Форма № 24

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет "Запорізька політехніка"

Факультет інформаційної безпеки та електронних комунікацій
(повне найменування інституту, факультету)

Кафедра інформаційної безпеки та наноелектроніки
(повне найменування кафедри)

Пояснювальна записка

до дипломного проєкту (роботи)
магістра
(ступінь вищої освіти)

на тему Розроблення та аналіз алгоритму приховування інформації в графічних файлах із підвищеною стійкістю до стеганоаналізу
(назва теми)

Виконав(ла): студент(ка) 2 курсу, групи БКз-814м

Спеціальності 125 Кібербезпека та захист інформації
(код і найменування спеціальності)

Освітня програма (спеціалізація) Безпека інформаційних і комунікаційних систем

НАПАДАЙЛО С.А.
(ПРІЗВИЩЕ та ініціали)

Керівник КОРОТУН А.В.
(ПРІЗВИЩЕ та ініціали)

Рецензент МОРОЗ Г.В.
(ПРІЗВИЩЕ та ініціали)

2025

Форма № 25

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»

Факультет інформаційної безпеки та електронних комунікацій
Кафедра інформаційної безпеки та наноелектроніки
Ступінь вищої освіти магістр
Спеціальність 125 Кібербезпека та захист інформації
(код і найменування)
Освітня програма (спеціалізація) Безпека інформаційних і комунікаційних систем
(назва освітньої програми (спеціалізації))

ЗАТВЕРДЖУЮ
Завідувач кафедри ІБтаН
Андрій КОРОТУН
«____» _____2025 року

З А В Д А Н Н Я

НА ДИПЛОМНИЙ ПРОЄКТ (РОБОТУ) СТУДЕНТА

НАПАДАЙЛО Сергій Анатолійович
(ПРІЗВИЩЕ, ім'я, по батькові)

1. Тема проєкту (роботи) Розроблення та аналіз алгоритму приховування інформації в графічних файлах із підвищеною стійкістю до стеганоаналізу
керівник проєкту (роботи) канд. фіз.-мат. наук., доц., КОРОТУН Андрій Віталійович
(науковий ступінь, вчене звання, ПРІЗВИЩЕ, ім'я, по батькові)
затверджені наказом закладу вищої освіти від «26» листопада 2025 року № 530
2. Строк подання студентом проєкту (роботи) 22.12.2025
3. Вихідні дані до проєкту (роботи) Дипломний проєкт присвячено розробленню алгоритму приховування інформації у графічних файлах із підвищеною стійкістю до стеганоаналізу. Об'єктом і предметом дослідження є методи стеганографічного приховування даних та їх виявлення. Метою роботи є підвищення стеганостійкості шляхом аналізу існуючих методів і розроблення нового алгоритму. У роботі використано методи НЗР, Коха-Жао, Хі-квадрат і RS, реалізовані програмно мовою Python. Результатом є алгоритм приховування інформації з підвищеною стійкістю до виявлення.
4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) Розрахунково-пояснювальна записка містить огляд методів стеганографії та стеганоаналізу графічних файлів, аналіз методів Хі-квадрат, RS і Коха-Жао, а також результати їх програмної реалізації та експериментального дослідження. Запропоновано комбінований підхід до стеганоаналізу й алгоритм приховування інформації з підвищеною стеганостійкістю.
5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, кількість слайдів, плакатів) Презентація доповіді (в MS PowerPoint), 21 слайд.

6. Консультанти розділів проєкту (роботи)

Розділ	ПРІЗВИЩЕ, ініціали та посада Консультанта	Підпис, дата	
		завдання видав	прийняв виконане завдання
1-3	КОРОТУН А.В., доцент кафедри ІБтаН	04.09.2025	19.12.2025
Нормоконтроль	КОРОЛЬКОВ Р.Ю., доцент кафедри ІБтаН		22.12.2025

7. Дата видачі завдання «04» вересня 2025 року.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проєкту (роботи)	Строк виконання етапів проєкту (роботи)	Примітка
1.	Вибір теми та аналіз літературних джерел.	04.09.25 – 20.09.25	Виконано
2.	Аналіз методів стеганографії та стеганоаналізу.	21.09.25 – 30.09.25	Виконано
3.	Розробка алгоритмів і математичних моделей.	01.10.25 – 15.10.25	Виконано
4.	Програмна реалізація розроблених методів.	16.10.25 – 10.11.25	Виконано
5.	Проведення експериментальних досліджень.	11.11.25 – 01.12.25	Виконано
6.	Аналіз результатів та оформлення пояснювальної записки.	02.12.25 – 10.12.25	Виконано
7.	Підготовка матеріалів і захист дипломного проєкту.	11.12.25 – 22.12.25	Виконано

Студент(ка) _____ Сергій НАПАДАЙЛО
(підпис) (Ім'я ПРІЗВИЩЕ)

Керівник проєкту (роботи) _____ Андрій КОРОТУН
(підпис) (Ім'я ПРІЗВИЩЕ)

АНОТАЦІЯ

Пояснювальна записка до магістерської роботи: 69 с., 3 табл., 9 рис., 1 дод., 20 джерел.

СТЕГАНОГРАФІЯ, СТЕГАНОАНАЛІЗ, ГРАФІЧНІ ФАЙЛИ, ПРИХОВУВАННЯ ІНФОРМАЦІЇ, СТЕГАНОСТІЙКІСТЬ, ІНФОРМАЦІЙНА БЕЗПЕКА.

