Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

Факультет навчально-науковий центр заочної форми навчання
(повна назва)

Кафедра електронних обчислювальних машин
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

Рівень вищої освіти перший (бакалаврський)

Інструмент виявлення стеганографії в зображеннях на основі сигнатур

(тема)

Виконав:
здобувач 4 року навчання,
групи КІУКІз-21-1
Дмитро КЛЮЧНИК
(власне ім'я, прізвище)

Спеціальність
123 «Комп'ютерна інженерія»
(код і повна назва спеціальності)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма
Комп'ютерна інженерія
(повна назва освітньої програми)

Керівник: ас. Дмитро ДЯЧЕНКО
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ЕОМ _____ Андрій КОВАЛЕНКО
(підпис) (власне ім'я, прізвище)

2025 р.

Харківський національний університет радіоелектроніки

Факультет навчально-науковий центр заочної форми навчання

Кафедра електронних обчислювальних машин

Рівень вищої освіти перший (бакалаврський)

Спеціальність 123 «Комп'ютерна інженерія»
(код і повна назва)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Комп'ютерна інженерія
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
"_____"_____ 20___ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Ключнику Дмитру Павловичу
(прізвище, ім'я, по батькові)

1. Тема роботи Інструмент виявлення стеганографії в зображеннях на основі сигнатур

затверджена наказом по університету від "05" травня 2025 р. № 72 СТз

2. Термін подання здобувачем роботи до екзаменаційної комісії 16 червня 2025 р.

3. Вхідні дані до роботи
стеганографія
застосунок для мобільних пристроїв
PixelKnot
LSB

4. Перелік питань, що потрібно опрацювати у роботі
Теоретичні основи моніторингу мережного трафіка
Огляд сучасних програмних засобів моніторингу мережного трафіка
Порівняльний аналіз програмних засобів моніторингу
Розробка програмних засобів моніторингу

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій 11 слайдів

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1	Отримання завдання та аналіз літератури	05.05.2025–07.05.2025	
2	Огляд існуючих аналогів	08.05.2025–13.05.2025	
3	Вибір методів для реалізації	13.05.2025–17.05.2025	
4	Вибір програмних засобів	18.05.2025–23.05.2025	
5	Програмна реалізація	24.05.2025–26.05.2025	
6	Аналіз отриманих результатів	27.05.2025–28.05.2025	
7	Оформлення записки	28.05.2025–16.06.2025	

Дата видачі завдання "05" травня 2025 р.

Здобувач _____
(підпис)

Керівник роботи _____ ас. Дмитро ДЯЧЕНКО
(підпис) (посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 55 с., 7 рис., 1 дод., 8 джерел.

СТЕГАНОГРАФІЯ, ДЕТЕКЦІЯ СТЕГАНОГРАФІЇ, МОБІЛЬНІ ДОДАТКИ, ANDROID-ЗАСТОСУНКИ, ПРИХОВАНІ ДАНІ, СИМВОЛЬНЕ ВИКОНАННЯ, СИГНАТУРНИЙ АНАЛІЗ, АВТОМАТИЗОВАНА ДЕТЕКЦІЯ, ОБРОБКА ЗОБРАЖЕНЬ, ПРОГРАМНИЙ АНАЛІЗ, ЦИФРОВА КРИМІНАЛІСТИКА, ІНСТРУМЕНТАРІЙ ДЕТЕКЦІЇ.

Метою кваліфікаційної роботи є аналіз та розробка програмних засобів виявлення стеганографії в застосунках для мобільних пристроїв.

У ході виконання кваліфікаційної роботи виконано детальний аналіз існуючих програмних рішень для детекції стеганографії в мобільних платформах. Дослідження включало огляд інструментів для вбудовування прихованої інформації у зображення, вивчення сучасних методів виявлення стеганографічного контенту та аналіз технологій розпізнавання стего-елементів через сигнатурні маркери. Розроблено програмний інструментарій для виявлення стеганографічних додатків на мобільних пристроях. Імплементовано автоматизовану систему детекції оригінальних стеганографічних зображень у Android-застосунках. Для абстрагування операційної логіки програм при обробці зображень застосовано символьне виконання з представленням через структуровану ієрархію математичних виразів.

ABSTRACT

Bachelor's thesis: 55 pages, 7 figures, 1 appendices, 8 sources.

STEGANOGRAPHY, STEGANOGRAPHY DETECTION, MOBILE APPLICATIONS, ANDROID APPS, HIDDEN DATA, SYMBOLIC EXECUTION, SIGNATURE ANALYSIS, AUTOMATED DETECTION, IMAGE PROCESSING, PROGRAM ANALYSIS, DIGITAL FORENSICS, DETECTION TOOLKIT.

The major goal of this thesis is to analyze and develop software tools for detecting steganography within mobile applications.

In order to a comprehensive analysis of existing software solutions for steganography detection on mobile platforms was conducted. The study included a review of tools used to embed hidden information within images, an examination of current methods for identifying steganographic content, and an analysis of technologies for recognizing stego-elements based on signature markers. A specialized software toolkit was developed to identify steganographic applications on mobile devices. An automated detection system was implemented to locate original steganographic images embedded in Android applications. To abstract the operational logic of image-handling software, symbolic execution was utilized, representing application behavior through a structured hierarchy of mathematical expressions.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ 7
ВСТУП 8
1 ТЕОРЕТИЧНІ ОСНОВИ СТЕГАНОГРАФІЇ ТА МЕТОДІВ ЇЇ ВИЯВЛЕННЯ 11
1.1 Загальні поняття стеганографії 11
1.2 Цифрова стеганографія в зображеннях 13
1.3 Стеганоаналіз та методи виявлення прихованої інформації 16
2 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ ТА ПОСТАНОВКА ЗАВДАННЯ 22
2.1 Огляд існуючих інструментів виявлення стеганографії 22
2.2 Порівняльний аналіз методів виявлення 26
2.3 Виявлення недоліків існуючих рішень 30
3 РОЗРОБКА ІНСТРУМЕНТУ ВИЯВЛЕННЯ СТЕГАНОГРАФІЇ 34
3.1 Архітектура системи 34
3.2 База сигнатур стеганографічних методів 36
3.3 Алгоритм виявлення стеганографії 39
3.4 Модуль аналізу метаданих 43
3.5 Тестування та анліз результатів 44
ВИСНОВКИ 47
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ 48
ДОДАТОК А Графічний матеріал кваліфікаційної роботи 49

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

AI – штучний інтелект

APK – Android Package Kit

DEX – Dalvik Executable

DF – цифрова криміналістика

EXIF – Exchangeable Image File Format

GUI – графічний інтерфейс користувача

JPEG – Joint Photographic Experts Group

MOBSEC – безпека мобільних застосунків

MOBAPP – мобільний застосунок

OSINT – розвідка з відкритих джерел

PNG – Portable Network Graphics

ПЗ – програмне забезпечення

SDK – Software Development Kit

SoS – рівень обслуговування

STEG – стеганографія

STEG-DET – система виявлення стеганографії

UI – користувацький інтерфейс

XML – Extensible Markup Language

ВСТУП

Сучасний цифровий світ характеризується інтенсивним зростанням обсягів інформації, що передається та зберігається в електронному вигляді, а також постійним вдосконаленням технологій обробки мультимедійного контенту. В умовах глобальної цифровізації та широкого розповсюдження інтернет-технологій особливої актуальності набувають питання інформаційної безпеки, захисту персональних даних та протидії кіберзагрозам. Серед різноманітних методів несанкціонованого використання інформаційних ресурсів особливе місце займає стеганографія – наука про приховування інформації всередині інших, на перший погляд нейтральних, об'єктів.

Стеганографічні технології, маючи багатовікову історію розвитку, в епоху цифрових технологій отримали принципово нові можливості та сфери застосування. Цифрові зображення, завдяки своїй поширеності в інтернет-просторі та значній кількості надлишкової інформації, стали одним з найпопулярніших носіїв для приховування даних. Легкість створення, передачі та зберігання графічних файлів, поєднана з відносною простотою реалізації базових алгоритмів стеганографії, робить цифрові зображення привабливим інструментом як для легітимного захисту інформації, так і для зловмисної діяльності.

Водночас розвиток стеганографічних методів супроводжується паралельним удосконаленням технологій стеганоаналізу – науки про виявлення прихованої інформації. Ця постійна конкуренція між методами приховування та виявлення створює динамічне середовище, де кожна сторона прагне перевершити досягнення опонента. Сучасні алгоритми стеганографії стають дедалі більш витонченими, використовуючи складні математичні моделі, адаптивні стратегії вбудовування та міждисциплінарні підходи, що значно ускладнює процес їх виявлення традиційними методами.

Практичне значення ефективних інструментів виявлення стеганографії важко переоцінити в контексті забезпечення кібербезпеки. Правоохоронні органи потребують надійних засобів для виявлення прихованої інформації в рамках розслідування кіберзлочинів та протидії тероризму. Корпоративні системи безпеки мають забезпечувати захист від витоку конфіденційної інформації через приховані канали передачі даних. Освітні установи та дослідницькі центри потребують інструментів для навчання та наукової роботи в галузі інформаційної безпеки.

Аналіз сучасного стану технологій стеганоаналізу виявляє кілька критичних проблем, що обмежують ефективність існуючих рішень. Більшість доступних інструментів розроблені для виявлення застарілих алгоритмів стеганографії та демонструють низьку ефективність при роботі з сучасними адаптивними методами. Обмежена підтримка нових форматів зображень створює прогалини в системах безпеки, які можуть бути використані зловмисниками. Складність налаштування та використання багатьох професійних інструментів обмежує їх доступність для широкого кола користувачів.

Сигнатурний підхід до виявлення стеганографії представляє перспективний напрямок розвитку технологій стеганоаналізу, що поєднує високу швидкість обробки з можливістю точної ідентифікації конкретних стеганографічних інструментів та алгоритмів. Цей метод базується на виявленні характеристичних "відбитків пальців", що залишають різні стеганографічні програми в процесі вбудовування інформації. Ефективність сигнатурного підходу доведена в суміжних галузях, зокрема в системах виявлення комп'ютерних вірусів, де він забезпечує швидку та точну ідентифікацію відомих загроз.

Метою роботи є аналіз та розробка програмних засобів виявлення стеганографії в застосунках для мобільних пристроїв.

