience Of Information Systems, Protection Strategies.*



Publishing, cover design: V.N. Karazin Kharkiv National University, 2023

Наукове видання

КОМП’ЮТЕРНІ НАУКИ ТА КІБЕРБЕЗПЕКА

Випуск2(24) 2023

Міжнародний електронний науково-теоретичний журнал

Англійською, українською, *та ін. мовами*

Комп’ютерне верстання –Єсіна М.В.,Федоренко В.В.

61022, Харків, майдан Свободи, 6
Харківський національний університет імені В.Н. Каразіна

V. N. Karazin Kharkiv National University Publishing

CS&CS

2023