Дипломний проєкт присвячено розробленню та аналізу алгоритму приховування інформації у графічних файлах із підвищеною стійкістю до методів стеганоаналізу. Актуальність роботи зумовлена зростанням обсягів передавання мультимедійних даних у відкритих мережах та необхідністю забезпечення прихованості інформаційних повідомлень і захисту від несанкціонованого виявлення прихованих даних.

У роботі проведено аналіз сучасного стану стеганографії графічних файлів, розглянуто основні методи приховування інформації у просторовій та частотній областях, зокрема метод найменш значущих розрядів і метод Коха-Жао. Детально досліджено методи стеганоаналізу, зокрема статистичний метод Хі-квадрат, регулярно-сингулярний метод та підходи до їх поєднання. Наведено особливості застосування кожного з методів, їх переваги, недоліки та області ефективного використання.

У межах дипломного проєкту виконано програмну реалізацію досліджуваних методів стеганоаналізу із застосуванням мови програмування Python та проведено експериментальне тестування для різних способів вбудовування інформації. На основі отриманих результатів здійснено порівняльний аналіз точності та ефективності методів виявлення прихованих повідомлень.

За результатами дослідження запропоновано комбінований підхід до стеганоаналізу, який дозволяє підвищити надійність виявлення прихованої інформації, а також розроблено новий алгоритм приховування даних у графічних файлах із підвищеною стеганостійкістю. Проведено аналіз ефективності запропонованого алгоритму та сформульовано практичні рекомендації щодо його застосування.

Отримані результати можуть бути використані у системах захисту інформації, комп'ютерній криміналістиці, автоматизованих системах аналізу мультимедійних даних, а також у навчальному процесі при підготовці фахівців з інформаційної безпеки та телекомунікацій.

ABSTRACT

Explanatory note to the Master's thesis: 69 pages, 3 tables, 9 figures, 20 references.

STEGANOGRAPHY, STEGANALYSIS, GRAPHIC FILES, INFORMATION HIDING, STEGANOGRAPHIC ROBUSTNESS, INFORMATION SECURITY.

The diploma project is devoted to the development and analysis of an algorithm for hiding information in graphic files with increased resistance to steganalysis methods. The relevance of the work is determined by the growing volume of multimedia data transmission in open networks and the need to ensure the concealment of information messages and protection against unauthorized detection of hidden data.

The paper analyzes the current state of graphic file steganography and considers the main methods of information hiding in the spatial and frequency domains, in particular the Least Significant Bit (LSB) method and the Koch-Zhao method. Steganalysis methods are studied in detail, including the statistical Chi-square method, the Regular-Singular (RS) method, and approaches to their combination. The features of application of each method, their advantages, disadvantages, and areas of effective use are presented.

Within the framework of the diploma project, a software implementation of the investigated steganalysis methods was developed using the Python programming language, and experimental testing was carried out for various information embedding techniques. Based on the obtained results, a comparative analysis of the accuracy and efficiency of hidden message detection methods was performed.

As a result of the research, a combined steganalysis approach was proposed,

which improves the reliability of hidden information detection, and a new algorithm for hiding data in graphic files with increased steganographic robustness was developed. The effectiveness of the proposed algorithm was analyzed, and practical recommendations for its application were formulated.

The obtained results can be used in information security systems, computer forensics, automated multimedia data analysis systems, as well as in the educational process for training specialists in information security and telecommunications.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

НЗР – Least Significant Bit, LSB – найменший значущий розряд

СПІ – стеганографічне приховування інформації

СА – стеганоаналіз

ДКП – дискретне косинусне перетворення

ПЗ – пара значень

RS-метод – регулярно-сингулярний метод стеганоаналізу

DC-коефіцієнт – нульовий коефіцієнт ДКП (низькочастотна складова)

AC-коефіцієнти – Least Significant Bit, LSB – найменший значущий розряд

SM – Singular Mask – кількість сингулярних груп

p – відносна довжина прихованого повідомлення (ступінь заповнення контейнера)

ВСТУП

На сьогоднішній день питання забезпечення інформаційної безпеки набуває особливої актуальності у зв’язку зі стрімким розвитком інформаційних технологій, зростанням обсягів передавання цифрових даних та ускладненням методів несанкціонованого доступу до інформаційних систем. Однією з найбільш прихованих і водночас небезпечних форм інформаційних атак є використання методів стеганографії, що дозволяють приховувати сам факт передавання інформації.

Особливу загрозу становлять атаки, у яких як контейнери для прихованих повідомлень використовуються звичайні файли даних, насамперед файли графічних форматів. Такі файли широко застосовуються в мережевому обміні, мультимедійних сервісах та інформаційних системах, що робить їх зручним середовищем для приховування як шкідливого програмного коду, так і конфіденційної або витокової інформації без помітних візуальних ознак модифікації.

Розробкою методів виявлення прихованої інформації займається спеціалізований розділ стеганографії – стеганоаналіз, який досліджує способи детектування та оцінювання прихованих повідомлень у цифрових контейнерах. Методи стеганоаналізу повинні активно застосовуватися у практичній діяльності:

– комп’ютерними криміналістами під час розслідування інцидентів інформаційної безпеки;

– автоматизованими системами захисту інформації, оснащеними модулями аналізу файлів даних на предмет наявності в них небезпечної або шкідливої інформації.