Завдання:

- аналіз існуючих програмних засобів створення стеганографічних

зображень;

- аналіз існуючих програмних засобів виявлення стеганографії в зображеннях;
- аналіз методу виявлення стегозображень на основі сигнатур;
- розробка інструменту виявлення стеганографічних додатків для мобільних пристроїв.

1 ТЕОРЕТИЧНІ ОСНОВИ СТЕГАНОГРАФІЇ ТА МЕТОДІВ ЇЇ ВИЯВЛЕННЯ

1.1 Загальні поняття стеганографії

Стеганографія як наука про приховування інформації має багатовікову історію, сягаючи корінням до давньогрецьких часів, коли термін було вперше введено Йоганнесом Тритеміусом у його праці "Стеганографія" 1499 року. Етимологічно слово походить від грецьких слів "steganos", що означає "прихований" або "покритий", та "graphein" - "писати", що в сукупності утворює поняття "прихованого письма". У сучасному розумінні стеганографія представляє собою мистецтво та науку приховування повідомлень або інформації всередині інших неприховуваних повідомлень чи фізичних об'єктів таким чином, щоб сам факт існування прихованого повідомлення не було помітно.

Розвиток стеганографії протягом історії демонструє еволюцію від простих фізичних методів до складних цифрових технологій. У давнину використовувалися такі методи, як написання повідомлень на дерев'яних табличках з подальшим покриттям їх воском з нейтральним текстом, татуювання повідомлень на бритій голові посланця з подальшим відрощуванням волосся, або використання невидимого чорнила з лимонного соку. Середньовічний період характеризувався розробкою більш витончених методів, включаючи акростичні вірші, де перші літери кожного рядка складали приховане повідомлення, та використання мікрографії для створення надзвичайно дрібного письма.

Принципові основи сучасної стеганографії базуються на концепції непомітності приховування інформації. Фундаментальним принципом є те, що носій інформації після вбудовування прихованих даних повинен залишатися візуально та статистично неможливим до відрізнення від оригінального носія. Це досягається шляхом мінімізації модифікацій, що

вносяться в носій, та забезпечення того, що ці модифікації знаходяться в межах природних варіацій даних носія.

Класифікація стеганографічних методів може здійснюватися за різними критеріями. За типом носія розрізняють текстову стеганографію, де повідомлення приховуються в текстових документах шляхом модифікації форматування, використання синонімів або зміни структури речень; графічну стеганографію, що працює із статичними зображеннями; аудіостеганографію для звукових файлів; відеостеганографію для відеоматеріалів; та мережеву стеганографію, що використовує особливості мережевих протоколів для приховування даних.

За принципом роботи методи поділяються на просторові, що модифікують дані безпосередньо в просторовій області носія, та частотні, які працюють у трансформованій області після застосування математичних перетворень типу дискретного косинусного перетворення або вейвлет-перетворення. Адаптивні методи враховують локальні характеристики носія для оптимального розміщення прихованої інформації, тоді як неадаптивні застосовують однакову стратегію для всього носія.

Відмінності між стеганографією та криптографією полягають в їхніх цілях та підходах до забезпечення безпеки інформації. Криптографія зосереджується на перетворенні даних таким чином, щоб вони стали нечитабельними без відповідного ключа дешифрування, при цьому сам факт існування зашифрованого повідомлення є очевидним. Стеганографія, навпаки, прагне повністю приховати існування повідомлення, роблячи його непомітним для стороннього спостерігача. Ці підходи можуть використовуватися комплементарно, коли спочатку повідомлення шифрується, а потім приховується за допомогою стеганографічних методів, забезпечуючи подвійний рівень захисту.

1.2 Цифрова стеганографія в зображеннях

Цифрова стеганографія в зображеннях (рисунок 1.1) представляє собою один з найпоширеніших та найефективніших підходів до приховування інформації в сучасному цифровому світі. Популярність цього методу обумовлена кількома факторами: широким розповсюдженням цифрових зображень в інтернеті та повсякденному житті, великою кількістю надлишкової інформації в графічних файлах, яка може бути використана для вбудовування даних, та відносною простотою реалізації базових алгоритмів.

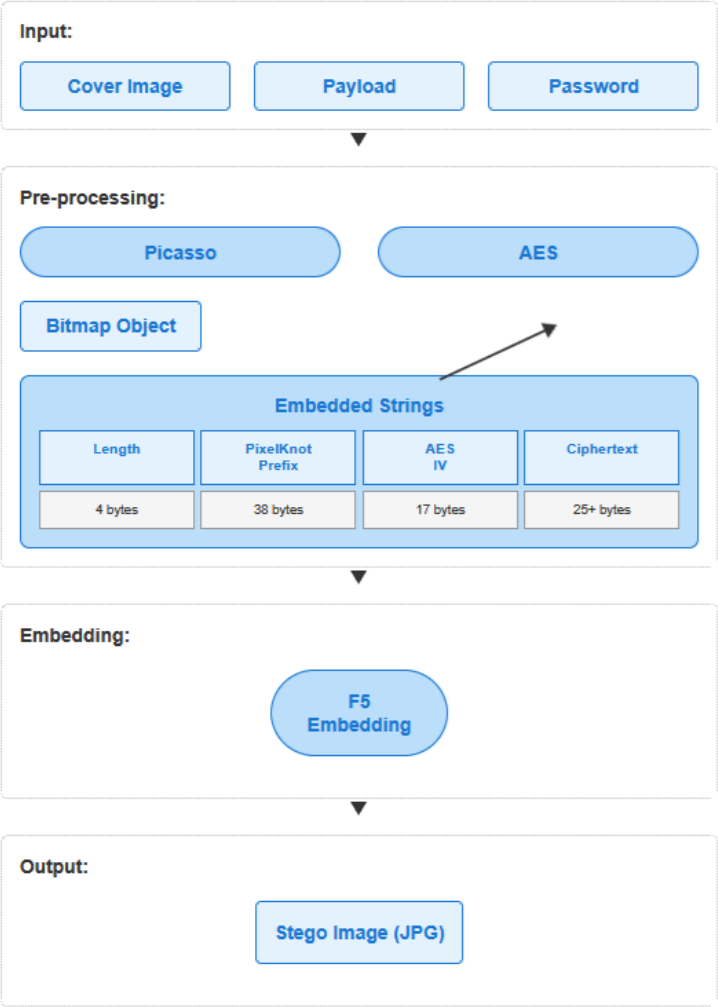

Рисунок 1.1 – Цифрова стеганографія в зображеннях

Структура цифрових зображень створює фундаментальні можливості для стеганографії. Растрові зображення складаються з пікселів, кожен з яких

представлений числовими значеннями, що описують інтенсивність кольору. У випадку чорно-білих зображень кожен піксель характеризується одним значенням яскравості, зазвичай в діапазоні від 0 до 255 для 8-бітного представлення. Кольорові зображення використовують різні колірні моделі, найпоширенішою з яких є RGB, де кожен піксель описується трьома значеннями для червоного, зеленого та синього каналів.

Важливою характеристикою цифрових зображень є їхня надлишковість, яка проявляється в кореляції між сусідніми пікселями та в тому факті, що людське око не здатне сприймати дрібні зміни в інтенсивності кольору. Ця особливість зорового сприйняття людини створює можливості для модифікації даних зображення без помітного погіршення візуальної якості, що і використовується в стеганографічних алгоритмах.

Методи вбудовування інформації в зображення можна категоризувати за областю, в якій відбувається модифікація даних. Просторові методи працюють безпосередньо з пікселями зображення, модифікуючи їхні значення для вбудовування інформації. Ці методи характеризуються простотою реалізації та високою швидкістю роботи, проте можуть бути вразливими до простих статистичних атак. Частотні методи спочатку трансформують зображення в частотну область за допомогою математичних перетворень, модифікують коефіцієнти перетворення, а потім виконують зворотне перетворення для отримання стего-зображення.

LSB-метод представляє собою найпростіший та найпоширеніший підхід до стеганографії в зображеннях. Принцип методу базується на заміні найменш значущих бітів пікселів зображення бітами прихованого повідомлення. Оскільки зміна найменш значущого біта призводить до мінімальної зміни значення пікселя (максимум на одиницю), така модифікація зазвичай непомітна для людського ока. У простій реалізації LSB-методу кожен біт прихованого повідомлення записується в найменш значущий біт відповідного байта зображення, що дозволяє вбудувати один біт інформації на кожен канал кольору кожного пікселя

Розширені варіанти LSB-методу включають використання кількох молодших бітів одночасно, що збільшує пропускну здатність, але підвищує ймовірність виявлення. Адаптивні LSB-методи вибирають піксели для модифікації на основі локальних характеристик зображення, надаючи перевагу областям з високою текстурою або шумом, де зміни будуть менш помітними. Рандомізовані підходи використовують псевдовипадкові послідовності для вибору позицій пікселів, що ускладнює виявлення закономірностей в розподілі прихованої інформації.

DCT-методи базуються на дискретному косинусному перетворенні, яке широко використовується в алгоритмах стиснення зображень, зокрема в форматі JPEG. Перетворення розділяє зображення на блоки розміром 8×8 пікселів та застосовує двовимірне DCT до кожного блоку, в результаті чого отримуються коефіцієнти, що представляють частотні компоненти блоку. Низькочастотні коефіцієнти, розташовані в лівому верхньому куті матриці DCT, містять основну візуальну інформацію, тоді як високочастотні коефіцієнти відповідають за деталі та можуть бути модифіковані з меншим впливом на візуальну якість.

Стеганографічні алгоритми на основі DCT зазвичай модифікують середньочастотні коефіцієнти, оскільки низькочастотні коефіцієнти критично важливі для візуального сприйняття, а високочастотні часто втрачаються під час стиснення. Популярним підходом є квантування коефіцієнтів DCT з подальшою модифікацією їхніх значень відповідно до біт прихованого повідомлення. Такі методи забезпечують кращу стійкість до стиснення порівняно з просторовими методами, оскільки працюють в тій же області, що й алгоритми стиснення.

Методи на основі вейвлет-перетворення використовують математичний апарат вейвлетів для розкладання зображення на компоненти різних масштабів та орієнтацій. Дискретне вейвлет-перетворення створює ієрархічне представлення зображення, де кожен рівень декомпозиції містить апроксимуючі коефіцієнти, що представляють загальну структуру

зображення, та детальні коефіцієнти, що відповідають за горизонтальні, вертикальні та діагональні деталі.