Важливою особливістю стеганоаналітичних досліджень є необхідність комплексного аналізу різних форм і представлень файлів-контейнерів. Зокрема, для графічних зображень недостатньо перевіряти лише безпосередні значення

кольорів пікселів у просторовій області. У багатьох випадках приховування інформації здійснюється шляхом модифікації частотних характеристик зображення, що вимагає застосування методів аналізу у частотній області.

У зв’язку з цим постає низка важливих питань, пов’язаних із вибором оптимальних алгоритмів стеганоаналізу, оцінюванням їх ефективності, а також формуванням коректних і достовірних висновків за результатами їх застосування. Різні методи стеганоаналізу мають різну чутливість до способу вбудовування, обсягу прихованого повідомлення та характеристик контейнера.

Для виявлення приховування інформації, здійсненого в найменші значущі біти (НЗР) пікселів зображення, широко використовується метод Хі-квадрат, що базується на статистичному аналізі з використанням критерію Пірсона, а також регулярно-сингулярний (RS) метод, який застосовує сигнатурний аналіз груп пікселів і елементи аналітичної геометрії для оцінювання відносного обсягу прихованого повідомлення.

Для виявлення вбудовування інформації, здійсненого у частотному поданні зображення, зокрема після дискретного косинусного перетворення, використовується метод стеганоаналізу Коха-Жао, який дозволяє не лише встановити факт приховування, але й оцінити параметри, необхідні для подальшого витягу схованого повідомлення.

Таким чином, дослідження та порівняльний аналіз ефективності різних методів стеганоаналізу графічних файлів є важливим науково-практичним завданням, спрямованим на підвищення рівня захисту інформаційних систем і вдосконалення засобів виявлення прихованих інформаційних загроз.

1 ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Стеганографія

Файли та дані графічних форматів на сьогоднішній день займають значну частку мережного трафіку та використовуються практично в усіх сучасних інформаційних системах. Вони присутні як складові елементи більш комплексних форматів даних, застосовуються в графічних інтерфейсах користувача, розміщуються на різноманітних мережних ресурсах, а також широко використовуються у дизайнерських і мультимедійних рішеннях. Така повсюдність графічних даних робить їх зручним та малопомітним середовищем для прихованого передавання інформації.

Одночасно з цим упродовж останніх років спостерігається стале зростання кількості шкідливого програмного забезпечення, яке використовує у своїх атаках методи стеганографії, тобто методи приховування самого факту передавання таємного повідомлення. У більшості випадків саме файли зображень обираються як контейнери для доставки потенційно небезпечних даних, оскільки вони дозволяють приховувати значні обсяги інформації без внесення візуально помітних для людського ока спотворень. Це суттєво ускладнює виявлення таких загроз традиційними засобами захисту інформації.

Виявлення прихованої інформації у цифрових контейнерах є предметом дослідження стеганоаналізу – спеціалізованого розділу стеганографії, який вивчає методи детектування та, у ряді випадків, витягу прихованих повідомлень з інформаційних об’єктів [1]. Під прихованою інформацією, як правило, розуміють дані, вбудовані за допомогою різних стеганографічних алгоритмів. Окрім встановлення факту приховування, стеганоаналіз також досліджує можливості відновлення прихованої інформації в умовах відсутності ключових параметрів або апріорних відомостей про застосований метод вбудовування.

Незважаючи на велику кількість відомих алгоритмів стеганографічного

приховування інформації в графічних файлах, більшість із них зводиться до декількох базових підходів. До таких належать, зокрема, метод Коха-Жао, який здійснює кодування інформації в частотному поданні зображення, а також метод приховування даних у найменших значущих бітах пікселів. Для виявлення інформації, прихованої цими способами, було розроблено низку методів стеганоаналізу, програмна реалізація яких дозволяє автоматизувати процес аналізу та здійснювати його без участі людини, на відміну від суб'єктивних візуальних методів оцінювання. Значна частина інших стеганографічних алгоритмів є лише модифікаціями або варіаціями зазначених базових методів.

Комп'ютерна стеганографія загалом являє собою сукупність методів прихованого вбудовування захищених даних у структуру інших даних, що називаються контейнерами та можуть зберігатися у відкритому вигляді або передаватися незахищеними каналами зв'язку. Практичне застосування методів стеганографічного приховування інформації охоплює широкий спектр задач, зокрема реалізацію DRM-систем, захист авторських прав, створення цифрових водяних знаків, контроль незаконного поширення цифрового контенту, формування унікальних «відбитків пальців», підтвердження автентичності цифрових документів, а також приховане анотування різнорідної мультимедійної інформації [2, 3, 18].

Найбільш розповсюджені методи комп'ютерної стеганографії базуються на використанні властивостей надмірності аудіо- та візуальної інформації. У контексті графічних форматів до найпоширеніших класів алгоритмів приховування інформації належать алгоритми модифікації молодших бітів колірних значень пікселів із псевдовипадковим вибором елементів контейнера та корекцією його статистичних характеристик, алгоритми сегментації бітових площин растрових зображень за рівнем складності (BPCS-алгоритми), алгоритми модифікації таблиць квантування JPEG, методи зміни молодших бітів спектральних коефіцієнтів із корекцією статистик заповненого контейнера, а також алгоритми модифікації колірної палітри [4-6].

Слід зазначити, що практично всі розглянуті алгоритми приховування інформації призводять до спотворення графічного контейнера, яке має адитивний або мультиплікативний характер. До основних недоліків таких підходів належать відносно низька стійкість до типових операцій перетворення маркованих файлів, а також алгоритмічний характер процедури вбудовування, що реалізує логічно визначені послідовності операцій над наперед відомими елементами контейнера. Це, як правило, призводить до зміни статистичних характеристик заповненого контейнера та створює передумови для успішного застосування стеганоаналітичних атак. Корекція статистик, яка використовується в окремих алгоритмах, потребує додаткових обчислювальних витрат і не завжди гарантує непомітність приховування з погляду стеганоаналізу.

У більшості наукових робіт, присвячених даній тематиці, описується підхід до стеганографічного приховування інформації, заснований на внесенні малопомітних низькочастотних деформуючих спотворень у фрагменти повнокольорових зображень-контейнерів та використанні кореляції між колірними каналами для подальшого відновлення схованих даних. Відмінність сучасних підходів полягає у застосуванні нестандартних способів модифікації контейнера, які не мають чітко вираженого адитивного або мультиплікативного характеру, а ґрунтуються на плавній еластичній деформації непересічних фрагментів зображення. Такий підхід передбачає подання цифрового зображення у вигляді неперервної функції просторових координат та використання спеціальних процедур апроксимації, що дозволяє здійснювати нецілочисельні зсуви опорних точок деформованих областей [7-9].

Реалізація зазначеного підходу ускладнює виявлення факту стеганографічного приховування інформації як за допомогою відомих статистичних стеганоаналітичних атак, орієнтованих на класичні схеми стеганографії, так і під час первинного візуального аналізу. У ряді алгоритмів процедура відновлення прихованої інформації базується на аналізі порушеної кореляції колірних компонентів фрагментів контейнера та реалізується із

залученням апарата штучних нейронних мереж [10, 11]. Їх адаптивність і здатність до донавчання для роботи з графічними контейнерами з різними статистичними характеристиками дозволяє розглядати нейронні мережі як універсальний інструмент стеганографічного декодування.

Аналіз сучасних публікацій свідчить [3, 12, 16-18,] про високу актуальність використання нейронних мереж у задачах стеганографії, де вони застосовуються як засоби прийняття рішень щодо наявності прихованої інформації, як класифікатори контейнерів, а також як декодери прихованих повідомлень.

1.2 Мета дослідження

Метою дипломної роботи є розробка нового алгоритму утаємничування інформації у графічних файлах, який забезпечує підвищену стійкість до сучасних методів стеганоаналізу та мінімізує ймовірність виявлення факту прихованого передавання даних. Актуальність поставленої мети зумовлена постійним удосконаленням методів стеганоаналітичного аналізу, зокрема статистичних і сигнатурних, які здатні ефективно виявляти класичні схеми приховування інформації, що базуються на модифікації найменших значущих бітів або частотних коефіцієнтів зображень. У зв’язку з цим виникає необхідність створення нових підходів до вбудовування інформації, орієнтованих на збереження візуальної якості графічного контейнера та його статистичних характеристик.

Досягнення поставленої мети передбачає проведення комплексного аналізу ефективності існуючих методів стеганографічного приховування інформації у графічних файлах і дослідження можливостей їх виявлення за допомогою відомих методів стеганоаналізу. На основі отриманих результатів має бути розроблено новий алгоритм утаємничування інформації, принципи

роботи якого враховують характерні особливості просторового та частотного подання зображень і спрямовані на зменшення статистичної помітності змін контейнера. Реалізація такого алгоритму повинна ускладнювати застосування типових стеганоаналітичних атак та підвищувати загальний рівень інформаційної безпеки при прихованому передаванні даних у графічних контейнерах.

1.3 Стеганоаналіз методом Хі-квадрат

Метод атаки на стеганосистеми, заснований на аналізі критерію Хі-квадрат, був запропонований Андросом Вестфельдом і детально описаний Андросом Фрідріхом у 1999 році. Даний метод орієнтований насамперед на виявлення інформації, прихованої за допомогою методу найменших значущих розрядів (НЗР), який є одним із найпоширеніших способів стеганографічного вбудовування у графічних файлах [12].

Основним елементом аналізу в межах даного підходу є так звані пари значень (ПЗ або PO – Pair of Values), під якими розуміють пари байтів, що кодують інтенсивності кольору пікселів і відрізняються між собою лише на один найменший значущий біт. Метод НЗР фактично здійснює модифікацію значень усередині таких пар, змінюючи, за необхідності, початкове значення байта на суміжне значення з тієї ж пари, не виходячи за межі допустимого діапазону інтенсивностей.

Ідея методу Хі-квадрат ґрунтується на статистичному припущенні про характер розподілу інтенсивностей у незаповненому контейнері. Для зображення, в яке не було вбудовано приховане повідомлення, імовірність одночасної появи обох значень кожної пари є малою, унаслідок чого кількість поява значень, що відрізняються на один НЗР, суттєво різниться. Іншими словами, у незаповненому контейнері для кожної пари значень спостерігається

значна асиметрія між частотами появи її елементів. Кількість появ конкретного значення інтенсивності називається його частотою і є базовою статистичною характеристикою, що використовується в подальшому аналізі.