Стеганографічні методи на основі вейвлетів модифікують детальні коефіцієнти на різних рівнях декомпозиції, враховуючи їхню важливість для візуального сприйняття. Коефіцієнти з більшими абсолютними значеннями зазвичай більш важливі для якості зображення, тому алгоритми часто використовують адаптивні стратегії, що враховують локальну важливість коефіцієнтів. Перевагою вейвлет-методів є їхня здатність забезпечувати багатомасштабний аналіз та кращу локалізацію змін порівняно з DCT-методами.

1.3 Стеганоаналіз та методи виявлення прихованої інформації

Стеганоаналіз представляє собою науку та мистецтво виявлення прихованої інформації в об'єктах-носіях, де така інформація може бути вбудована за допомогою стеганографічних методів. Розвиток стеганоаналізу йде пліч-о-пліч з еволюцією стеганографічних технік, створюючи постійну конкуренцію між методами приховування та виявлення інформації. Фундаментальний принцип стеганоаналізу базується на тому, що будь-яке вбудовування інформації в носій неминуче залишає сліди або створює аномалії, які можуть бути виявлені за допомогою відповідних аналітичних методів

Основні принципи стеганоаналізу ґрунтуються на статистичному аналізі властивостей досліджуваних об'єктів. Природні зображення мають певні статистичні характеристики, які змінюються при вбудовуванні стеганографічної інформації. Ці зміни можуть проявлятися в різних формах: порушення природних кореляцій між пікселями, зміна гістограм розподілу значень, поява артефактів у частотній області, або порушення природних закономірностей текстури зображення

Класифікація методів стеганоаналізу може здійснюватися за кількома

критеріями. За рівнем знань про стеганографічний метод розрізняють специфічний стеганоаналіз, що розроблений спеціально для виявлення конкретного стеганографічного алгоритму та використовує знання про його особливості, та універсальний стеганоаналіз, що прагне виявити факт наявності прихованої інформації незалежно від використаного методу вбудовування.

Статистичні методи виявлення складають основу сучасного стеганоаналізу та базуються на аналізі статистичних властивостей зображень. Гістограмний аналіз вивчає розподіл значень пікселів або коефіцієнтів перетворення в зображенні. LSB-стеганографія, наприклад, може призводити до характерних змін в гістограмі, таких як вирівнювання частот парних та непарних значень пікселів. Хі-квадрат тест є одним з класичних статистичних методів, що використовується для виявлення LSB-стеганографії шляхом аналізу розподілу пар значень пікселів.

Аналіз кореляційних властивостей досліджує взаємозв'язки між сусідніми пікселями або коефіцієнтами. Природні зображення характеризуються високою кореляцією між сусідніми пікселями, оскільки об'єкти в реальному світі мають гладкі поверхні та поступові переходи кольорів. Стеганографічне вбудовування може порушувати ці кореляції, особливо у випадку просторових методів типу LSB. Статистики вищих порядків аналізують моменти розподілу значень пікселів, які можуть бути чутливими до стеганографічних модифікацій.

Частотний аналіз застосовується для виявлення стеганографії в трансформованій області. Методи на основі DCT або вейвлет-перетворення залишають специфічні сліди в частотній області, які можуть бути виявлені шляхом аналізу коефіцієнтів перетворення. Аналіз розподілу коефіцієнтів DCT може виявити аномалії, створені стеганографічним вбудовуванням, особливо в середньочастотних компонентах, де зазвичай розміщується прихована інформація

Візуальні методи аналізу використовують особливості людського

зорового сприйняття для виявлення артефактів, створених стеганографічним вбудовуванням. Різницевий аналіз порівнює підозріле зображення з його версією після застосування фільтрів або перетворень, що можуть виявити або підкреслити стеганографічні модифікації. Градієнтний аналіз досліджує зміни інтенсивності пікселів та може виявити неприродні переходи, створені вбудовуванням інформації

Спектральний аналіз використовує перетворення Фур'є для дослідження частотного складу зображення. Стеганографічне вбудовування може створювати характерні періодичні структури або аномалії в спектрі, особливо якщо використовується регулярний паттерн розміщення прихованих даних. Фазовий аналіз досліджує фазові характеристики спектра, які можуть бути чутливими до певних типів стеганографічних модифікацій.

Методи на основі машинного навчання представляють сучасний підхід до стеганоаналізу, що використовує можливості автоматичного виявлення складних закономірностей в даних. Контрольовані методи навчання потребують навчальної вибірки, що містить як "чисті" зображення, так і зображення з вбудованою інформацією. Алгоритми машинного навчання автоматично вилучають ознаки, що дозволяють розрізняти ці два класи зображень.

Нейронні мережі, зокрема згорткові нейронні мережі, показують високу ефективність в задачах стеганоаналізу завдяки їхній здатності автоматично вивчати ієрархічні представлення даних. Архітектури нейронних мереж для стеганоаналізу зазвичай включають спеціалізовані шари для попередньої обробки зображень, такі як фільтри високих частот, що підкреслюють стеганографічні артефакти.

Метричне навчання та методи глибокого навчання дозволяють створювати універсальні детектори, здатні виявляти різні типи стеганографії без попереднього знання про конкретний алгоритм вбудовування. Ансамблеві методи комбінують результати кількох різних детекторів для підвищення надійності та точності виявлення.

1.4 Сигнатурний аналіз в стеганографії

Сигнатурний підхід до виявлення стеганографії представляє собою метод ідентифікації прихованої інформації на основі характерних ознак або "відбитків пальців", що залишаються різними стеганографічними інструментами та алгоритмами в процесі вбудовування даних. Цей підхід аналогічний до методів виявлення комп'ютерних вірусів, де антивірусні програми використовують бази сигнатур для ідентифікації відомих загроз. У контексті стеганографії сигнатури представляють собою унікальні характеристики, що дозволяють ідентифікувати конкретний стеганографічний інструмент або алгоритм, використаний для створення стего-об'єкта.

Поняття сигнатури в стеганографії охоплює широкий спектр характеристик, що можуть служити для ідентифікації. Ці характеристики включають специфічні паттерни розміщення прихованих даних, особливості алгоритмів шифрування ключів, характерні зміни статистичних властивостей носія, та навіть метадані, що додаються стеганографічними програмами. Кожен стеганографічний інструмент має свої унікальні особливості реалізації, які проявляються в специфічних модифікаціях носія інформації.

Формування сигнатур відбувається шляхом детального аналізу роботи конкретних стеганографічних алгоритмів та інструментів. Процес включає вивчення вихідного коду програм (якщо він доступний), аналіз алгоритмів вбудовування, дослідження створених стего-об'єктів та ідентифікацію характерних ознак. Важливим аспектом є створення контрольованих умов для генерації зразків стего-об'єктів з відомими параметрами, що дозволяє точно ідентифікувати зміни, внесені конкретним алгоритмом.

Типи сигнатур стеганографічних алгоритмів можна класифікувати за різними критеріями. Структурні сигнатури базуються на особливостях організації даних в стего-об'єкті. Наприклад, деякі програми можуть додавати специфічні заголовки або маркери в зображення, використовувати

фіксовані розміри блоків для обробки, або мати характерні паттерни заповнення невикористаного простору. Алгоритмічні сигнатури відображають специфіку математичних операцій, використовуваних для вбудовування. Це може включати особливості генерації псевдовипадкових послідовностей, специфічні методи квантування коефіцієнтів, або характерні способи адаптації до локальних властивостей носія.

Статистичні сигнатури проявляються в характерних змінах статистичних властивостей стего-об'єктів. Різні алгоритми по-різному впливають на гістограми розподілу значень, кореляційні функції, моменти вищих порядків та інші статистичні характеристики. Ці зміни можуть бути настільки специфічними, що дозволяють не тільки виявити факт наявності прихованої інформації, але й ідентифікувати конкретний алгоритм або навіть версію програми, використану для вбудовування.

Часові сигнатури пов'язані з особливостями процесу створення стего-об'єктів. Деякі програми можуть залишати мітки часу в метаданих файлів, використовувати часові послідовності як частину алгоритму генерації ключів, або мати характерні часові характеристики обробки різних типів контенту. Версійні сигнатури дозволяють розрізняти різні версії одного й того ж стеганографічного інструменту, оскільки кожна версія може мати свої особливості реалізації або виправлення помилок попередніх версій.

Методи вилучення сигнатур включають як ручний аналіз, так і автоматизовані підходи. Ручний аналіз передбачає детальне дослідження стего-об'єктів експертами з використанням спеціалізованих інструментів для аналізу структури файлів, статистичних властивостей та метаданих. Автоматизовані методи використовують алгоритми машинного навчання для автоматичного виявлення характерних ознак шляхом аналізу великих наборів зразкі

Переваги сигнатурного підходу включають високу точність ідентифікації відомих стеганографічних інструментів, швидкість перевірки завдяки простоті порівняння з базою сигнатур, та можливість не тільки

виявлення факту наявності прихованої інформації, але й ідентифікації конкретного методу вбудовування. Це дозволяє застосовувати специфічні методи вилучення прихованих даних та отримувати додаткову інформацію про джерело та методи створення стего-об'єкта

Недоліки сигнатурного підходу пов'язані з його залежністю від попередньо відомих алгоритмів та інструментів. Нові або модифіковані стеганографічні програми можуть не мати відповідних сигнатур в базі, що робить їх невиявленими. Крім того, досвідчені зловмисники можуть навмисно модифікувати стандартні інструменти для уникнення виявлення або використовувати власні алгоритми, для яких сигнатури ще не створені.

Проблема адаптивності сигнатурного підходу вирішується через постійне оновлення баз сигнатур, створення узагальнених сигнатур, що можуть виявляти класи подібних алгоритмів, та комбінування сигнатурного аналізу з іншими методами стеганоаналізу. Гібридні підходи використовують сигнатурний аналіз як перший етап швидкої перевірки, після чого застосовуються більш складні статистичні або машинні методи для підтвердження результатів або виявлення невідомих типів стеганографії.

2 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ ТА ПОСТАНОВКА ЗАВДАННЯ

2.1 Огляд існуючих інструментів виявлення стеганографії

Сучасний ринок програмного забезпечення для виявлення стеганографії характеризується наявністю різноманітних рішень (рисунки 2.1, 2.2), що відрізняються за функціональністю, методами виявлення, підтримуваними форматами та цільову аудиторією. Аналіз існуючих інструментів дозволяє визначити поточний стан технологій стеганоаналізу, ідентифікувати їхні сильні та слабкі сторони, а також окреслити напрямки для подальшого розвитку.