Як теоретично очікуваний розподіл у методі Хі-квадрат використовується послідовність середніх арифметичних значень частот для всіх пар значень. Оскільки в процесі приховування інформації методом НЗР відбувається лише перерозподіл частот усередині кожної пари, загальна сума частот пари залишається незмінною. Відповідно, середнє арифметичне значення частот елементів пари також зберігає сталу величину як для заповненого, так і для незаповненого контейнера.

Під вибіркою, що спостерігається, у даному контексті розуміється послідовність, яка складається або лише з парних, або лише з непарних значень усіх пар інтенсивностей. Такий поділ зумовлений необхідністю коректного порівняння емпіричного розподілу з теоретично очікуваним у рамках обчислення критерію Хі-квадрат. Визначальною характеристикою при цьому є різниця між середнім значенням частот пари та фактичними частотами кожного з її елементів.

Таким чином, теоретично очікуваний розподіл, сформований на основі середніх значень частот пар, існує як для заповнених, так і для незаповнених контейнерів. Ступінь відповідності між спостережуваним та очікуваним розподілами слугує мірою ймовірності наявності стеганографічного вбудовування. Якщо значення критерію Хі-квадрат свідчить про незначні відхилення від теоретично очікуваного розподілу, це означає, що з високою ймовірністю у зображенні відбулося приховане вбудовування інформації.

Ефективність методу Хі-квадрат значно зростає у випадку застосування його не до всього зображення загалом, а до окремих його фрагментів. Найчастіше зображення розбивають або на умовні рядки матриці пікселів, або на блоки, розмір яких становить приблизно один відсоток від загальної площі зображення. Такий підхід дозволяє не лише виявити сам факт приховування, але й оцінити приблизні межі послідовно вбудованого повідомлення.

Хоча найменш помітні для людського зору спотворення зазвичай пов’язані зі змінами у синьому колірному каналі, методи приховування інформації в НЗР дозволяють здійснювати вбудовування одночасно у всі три колірні канали. Це пояснюється низькою чутливістю зорової системи людини до незначних змін кольору, спричинених інверсією молодших розрядів пікселів.

У зв’язку з цим доцільним є обчислення середнього значення ймовірності наявності прихованої інформації за всіма трьома колірними каналами для кожного аналізованого блоку зображення. Навіть у випадку, коли вбудовування здійснювалося лише в одному каналі, усереднене значення ймовірності суттєво відрізнятиметься від нуля, що дозволяє зробити обґрунтований висновок про наявність прихованого повідомлення у відповідному фрагменті графічного контейнера.

1.4 RS-метод стеганоаналізу

Регулярно-сингулярний метод (RS-метод) виявлення стеганографічно схованих повідомлень був запропонований Андросом Фрідріх, Джезікою Фрідріх та Мирославом Гол’яном у 2001 році і набув широкого застосування як ефективний інструмент аналізу приховування інформації, здійсненого методом найменших значущих розрядів. Даний метод ґрунтується на статистичному аналізі локальних структур зображення та дозволяє виявляти приховані повідомлення навіть у випадку псевдовипадкового розподілу змінених пікселів.

Основою RS-методу є поділ зображення на непересічні групи з n суміжних пікселів, де n є парним числом. Такий поділ забезпечує однозначність подальшої класифікації груп і дозволяє виконувати порівняльний аналіз їх властивостей до та після спеціальних операцій обернення. Кожна група пікселів розглядається як локальна структура зображення, що характеризується певним

рівнем однорідності або, навпаки, різкості переходів між сусідніми значеннями інтенсивності.

Для кількісної оцінки цих властивостей у RS-методі вводиться функція регулярності – числова функція, яка ставить у відповідність кожній групі пікселів одне дійсне значення та відображає ступінь її регулярності. Інтуїтивно функція регулярності характеризує «гладкість» або «шумність» групи: чим більшими є перепади значень інтенсивностей між сусідніми пікселями, тим менш регулярною і більш «гучною» вважається відповідна група.

Як функцію регулярності зазвичай обирається сума абсолютних різниць між значеннями інтенсивностей сусідніх пікселів у межах однієї групи. Такий вибір зумовлений простотою обчислення та високою чутливістю цієї характеристики до локальних змін яскравості, що виникають у результаті інверсії найменших значущих розрядів. Використання суми абсолютних різниць дозволяє ефективно виявляти навіть незначні порушення локальної структури зображення, які не сприймаються людським оком, але призводять до зміни статистичних властивостей піксельних груп.

Таким чином, RS-метод базується на припущенні, що стеганографічне вбудовування інформації в НЗР призводить до систематичних змін регулярності груп пікселів, які можна виявити шляхом порівняння значень функції регулярності до та після спеціально визначених операцій обернення. Подальша класифікація груп на регулярні, сингулярні та непридатні дозволяє оцінити не лише факт приховування інформації, але й отримати кількісну оцінку відносного обсягу вбудованого повідомлення:

$$f(G) = f(g_1, g_2, \ldots, g_n) = \sum_{i=1}^{n-1} |g_{i+1} - g_i|, \qquad (1.1)$$

де G – група пікселів;

g_i – і-й елемент групи пікселів G;

n – кількість пікселів у групі.