Комерційні рішення представляють сегмент професійних інструментів, призначених для корпоративного використання та правоохоронних органів. Steganalysis Detection Suite є одним з провідних комерційних продуктів, що пропонує комплексний підхід до виявлення прихованої інформації в різних типах медіафайлів. Система використовує комбінацію сигнатурного аналізу та статистичних методів для виявлення широкого спектра стеганографічних алгоритмів. Архітектура програми побудована за модульним принципом, що дозволяє розширювати функціональність шляхом додавання нових модулів виявлення.

Функціональні можливості Steganalysis Detection Suite включають автоматичне сканування директорій з медіафайлами, пакетну обробку великих обсягів даних, створення детальних звітів з результатами аналізу, таінтеграцію з системами управління інцидентами безпеки. Програма підтримує основні графічні формати включаючи JPEG, PNG, BMP, TIFF та GIF, а також має можливості для аналізу аудіо та відеофайлів. Статистичний движок системи реалізує понад тридцять різних алгоритмів стеганоаналізу, включаючи хі-квадрат тести, аналіз гістограм, кореляційний аналіз та методи на основі машинного навчання.

OutGuess Detection представляє собою спеціалізований інструмент, розроблений специфічно для виявлення стеганографії, створеної за допомогою алгоритму OutGuess. Цей інструмент демонструє підхід до створення високоспеціалізованих детекторів, що забезпечують максимальну точність виявлення конкретних типів стеганографії. Алгоритм виявлення базується на статистичному аналізі розподілу коефіцієнтів DCT в JPEG-зображеннях, оскільки OutGuess модифікує ці коефіцієнти характерним чином.

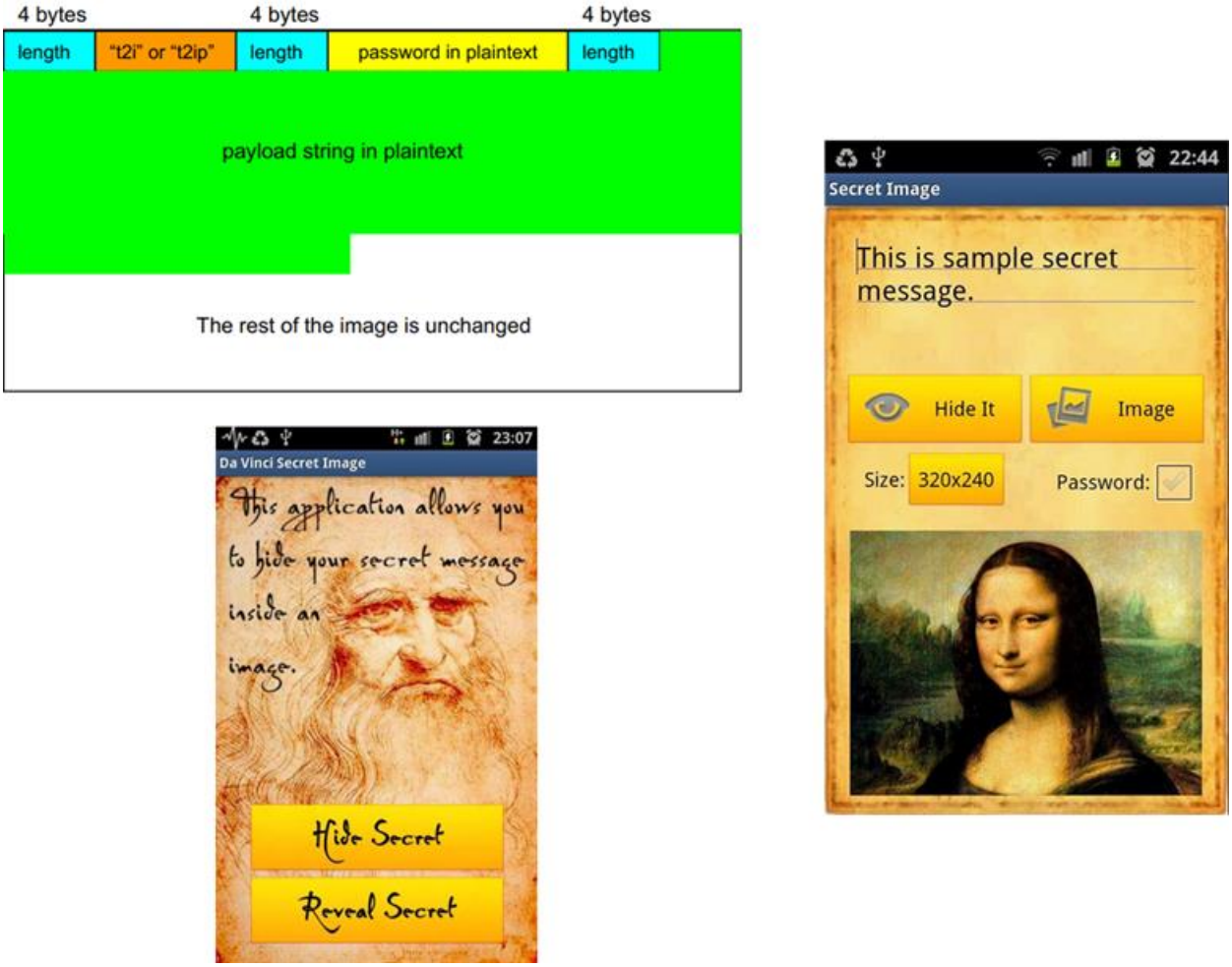

Рисунок 2.1 – Існуючі рішення

Технічна реалізація OutGuess Detection включає попередню обробку зображень для виділення блоків DCT, статистичний аналіз розподілу коефіцієнтів з використанням критеріїв максимальної правдоподібності, та

класифікацію результатів на основі порогових значень. Програма демонструє високу точність виявлення OutGuess-стеганографії з мінімальним рівнем помилкових спрацьовувань, проте її функціональність обмежена виявленням тільки цього конкретного типу алгоритмів

Відкриті рішення представляють альтернативний підхід до розробки інструментів стеганоаналізу, забезпечуючи доступність вихідного коду та можливість модифікації під специфічні потреби користувачів. StegSecret є одним з найпопулярніших відкритих інструментів, що поєднує функції створення та виявлення стеганографії. Архітектура програми побудована на основі плагінної системи, що дозволяє легко додавати підтримку нових алгоритмів стеганографії та методів їх виявлення.

Функціональність StegSecret включає підтримку множинних алгоритмів стеганографії включаючи LSB, DCT-based методи, та алгоритми на основі теорії інформації. Система виявлення використовує комбінацію статистичних тестів та евристичних правил для ідентифікації потенційної стеганографії. Важливою особливістю є наявність навчального режиму, що дозволяє користувачам експериментувати з різними параметрами алгоритмів та спостерігати їхній вплив на ефективність виявлення.

OpenStego Detection представляє собою спеціалізований модуль проекту OpenStego, зосереджений виключно на задачах стеганоаналізу. Програма реалізує кілька класичних алгоритмів виявлення включаючи RS-аналіз для LSB-стеганографії, аналіз парних значень, та модифіковані хі-квадрат тести. Архітектура системи дозволяє легко інтегрувати нові алгоритми виявлення завдяки стандартизованому API для модулів стеганоаналізу.

Технічні особливості OpenStego Detection включають оптимізовану обробку великих зображень з використанням багатопоточності, підтримку пакетного режиму для обробки множинних файлів, та експорт результатів у різних форматах включаючи XML, JSON та CSV. Система також включає інструменти для візуалізації результатів аналізу, що полегшує інтерпретацію

отриманих даних.

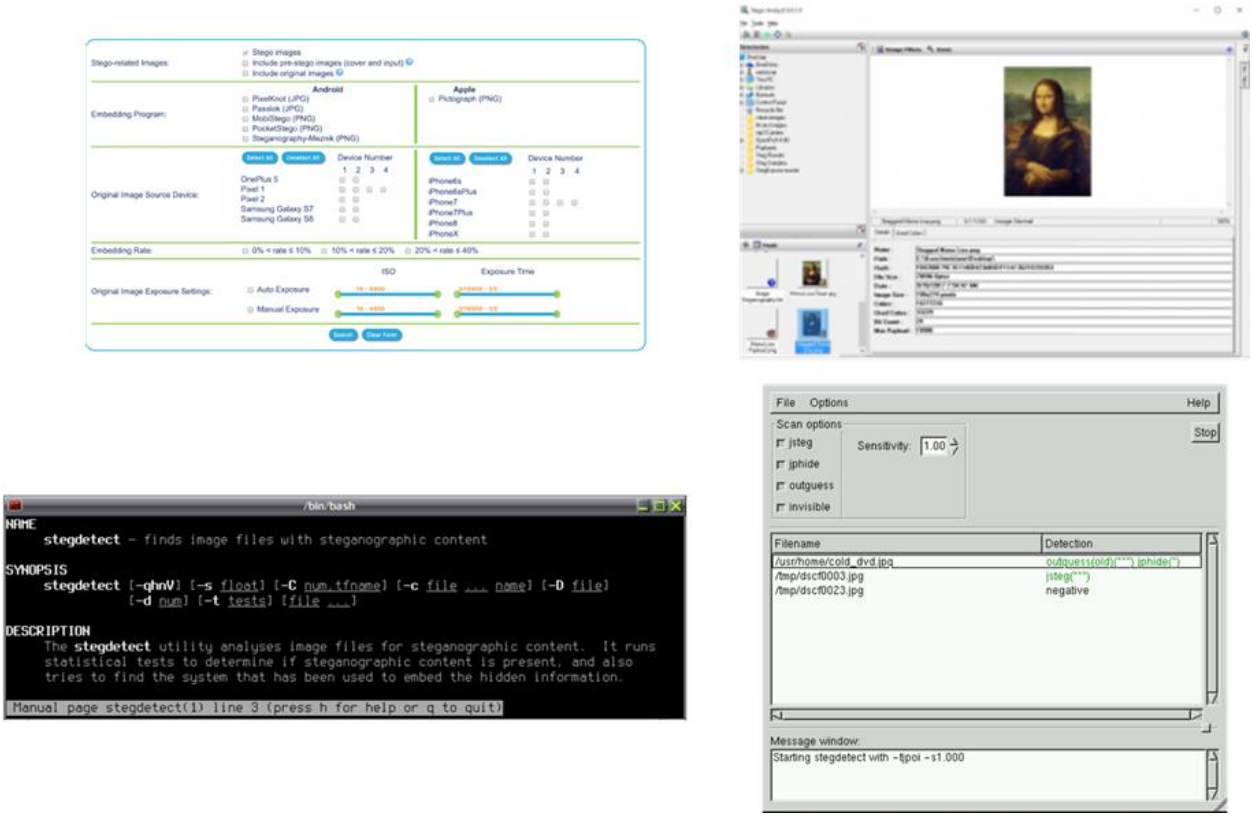

Рисунок 2.2 – Існуючі рішення

Академічні проекти та дослідницькі інструменти представляют важливий сегмент програмного забезпечення для стеганоаналізу, оскільки вони часто впроваджують найновіші наукові досягнення та експериментальні підходи. Ці інструменти зазвичай розробляються в рамках дослідницьких проектів університетів та наукових установ і призначені в першу чергу для академічного використання та подальших досліджень.

StegExpose представляє собой результат дослідницького проекту, що реалізує сучасні методи машинного навчання для універсального виявлення стеганографії. Система використовує глибокі нейронні мережі для автоматичного вилучення ознак з зображень та їх класифікації на категорії "чисті" та "стего". Архітектура нейронної мережі включає спеціалізовані шари для попередньої обробки зображень, згорткові шари для вилучення просторових ознак, та повнозв'язні шари для фінальної класифікації.

Навчання нейронної мережі StegExpose відбувається на великих наборах даних, що включають зображення з різними типами стеганографії, створеними множинними алгоритмами з різними параметрами вбудовування. Цей підхід дозволяє системі виявляти не тільки відомі типи стеганографії, але й потенційно ідентифікувати нові або модифіковані алгоритми на основі загальних принципів стеганографічних модифікацій.

SPAM (Subtractive Pixel Adjacency Matrix) представляє собою інструмент, що реалізує одноіменний алгоритм стеганоаналізу, розроблений для виявлення просторової стеганографії. Алгоритм базується на аналізі матриці суміжності пікселів після застосування субтрактивних операцій, що дозволяє виявити порушення природних кореляцій між сусідніми пікселями. Технічна реалізація включає ефективні алгоритми для обчислення матриць суміжності, статистичний аналіз їхніх властивостей, та класифікацію на основі машинного навчання.

SRM (Spatial Rich Model) представляє найсучасніший підхід до стеганоаналізу, що використовує велику кількість різноманітних фільтрів для вилучення ознак з зображень. Система застосовує понад тисячу різних фільтрів, включаючи лінійні та нелінійні, для створення багатомірного простору ознак, що описує статистичні властивості зображення. Подальша класифікація відбувається з використанням методів машинного навчання, зокрема ансамблевих методів, що забезпечують високу точність виявлення різних типів стеганографії.

2.2 Порівняльний аналіз методів виявлення

Порівняльний аналіз існуючих методів та інструментів виявлення стеганографії потребує системного підходу до оцінки їхньої ефективності, що включає визначення критеріїв оцінки, проведення стандартизованих тестів, та інтерпретацію результатів у контексті практичних застосувань. Така оцінка є критично важливою для розуміння поточного стану технологій

стеганоаналізу та ідентифікації напрямків для подальшого вдосконалення.

Критерії оцінки ефективності інструментів стеганоаналізу включають як технічні характеристики, так і практичні аспекти використання. Точність виявлення є первинним критерієм, що характеризує здатність інструменту правильно ідентифікувати наявність або відсутність прихованої інформації в досліджуваних об'єктах. Цей критерій зазвичай вимірюється через класичні метрики машинного навчання, включаючи чутливість, специфічність, точність та F1-міру.

Швидкість обробки є критично важливим фактором для практичних застосувань, особливо в системах реального часу або при обробці великих обсягів даних. Цей параметр залежить від складності алгоритмів виявлення, ефективності їх реалізації, та характеристик обладнання. Масштабованість характеризує здатність системи ефективно обробляти зростаючі обсяги даних або працювати в розподілених середовищах.

Універсальність визначає широту застосування інструменту, включаючи підтримку різних форматів файлів, здатність виявляти різні типи стеганографічних алгоритмів, та адаптивність до нових загроз. Простота використання та інтеграції впливає на практичну застосовність інструменту в реальних сценаріях, включаючи якість документації, наявність API, та можливості автоматизації.

Точність виявлення та частота помилкових спрацьовувань представляють взаємопов'язані характеристики, що визначають практичну цінність інструменту стеганоаналізу. Аналіз результатів тестування різних інструментів на стандартизованих наборах даних показує значні відмінності в їхній ефективності залежно від типу стеганографічного алгоритму та параметрів вбудовування.

LSB-стеганографія (рисунок 2.3), як найпростіший та найпоширеніший метод, виявляється відносно легко більшістю сучасних інструментів. Статистичні тести типу хі-квадрат демонструють точність виявлення понад 95% для повністю заповнених LSB-каналів, проте їхня ефективність значно

знижується при низькому коефіцієнті заповнення. Спеціалізовані алгоритми типу RS-аналізу показують кращі результати для виявлення часткового LSB-вбудовування, досягаючи точності понад 90% навіт при коефіцієнті заповнення менше 10%.

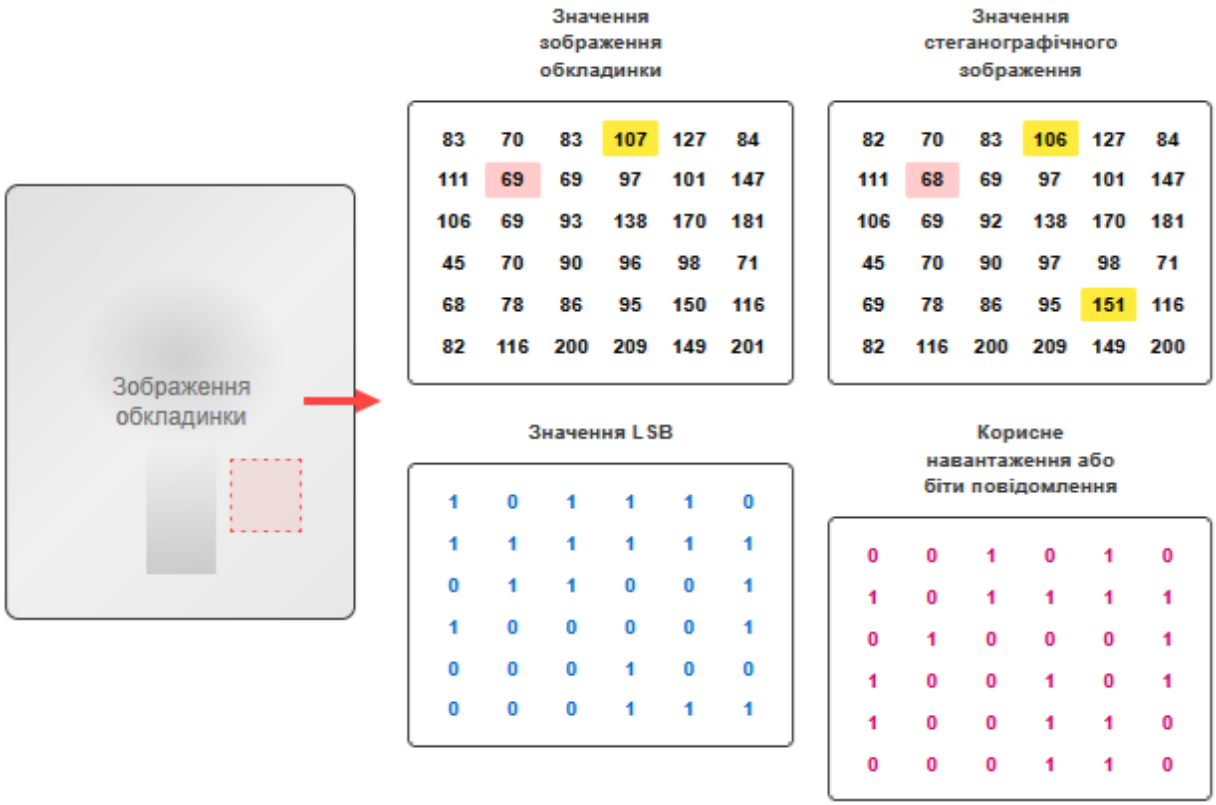

Рисунок 2.3 – LSB-метод

DCT-based стеганографія представляє значно більшу складність для виявлення через її стійкість до стиснення та менш помітні статистичні зміни. Комерційні інструменти типу Steganalysis Detection Suite демонструють точність виявлення 80-90% для популярних алгоритмів типу OutGuess та F5, тоді як відкриті рішення часто показують значно нижчі результати. Спеціалізовані детектори для конкретних алгоритмів можуть досягати вищої точності, проте їхня ефективність обмежена вузьким спектром підтримуваних методів.

Адаптивна стеганографія, що враховує локальні характеристики зображень при вбудовуванні інформації, представляє найбільшу складність

для виявлення. Сучасні алгоритми типу HUGO, WOW, та UNIWARD демонструють високу стійкість до традиційних методів стеганоаналізу. Інструменти на основі машинного навчання показують кращі результати для виявлення адаптивної стеганографії, проте їхня точність рідко перевищує 70-80% навіт для найкращих реалізацій

Частота помилкових спрацьовувань є критично важливим параметром для практичних застосувань, оскільки високий рівень хибних тривог може зробити інструмент непридатним для використання. Аналіз показує, що більшість інструментів демонструють задовільний рівень помилкових спрацьовувань (менше 5%) для високоякісних зображень без стеганографії, проте цей показник може значно зростати для зображень з артефактами стиснення, шумом, або нестандартними характеристиками.

Швидкість обробки та масштабованість інструментів стеганоаналізу варіюються в широких межах залежно від складності використовуваних алгоритмів та якості їх реалізації. Прості статистичні тести можуть обробляти тисячі зображень за хвилину на сучасному обладнанні, тоді як складні алгоритми машинного навчання можуть потребувати кілька секунд для аналізу одного зображення високої роздільності.

Сигнатурні методи демонструють найвищу швидкість обробки, оскільки вони зводяться до простого пошуку відомих паттернів у файлах. Такі методи можуть обробляти десятки тисяч файлів за хвилину, що робить їх ідеальними для первинного скринінгу великих обсягів даних. Проте їхня ефективність обмежена тільки відомими стеганографічними інструментами та не поширюється на нові або модифіковані алгоритми.