Після підрахунку значень функції регулярності для всіх груп аналізованого зображення визначається група функцій обернення. Ці функції відповідають наступному набору властивостей:

$$F(F(x)) = x, x \in P, \tag{1.2}$$

де P = 0, 1, …, 255,

$$F_{ПР}: 0 \to 1,1 \to 0,2 \to 3,\ldots,254 \to 255, \tag{1.3}$$

$$F_{ПР}: -1 \to 0,1 \to 2,\ldots,255 \to 256. \tag{1.4}$$

Група функцій обернення складається з:

– прямої $F_{пр}$;

– зворотної $F_{зв}$;

– нульової F_0.

Застосування функцій обернення імітує додавання оборотного шуму, підсилюючи сплески значень у групі і зменшуючи її регулярність [13].

Для застосування даних функцій, до значень пікселів групи вводиться маска, що описує групу функцій обернення, які було застосовано до групи пікселів.

Маска – це група з n значень, кожне з яких вибирається серед чисел:

а) -1;

б) 0;

в) 1.

Кожне з них кодує одну з трьох функцій обернення:

– значення -1 відповідає функції $F_{зв}$;

– значення 0 відповідає функції F_0;

– значення 1 відповідає функції $F_{пр}$.

До пікселю групи за проведення обернення, таким чином, застосовується

відповідна йому закодована в масці функція обернення.

Після застосування функцій обернення до групи здійснюється порівняння значень функції регулярності зі значеннями до обернення.

На основі даного порівняння група відноситься до одного з класів:

– звичайні (або регулярні):

$$f(F(G)) > f(G), \tag{1.5}$$

де $f \in R$;

– незвичайні (або сингулярні):

$$f(F(G)) < f(G), \tag{1.6}$$

де $f \in S$,

– непридатні (тобто, які не можуть застосовуватися):

$$f(F(G)) = f(G), \tag{1.7}$$

де $f \in U$.

Для кожної групи обернення здійснюється двічі:

– з прямою маскою;

– з інвертованою маскою.

Після проведення операцій класифікації для всіх груп виконується підрахунок ряду кількісних характеристик:

– S_M – кількість незвичайних груп для маски M;

– R_M – кількість звичайних груп для маски M,

– S_{-M} – кількість незвичайних груп для інверсної маски -M;

– R_{-M} – кількість звичайних груп для інверсної маски -M.

Всі описані характеристики задаються як відносні величини, тобто у відсотках від загального числа груп k. Таким чином: $R_M+S_M<=1$, $R_{-M}+S_{-M}<=1$.

Основною гіпотезою даного методу є припущення про те, що в незаповненому контейнері кількості груп одного класу для звичайної й інверсної маски рівні:$R_M = R_{-M}, S_M = S_{-M}$.

На рисунку 1.1 наведено типовий вид RS-діаграми – графіків значень:

– R_M;

– S_M;

– R_{-M};

– S_{-M}

у залежності від кількості пікселів з інвертованими НЗР у зображенні [17].

Вісь x – це відносна кількість пікселів з перевернутими молодшими розрядами, вісь y – відносна кількість регулярних та сингулярних груп з масками M та -M, відповідно, де M = [0110].