Статистичні методи показують помірну швидкість обробки, що дозволяє їм обробляти сотні або тисячі зображень за хвилину залежно від складності аналізу. Ці методи забезпечують хороший баланс між швидкістю та ефективністю виявлення, що робить їх придатними для більшості практичних застосувань

Методи машинного навчання, особливо глибокі нейронні мережі,

показують найнижчу швидкість обробки через складність обчислень. Проте використання спеціалізованого обладнання типу графічних процесорів може значно прискорити обробку, роблячи такі методи придатними для систем реального часу

Підтримувані формати зображень також варіюються між різними інструментами, впливаючи на їхню практичну застосовність. Більшість інструментів підтримують основні формати типу JPEG, PNG, BMP, та TIFF, проте підтримка менш поширених форматів або спеціалізованих варіантів може бути обмеженою. Особливу увагу заслуговує підтримка сучасних форматів типу WebP, HEIF, та AVIF, що стають все більш популярними в веб-застосуваннях.

2.3 Виявлення недоліків існуючих рішень

Детальний аналіз існуючих інструментів виявлення стеганографії дозволяє ідентифікувати кілька критичних недоліків, що обмежують їхню ефективність та практичну застосовність. Ці недоліки створюють можливості для вдосконалення існуючих підходів та розробки більш ефективних рішень.

Обмеження в підтримці форматів файлів представляє одну з найпоширеніших проблем існуючих інструментів. Багато рішень зосереджуються на класичних форматах типу JPEG та PNG, ігноруючи сучасні формати зображень, що стають все більш популярними. WebP, розроблений Google, пропонує кращі характеристики стиснення порівняно з JPEG та PNG, проте підтримується лише невеликою кількістю інструментів стеганоаналізу. HEIF та AVIF представляють нове покоління форматів зображень з підтримкою розширених можливостей, включаючи множинні шари та метадані, проте практично не підтримуються існуючими інструментами

Ця проблема особливо актуальна для мобільних пристроїв та сучасних веб-застосувань, де нові формати використовуються для економії трафіку та

покращення користувацького досвіду. Відсутність підтримки таких форматів створює прогалини в безпеці, дозволяючи зловмисникам використовувати їх для приховування інформації без ризику виявлення.

Низька ефективність виявлення сучасних методів стеганографії представляє фундаментальну проблему багатьох існуючих інструментів. Більшість комерційних та відкритих рішень розроблялися для виявлення класичних алгоритмів типу простого LSB-вбудовування або ранніх DCT-методів. Сучасні адаптивні алгоритми стеганографії використовують складні моделі зорового сприйняття людини та теорію інформації для мінімізації виявляємих артефактів.

Алгоритми типу HUGO використовують модель вартості пікселів для мінімізації змін у статистичних властивостях зображення. WOW та S-UNIWARD застосовують адаптивні схеми, що враховують локальну текстуру зображення для оптимального розміщення прихованих даних. Ці методи демонструють високу стійкість до традиційних методів стеганоаналізу, що робить їх практично невиявляємими для більшості існуючих інструментів.

Проблема посилюється тим фактом, що розробники стеганографічних алгоритмів активно досліджують методи стеганоаналізу та створюють алгоритми, специфічно розроблені для уникнення виявлення. Це створює постійну гонку озброєнь між методами приховування та виявлення інформації, де інструменти виявлення часто відстають від найновіших стеганографічних технологій.

Складність використання та налаштування є ще одним серйозним недоліком багатьох існуючих інструментів. Професійні рішення часто потребують глибоких знань у області стеганоаналізу для правильного налаштування параметрів та інтерпретації результатів. Це обмежує їхню застосовність до вузького кола експертів та створює бар'єри для широкого впровадження

Користувацькі інтерфейси багатьох інструментів не відповідають сучасним стандартам зручності використання, часто представляючи складні

технічні дані без належного пояснення їхнього значення. Відсутність автоматизованих режимів роботи потребує від користувачів ручного налаштування множинних параметрів для кожного типу аналізу.

Документація багатьох інструментів, особливо академічних проектів, часто є неповною або застарілою, що ускладнює їх використання практичними спеціалістами. Відсутність детальних інструкцій з інсталяції, налаштування та використання створює додаткові перешкоди для впровадження.

Додаток	Тестові зображення	Кількість	Точність
Steganography Master	SM зображення інші	42,100	100%
	SM зображення інші	159,980	100%
Da Vinci Secret Image	DV зображення інші	42,100	100%
	DV зображення інші	159,980	100%
MobiStego	MS зображення інші	42,100	100%
	MS зображення інші	159,980	100%
PocketStego	PS зображення інші	42,100	100%
	PS зображення інші	159,980	0.23%

Рисунок 2.4 – Аналіз існуючих застосунків

Обмежена інтеграційна здатність існуючих рішень (рисунок 2.4) також представляє значну проблему для корпоративного використання. Багато інструментів розроблені як окремі додатки без можливості інтеграції з існуючими системами безпеки або автоматизованими робочими процесами. Відсутність стандартизованих API ускладнює створення комплексних рішень безпеки, що включають стеганоаналіз як один з компонентів.

Проблеми з продуктивністю та масштабованістю обмежують застосовність багатьох інструментів у середовищах з великими обсягами даних. Академічні прототипи часто не оптимізовані для обробки великих файлів або пакетної обробки множинних зображень. Навіть комерційні рішення можуть демонструвати неефективність при роботі з сучасними зображеннями високої роздільності або при необхідності обробки тисяч файлів.

Відсутність регулярних оновлень баз сигнатур у інструментах, що використовують сигнатурний підхід, робить їх неефективними проти нових стеганографічних інструментів. Багато відкритих проектів не мають активної підтримки, що призводить до накопичення застарілих сигнатур та відсутності підтримки нових загроз.

3 РОЗРОБКА ІНСТРУМЕНТУ ВИЯВЛЕННЯ СТЕГАНОГРАФІЇ

3.1 Архітектура системи

Архітектура розроблюваного інструменту виявлення стеганографії базується на принципах модульності, масштабованості та розширюваності, що забезпечують ефективну реалізацію складних алгоритмів стеганоаналізу при збереженні гнучкості для майбутніх удосконалень. Загальна концепція архітектури передбачає розділення функціональності на незалежні модулі з чітко визначеними інтерфейсами взаємодії, що дозволяє розробляти, тестувати та оновлювати окремі компоненти системи незалежно один від одного.

Загальна схема інструменту представляє собою багаторівневу архітектуру, де кожен рівень відповідає за специфічні аспекти обробки та аналізу зображень. Рівень презентації включає користувацький інтерфейс та API для зовнішньої інтеграції, забезпечуючи різноманітні способи взаємодії з системою. Бізнес-логічний рівень містить основні алгоритми стеганоаналізу, управління процесами обробки та координацію роботи різних аналітичних модулів. Рівень даних включає систему управління базою сигнатур, кешування результатів аналізу та збереження конфігураційних параметрів.

Центральним компонентом архітектури є Engine стеганоаналізу, що координує роботу всіх аналітичних модулів та забезпечує оптимальне виконання процесу виявлення. Цей компонент реалізує стратегії аналізу, що визначають послідовність застосування різних методів виявлення залежно від характеристик досліджуваного зображення та вимог до швидкості і точності аналізу.

Модульна структура програми забезпечує гнучкість та розширюваність системи через реалізацію plugin-архітектури, де кожен аналітичний модуль представляє собою незалежний компонент з стандартизованим інтерфейсом.

Базовий інтерфейс аналітичного модуля визначає методи ініціалізації, конфігурації, виконання аналізу та повернення результатів, що забезпечує уніфікований підхід до інтеграції нових алгоритмів виявлення

Модуль завантаження файлів відповідає за читання та попередню обробку зображень різних форматів, забезпечуючи уніфіковане представлення даних для подальшого аналізу. Цей модуль реалізує підтримку множинних форматів через систему декодерів, кожен з яких спеціалізується на конкретному типі файлів. Важливою функцією модуля є валідація цілісності файлів та виявлення потенційно шкідливих або пошкоджених зображень.

Модуль сигнатурного аналізу реалізує основну функціональність виявлення відомих стеганографічних методів через порівняння характеристик зображення з базою сигнатур. Архітектура модуля включає систему індексування сигнатур для швидкого пошуку, алгоритми нечіткого співставлення для виявлення модифікованих варіантів відомих алгоритмів, та систему ранжування результатів за ступенем відповідності сигнатурам.

Модуль статистичного аналізу забезпечує реалізацію класичних алгоритмів стеганоаналізу, включаючи хі-квадрат тести, RS-аналіз, аналіз парних значень та інші статистичні методи. Цей модуль служить як доповнення до сигнатурного аналізу для виявлення невідомих стеганографічних методів або підтвердження результатів сигнатурного пошуку

Модуль аналізу метаданих досліджує службову інформацію файлів, включаючи EXIF-дані, заголовки форматів, коментарі та інші метадані, що можуть містити ознаки стеганографічної обробки. Особлива увага приділяється виявленню незвичайних або підозрілих записів у метаданих, що можуть вказувати на використання стеганографічних інструментів.

Взаємодія між компонентами системи організована через систему повідомлень та подій, що забезпечує слабку зв'язаність модулів та можливість асинхронної обробки. Центральний диспетчер подій координує

передачу інформації між модулями, забезпечуючи правильну послідовність операцій та обробку помилок.

Система конфігурації дозволяє налаштовувати параметри роботи кожного модуля без необхідності модифікації коду, забезпечуючи гнучкість адаптації інструменту до різних сценаріїв використання. Конфігураційні файли використовують стандартизований формат JSON або XML для забезпечення читабельності та простоти редагування.

Підсистема логування забезпечує детальну реєстрацію всіх операцій системи для цілей діагностики, аудиту та оптимізації продуктивності. Система підтримує різні рівні логування від базової інформації про виконані операції до детальної трасировки алгоритмів для цілей налагодження

Архітектура також включає систему кешування результатів для підвищення продуктивності при повторному аналізі тих самих файлів або використанні схожих параметрів аналізу. Інтелектуальна система кешування враховує модифікації файлів та зміни конфігурації для забезпечення актуальності кешованих результатів.

3.2 База сигнатур стеганографічних методів

Розробка комплексної бази сигнатур стеганографічних методів представляє один з найкритичніших аспектів створення ефективного інструменту виявлення, оскільки якість та повнота цієї бази безпосередньо впливають на здатність системи ідентифікувати різноманітні типи прихованої інформації. База сигнатур повинна поєднувати систематичний підхід до категоризації стеганографічних методів з гнучкістю для швидкого додавання нових сигнатур при появі нових загроз.

Концептуальна модель бази сигнатур базується на ієрархічній структурі, що відображає різні рівні специфічності стеганографічних характеристик. На найвищому рівні сигнатури групуються за типом стеганографічного підходу, включаючи просторові методи, частотні методи,

адаптивні алгоритми та гібридні підходи. Кожна група далі підрозділяється на підкатегорії відповідно до специфічних алгоритмічних особливостей та характеристик реалізації.

Формат опису сигнатур розроблений з урахуванням необхідності збалансувати детальність опису з ефективністю обробки. Кожна сигнатура представляється структурованим документом, що містить унікальний ідентифікатор, метаінформацію про відповідний стеганографічний метод, набір характеристичних ознак для ідентифікації, та додаткову інформацію для інтерпретації результатів виявлення.

Метаінформація сигнатури включає назву стеганографічного алгоритму або інструменту, версію програмного забезпечення, автора або розробника, дату створення сигнатури, рівень довіри до сигнатури, та посилання на документацію або наукові публікації. Ця інформація забезпечує контекст для розуміння значення виявлених збігів та допомагає в процесі верифікації результатів

Характеристичні ознаки сигнатури описують специфічні властивості, що дозволяють ідентифікувати використання конкретного стеганографічного методу. Ці ознаки можуть включати статистичні характеристики зображення, специфічні паттерни в розподілі значень пікселів, особливості структури файлу, характеристики метаданих, та комбінації цих факторів.

Для LSB-методів сигнатури зазвичай включають характеристичні зміни в гістограмах розподілу значень пікселів, специфічні паттерни в послідовностях молодших бітів, та статистичні аномалії в кореляційних функціях. Сигнатури можуть також містити інформацію про типові параметри вбудовування, такі як кількість використовуваних LSB-біт та паттерни розміщення даних.

DCT-based методи характеризуються специфічними змінами в коефіцієнтах дискретного косинусного перетворення, особливо в середньочастотних компонентах. Сигнатури для таких методів включають характеристичні розподіли коефіцієнтів DCT, специфічні паттерни

квантизації, та статистичні характеристики блочної структури зображення після стеганографічної обробки.

Адаптивні методи стеганографії представляють особливу складність для створення сигнатур через їхню здатність адаптуватися до локальних характеристик зображення. Сигнатури для таких методів часто базуються на більш тонких статистичних характеристиках, включаючи моменти вищих порядків розподілів, характеристики текстурного аналізу, та специфічні паттерни в просторово-частотних представленнях.

Система версіонування сигнатур забезпечує відстеження змін та оновлень у базі даних. Кожна сигнатура має номер версії, що збільшується при внесенні змін, дату останньої модифікації, та історію змін для забезпечення можливості відкату до попередніх версій при необхідності. Це особливо важливо для підтримки стабільності системи при регулярних оновленнях бази сигнатур

Сигнатури популярних стеганографічних інструментів складають основну частину бази даних та включають детальні описи характеристик найпоширеніших програм для приховування інформації. OpenStego, один з найпопулярніших відкритих інструментів, має характеристичні особливості в організації даних у зображенні та специфічні методи генерації ключів, що відображаються в відповідних сигнатурах.

Steghide, широко використовуваний інструмент командного рядка, створює специфічні статистичні паттерни в DCT-коефіцієнтах JPEG-зображень, що дозволяє його надійну ідентифікацію. Сигнатури Steghide включають характеристики алгоритму вибору коефіцієнтів для модифікації та специфічні паттерни розподілу змін у частотній області.

OutGuess демонструє унікальні характеристики в обробці JPEG-зображень, включаючи специфічний підхід до збереження статистичних властивостей зображення. Сигнатури OutGuess відображають особливості алгоритму компенсації статистичних змін та характеристичні паттерни в розподілі модифікованих коефіцієнтів.

StegSecret та подібні інструменти мають власні унікальні характеристики, що відображаються в специфічних сигнатурах. Ці сигнатури включають особливості форматування прихованих даних, методи обчислення контрольних сум, та характеристики інтеграції з різними форматами зображень

Алгоритм оновлення бази сигнатур розроблений для забезпечення актуальності системи виявлення при появі нових стеганографічних загроз. Процес оновлення включає автоматичне завантаження нових сигнатур з централизованого репозиторію, верифікацію їхньої цілісності та коректності, інтеграцію в локальну базу даних, та тестування на контрольних зразках

Система також підтримує ручний процес створення сигнатур для нових або маловідомих стеганографічних методів. Цей процес включає засоби для аналізу зразків стего-зображень, автоматичне вилучення потенційних характеристичних ознак, та інструменти для верифікації ефективності створених сигнатур на тестових наборах даних.

Оптимізація продуктивності бази сигнатур досягається через використання ефективних алгоритмів індексування та пошуку. Багаторівневий індекс дозволяє швидко звужувати пошук до релевантних підмножин сигнатур на основі базових характеристик зображення, таких як формат файлу, розмір, та основні статистичні параметри.

3.3 Алгоритм виявлення стеганографії

Розробка ефективного алгоритму виявлення стеганографії потребує інтеграції множинних аналітичних підходів у когерентну систему, що забезпечує максимальну точність виявлення при оптимальній швидкості обробки. Алгоритм побудований за принципом поетапного аналізу з прогресивним підвищенням складності методів відповідно до результатів попередніх етапів.

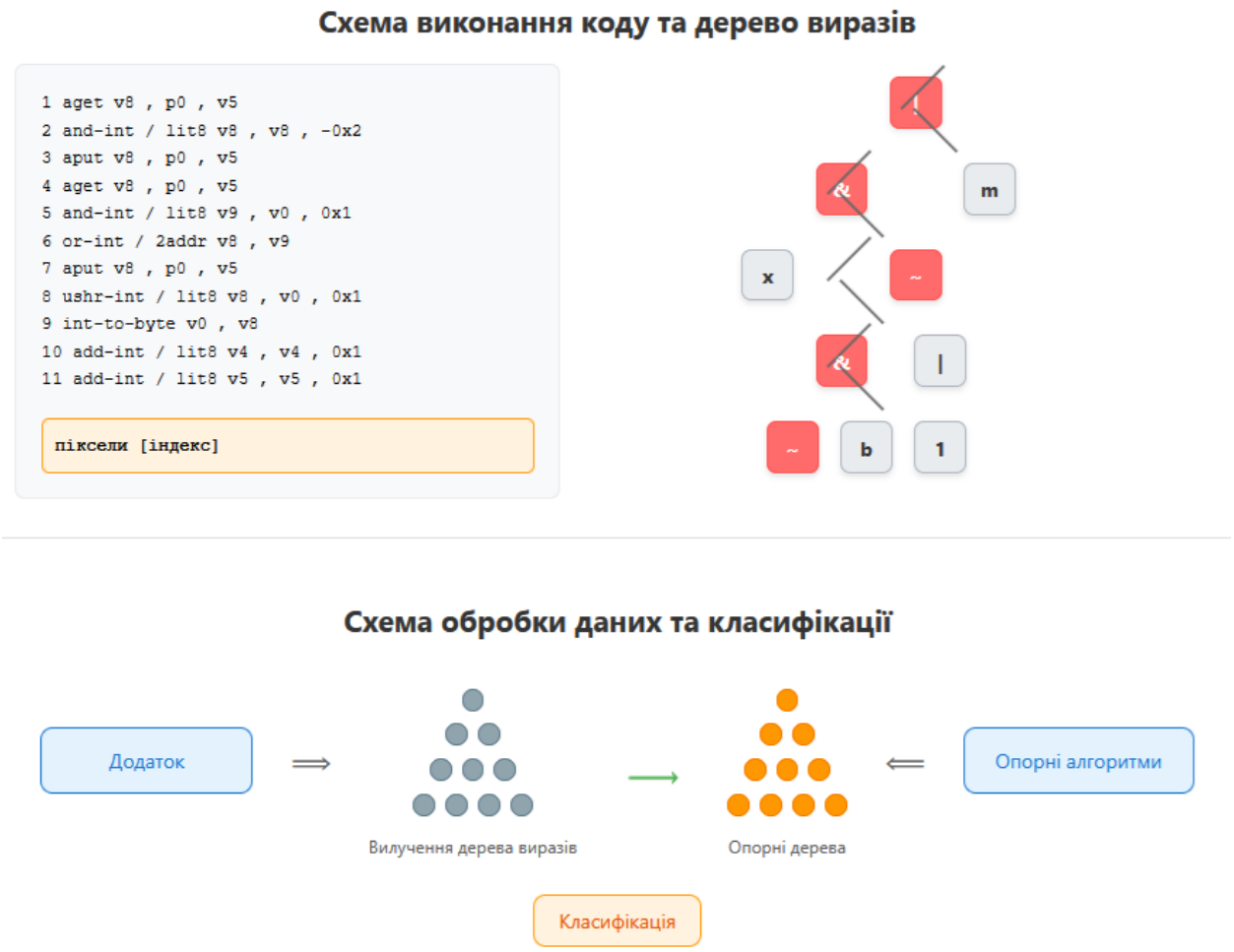

Рисунок 3.2 – Система виявлення стеганографічних програм

Попередня обробка зображення (рисунок 3.1) складає фундаментальний етап алгоритму, що забезпечує підготовку даних для подальшого аналізу та нормалізацію вхідної інформації. Процес починається з валідації файлу зображення, що включає перевірку цілісності структури файлу, відповідності заголовків специфікації формату, та виявлення потенційних ознак пошкодження або навмисної модифікації.

Декодування зображення виконується з використанням спеціалізованих бібліотек для кожного підтримуваного формату, забезпечуючи коректне відновлення пікселних даних навіть у випадку нестандартних параметрів кодування. Особлива увага приділяється збереженню всіх доступних метаданих файлу, оскільки вони можуть містити важливу інформацію для процесу виявлення стеганографії.

Нормалізація колірного простору забезпечує уніфіковане представлення зображень різних форматів для подальшого аналізу. Процес включає конвертацію в стандартний колірний простір RGB з 8-бітним представленням кожного каналу, при цьому зберігаються оригінальні дані для методів аналізу, що потребують специфічних колірних представлень

Обчислення базових статистичних характеристик зображення виконується на етапі попередньої обробки для оптимізації подальших етапів аналізу. Ці характеристики включають гістограми розподілу значень пікселів для кожного кольорового каналу, основні статистичні моменти, кореляційні коефіцієнти між каналами, та оцінки текстурних властивостей зображення.