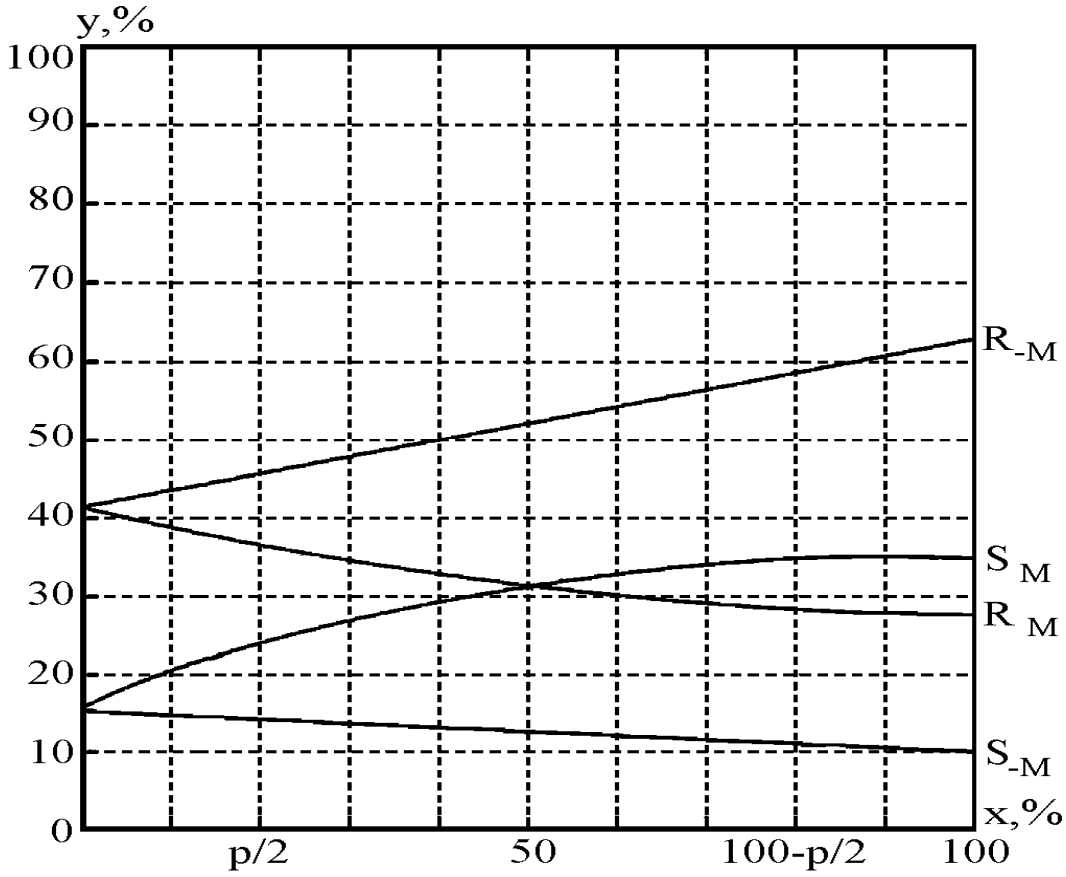


Рисунок 1.1 – RS-діаграма

Під величиною p на рисунку 1.1 і надалі розуміється відсоток заповнення стеганоконтейнера схованим повідомленням (відносна довжина повідомлення). На осі абсцис відмірюється відсоток пікселів з інвертованими НЗР, на осі ординат – відсоток груп звичайних і незвичайних класів прямої й інвертованої масок (від загальної кількості груп).

На основі інформації про типову поведінку залежностей кількісних характеристик груп здійснюються подальші обчислення, що дозволяють оцінити відносну довжину схованого повідомлення.

Якщо відносна довжина схованого повідомлення дорівнює p, то, оскільки p є випадковою течією розрядів, у середньому в зображенні інвертується p/2 найменших значущих розрядів.

У такому випадку, 100% – заповнювання стеганоконтейнера призведе до того, що p/2=50%. Інверсія половини найменших значущих розрядів означає, що різниця між кількістю звичайних і незвичайних груп зведеться до нуля. Тоді: $R_M = S_M$.

Вимірювання чисельних характеристик груп відповідають крапкам RS-діаграми:

– $R_M(p/2)$;

– $S_M(p/2)$;

– $R_{-M}(p/2)$;

– $S_{-M}(p/2)$.

Обчислені чисельні характеристики груп для того ж зображення після інвертування всіх його НЗР будуть відповідати крапкам:

– $R_M(1-p/2)$;

– $S_M(1-p/2)$;

– $R_{-M}(1-p/2)$;

– $S_{-M}(1-p/2)$.

Далі RS-метод передбачає апроксимацію кривих, що проходять крапками:

– $R_{-M}(p/2)$;

– $R_M(1\text{-}p/2)$;

– $S_{\text{-}M}(p/2)$;

– $S_M(1\text{-}p/2)$,

відповідно, прямими лініями.

Криві, що проходять через крапки:

– $R_M(p/2)$;

– $R_M(1\text{-}p/2)$;

– $S_M(p/2)$;

– $S_M(1\text{-}p/2)$,

апроксимуються квадратичними параболами з урахуванням наявності крапок перетинання прямих і парабол: відповідні одній масці парабола і пряма мають крапку перетинання на осі ординат RS-діаграми, а параболи перетинаються при $p/2 = 50\%$.

Для оцінки відносної довжини повідомлення p треба розв'язати систему 11 рівнянь з 11 невідомими:

– по два коефіцієнти для кожної з двох прямих;

– по три коефіцієнти для кожної з двох парабол;

– значення p.

Система містить у собі:

– 8 рівнянь прямих чи парабол для 8 знайдених раніше крапок;

– 2 рівняння крапок перетинань парабол з відповідними прямими;

– 1 рівняння для крапки перетинання парабол.

Розв'язання системи:

$$2 \cdot (d_1 + d_0) \cdot x^2 + (d_{-0} - d_{-1} - d_1 - 3 \cdot d_0) \cdot x + d_0 - d_{-0} = 0, \quad (1.8)$$

дозволяє знайти оцінку значення p [6]:

$$p = \frac{x}{x - \frac{1}{2}}, \quad (1.9)$$

де $d_0 = R_M\left(\frac{p}{2}\right) - S_M\left(\frac{p}{2}\right)$;

$d_{-0} = R_{-M}\left(\frac{p}{2}\right) - S_{-M}\left(\frac{p}{2}\right)$;

$d_1 = R_M\left(1 - \frac{p}{2}\right) - S_M\left(1 - \frac{p}{2}\right)$;

$d_{-1} = R_{-M}\left(1 - \frac{p}{2}\right) - S_{-M}\left(1 - \frac{p}{2}\right)$.

Ключова особливість RS-методу полягає в тому, що він аналізує кількісні характеристики невеликих груп пікселів. У зв'язку з чим він, хоча і не здатний визначати область потенційного вбудовування, але може знайти приховування, зроблене у випадкові біти, а не послідовно.

1.5 Метод Коха-Жао

Даний вид аналізу призначений для виявлення вбудовування до зображення-контейнера повідомлення за методом Коха-Жао. Цей метод шукає інформацію, закодовану в частотному поданні зображення.

Подавання зображення в частотній області формується шляхом обчислення коефіцієнтів дискретного косинусного перетворення (ДКП), для чого виконуються наступні операції [14]:

– зображення розбивається на блоки розміром 8х8 пікселів;

– після чого над кожним блоком здійснюється двовимірне дискретне косинусне перетворення;

– формується матриця з 64-х коефіцієнтів.