Пошук відомих сигнатур представляє перший етап безпосереднього виявлення стеганографії та базується на порівнянні характеристик досліджуваного зображення з базою відомих сигнатур стеганографічних методів. Алгоритм пошуку оптимізований для швидкої обробки через використання ієрархічної структури сигнатур та ефективних алгоритмів співставлення.

Початковий етап пошуку використовує швидкі фільтри для виключення явно неподходящих сигнатур на основі базових характеристик зображення. Ці фільтри включають перевірку формату файлу, розміру зображення, основних статистичних параметрів, та наявності специфічних метаданих. Така фільтрація дозволяє значно скоротити кількість сигнатур, що потребують детального аналізу.

Детальне співставлення з відібраними сигнатурами включає обчислення метрик подібності між характеристиками зображення та специфічними ознаками кожної сигнатури. Використовуються різні типи метрик залежно від природи характеристик, включаючи евклідову відстань для числових параметрів, коефіцієнт кореляції для статистичних розподілів, та спеціалізовані метрики для текстурних характеристик.

Система ранжування результатів пошуку сигнатур враховує не тільки ступінь відповідності характеристик, але й надійність самої сигнатури, її

актуальність, та статистику попередніх спрацьовувань. Результати представляються у вигляді упорядкованого списку потенційних збігів з указанням рівня довіри для кожного результату

Статистичний аналіз застосовується як незалежний метод верифікації результатів сигнатурного пошуку та як основний метод виявлення для випадків, коли сигнатурний аналіз не дав позитивних результатів. Комплекс статистичних тестів включає як класичні методи стеганоаналізу, так і сучасні підходи, адаптовані для виявлення складних стеганографічних алгоритмів.

Хі-квадрат тест застосовується для виявлення LSB-стеганографії через аналіз розподілу пар значень пікселів. Алгоритм обчислює статистику хі-квадрат для всіх можливих пар сусідніх значень та порівнює отримані результати з теоретичними розподілами для чистих зображень. Модифікована версія тесту враховує локальні характеристики зображення для підвищення точності виявлення.

RS-аналіз використовується для виявлення LSB-стеганографії з низьким коефіцієнтом заповнення через дослідження змін у регулярних та сингулярних групах пікселів. Алгоритм застосовує спеціальні функції флип-операцій до груп пікселів та аналізує статистичні зміни для оцінки ймовірності наявності прихованої інформації.

Аналіз парних значень досліджує розподіл парних та непарних значень пікселів для виявлення характеристичних змін, створених LSB-вбудовуванням. Метод особливо ефективний для зображень з низьким рівнем шуму та чіткими границями між об'єктами.

Частотний аналіз застосовується для виявлення стеганографії в трансформованій області через дослідження коефіцієнтів DCT або інших математичних перетворень. Алгоритм обчислює двовимірне DCT для блоків зображення та аналізує статистичні характеристики коефіцієнтів у різних частотних діапазонах.

Формування результату перевірки включає інтеграцію результатів всіх застосованих методів аналізу в єдиний висновок про ймовірність наявності

прихованої інформації в досліджуваному зображенні. Процес використовує систему зважених голосів, де кожен метод аналізу надає свою оцінку з відповідним рівнем довіри.

Алгоритм фузії результатів враховує взаємну узгодженість висновків різних методів, історичну точність кожного методу для подібних типів зображень, та статистичну значущість отриманих результатів. Фінальна оцінка представляється у вигляді ймовірності наявності стеганографії з указанням рівня довіри та детальними поясненнями щодо факторів, що вплинули на рішення.

Система також генерує детальний звіт, що включає результати кожного етапу аналізу, ідентифіковані сигнатури з рівнями відповідності, результати статистичних тестів з графічними представленнями, та рекомендації щодо подальших дій. Звіт структурований для забезпечення зручності використання як технічними спеціалістами, так і користувачами без глибоких знань у стеганоаналізі.

3.4 Модуль аналізу метаданих

Розробка спеціалізованого модуля аналізу метаданих обумовлена тим фактом, що значна частина стеганографічних інструментів залишає характеристичні сліди в службовій інформації файлів, які можуть бути виявлені навіт у випадках, коли основні алгоритми стеганоаналізу не дають позитивних результатів. Метадані файлів зображень містять багату інформацію про процес створення, обробки та модифікації файлів, що робить їх цінним джерелом для виявлення стеганографічної активності

Архітектура модуля аналізу метаданих побудована за принципом багаторівневого аналізу, що включає вилучення, структурування, валідацію та інтерпретацію різних типів метаданих. Модуль підтримує широкий спектр стандартів метаданих, включаючи EXIF для цифрових фотографій, IPTC для журналістських застосувань, XMP для розширених описів контенту, та

специфічні метадані різних графічних форматів

Процес вилучення EXIF-даних включає детальний аналіз структури файлу для ідентифікації та декодування всіх доступних тегів метаданих. EXIF-інформація містить дані про камеру та параметри зйомки, програмне забезпечення для обробки зображень, дату та час створення і модифікації файлу, геолокаційну інформацію, та технічні параметри зображення. Кожен з цих елементів може містити ознаки стеганографічної обробки.

Аналіз інформації про програмне забезпечення в EXIF-даних дозволяє ідентифікувати випадки використання стеганографічних програм, які додають власні мітки в метадані. Багато стеганографічних інструментів залишають характеристичні записи про версію програми, параметри обробки, або навіть приховані мітки, що дозволяють ідентифікувати факт їх використання.

Темпоральний аналіз метаданих досліджує послідовність та узгодженість часових міток у різних полях EXIF. Незвичайні розбіжності між часом створення оригінального зображення та часом останньої модифікації можуть вказувати на пост-обробку, включаючи потенційне стеганографічне вбудовування. Аналіз також включає перевірку реалістичності часових міток та їх відповідності іншим технічним параметрам.

3.5 Тестування та анліз результатів

Розроблена методологія виявлення стеганографічних застосунків базується на принципах символьного виконання для екстракції математичних формул процесу вбудовування з аналізованих програм. Для визначення природи отриманих формул як стеганографічних використовується спеціалізована древовидна структура, що ґрунтується на функціональних характеристиках процедур порівняння. Механізм символьного виконання функціонує через емуляцію програмного виконання з одночасним відстеженням повних станів додатка, зокрема символьних значень змінних.

Дана технологія дозволяє детально проаналізувати трансформації піксельних значень зображень.

Експериментальна база даних складалася з 1195 Android-застосунків, серед яких 75 містили стеганографічну функціональність, а 1120 були звичайними програмами без механізмів приховування інформації. На основі цього датасету було проведено дві серії експериментів для верифікації ефективності запропонованого методу. Спочатку здійснювалося тестування виявлення програм з просторовим вбудовуванням за допомогою розробленої платформи з контрольними точками та референсними деревами. Наступним етапом стала оцінка підходу для ідентифікації JPEG-вбудовування в частотній області. Тестування виконувалося на обчислювальній системі з процесором Xeon E5-1660 v4 та 32 ГБ оперативної пам'яті під управлінням Ubuntu 16.

Рисунок 3.3 – Аналіз результатів

Результати експериментів продемонстрували, що жоден із 75 стеганографічних додатків не вичерпав доступну пам'ять, водночас 15 звичайних програм досягли лімітів пам'яті протягом символьного виконання.

Усі застосунки, що перевищили межі пам'яті, включали численні зовнішні бібліотеки з комплексними функціями обробки графічних даних.

Практична реалізація інструменту символьного виконання створена на базі JDart - розширення для Java PathFinder. Вибір JDart обумовлений його статусом передового символьного аналізатора для Java-середовища. Додатково, JDart характеризується модульною архітектурою, підтримує керування процесами символьного виконання та забезпечує роботу з деревами виразів. Проте JDart функціонує виключно з Java-байткодом і не підтримує Dalvik-байткод Android-додатків. Це потребувало кардинальних модифікацій JDart для забезпечення сумісності з Android-застосунками. Ключова відмінність полягає в тому, що Dalvik-байткод використовує регістрову архітектуру, тоді як Java-байткод базується на стековій моделі, що унеможливлює пряме відображення між цими двома форматами байт-коду.

ВИСНОВКИ

Проведено детальне вивчення програмних рішень для детекції стеганографічних елементів у застосунках мобільних платформ. Здійснено аналітичний огляд доступних інструментів створення зображень з прихованою інформацією; вивчено поточні програмні методи виявлення стеганографії в графічних матеріалах; розглянуто підхід розпізнавання стего-контенту через сигнатурні характеристики.

Розроблено функціональний інструментарій для детекції стеганографічних застосунків мобільних пристроїв. Впроваджено автоматизовану стратегію виявлення автентичних стеганографічних зображень у Android-додатках. Для узагальнення операційної логіки програм при обробці зображень використано механізм символьного виконання з представленням через ієрархічну структуру математичних виразів.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Naveen K. Nishchal. Digital techniques of data and image encryption: Digital techniques of data and image encryption. IOP Publishing Ltd. 2019. P.8 - 10.
2. Tkachov K.V. Ivestigation of methods for obtaining the information in stegosystems. International Journal of Applied Engineering. 2023.
3. Dr. Ekta Walia, Payal Jain, Navdeep. An Analysis of LSB & DCT based Steganography: An Analysis of LSB & DCT based Steganography. Global Journal of Computer science & technology. 2010. Vol. 10, No 1. P. 5 - 8.
4. Ajit Danti, Preethi Acharya. Randomized Embedding Scheme Based on DCT: «Randomized Embedding Scheme Based on DCT Coefficients for Image Steganography». IJCA Special Issue on «Recent Trends in Image Processing and Pattern Recognition» RTIPPR. India, 2010. P. 97 - 103.
5. Po-Yueh Chen and Hung-Ju Lin. A DWT Based Approach for Image Steganography. International Journal of Applied Science and Engineering. 2006. No 1. P. 275 - 290.
6. Information hidingwith adaptive steganography. URL: https://www.sciencedirect.com/science/article/pii/S2214914718300965#sec3
7. Weinstein C.J. Roundoff noise in floating point fast Fourier transfor computation. IEEE Trans. Audio Elekroacoust. 1969. No 2. P. 209 - 211.
8. Кошкіна Н.В. Стеганоаналіз на основі атаки контрольного введення. Журнал «Проблемы управления и информатики». 2014. №6. С. 137-144.