В отриманій матриці коефіцієнт у лівому верхньому куті, що відповідає нульовій частоті (елемент матриці з індексами (0; 0)), називається DC-коефіцієнтом. Він визначає основний колірний відтінок (середню інтенсивність кольору) усього блоку. Всі інші отримані коефіцієнти називаються AC-

коефіцієнтами і виражають частоту зміни інтенсивності кольору по різних напрямках обраного блоку (по горизонталі і вертикалі) [15].

Таким чином, кожна матриця коефіцієнтів дискретного косинусного перетворення поділяється на три підмножини (від правого нижнього до лівого верхнього кута матриці):

– високочастотні;

– середньочастотні;

– низькочастотні.

Низькочастотні коефіцієнти мають більший вплив на інтенсивність кольору пікселів. У зв'язку з цим будь-які зміни і перетворення над коефіцієнтами ДКП здійснюються в:

– високочастотній;

– середньочастотній областях.

Однією з найважливіших задач, що вимагають розв’язання при спробі виявити вбудовування методом Коха-Жао, є отримання висновку про те, за рахунок яких коефіцієнтів ДКП відбувалося вбудовування. Оскільки застосування методу Коха-Жао припускає приховування інформації в одному з наборів середньочастотних компонент, основні операції аналізу виконуються для кожного з таких наборів окремо [16].

Оскільки кодування біт приховуваного повідомлення здійснюється за рахунок різниці абсолютних значень обраних коефіцієнтів, спершу варто обчислити абсолютні значення даних різниць для всіх блоків зображення:

$$C_i = \big||D_i(k_1, k_2)| - |D_i(k_2, k_1)|\big|, \qquad (1.10)$$

де $D_i(x,y)$ – значення коефіцієнта ДКП з індексами (x, y) у і-блоці.

На даному етапі обчислюється не просто різниця абсолютних значень коефіцієнтів, а їхні абсолютні значення. Це пов'язано з тим, що кодування розрядів визначається подоланням граничного значення P для нуля і -P – для одиниці [9, 20].

Незважаючи на флуктуації величини пікових значень C_i, у блоках, що не використовувалися для кодування біт схованого повідомлення (рис. 1.2), реально використані для вбудовування блоки відрізняє неперервна послідовна ділянка порівняно контрастних значень C_i (рис. 1.3).

На рисунках 1.2, 1.3 наведено значення коефіцієнтів C_i, для одного і того самого контейнера:

– порожнього (рис. 1.2);

– заповненого (рис. 1.3).

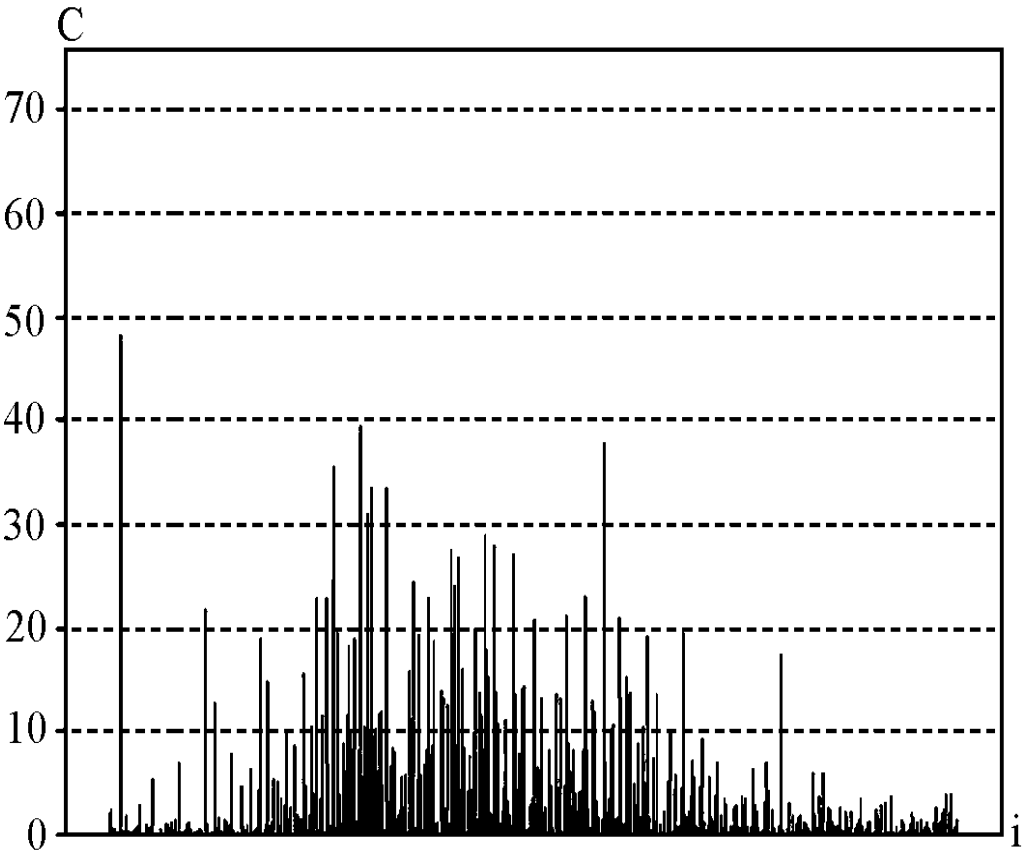


Рисунок 1.2 – Значення коефіцієнтів C порожнього контейнера

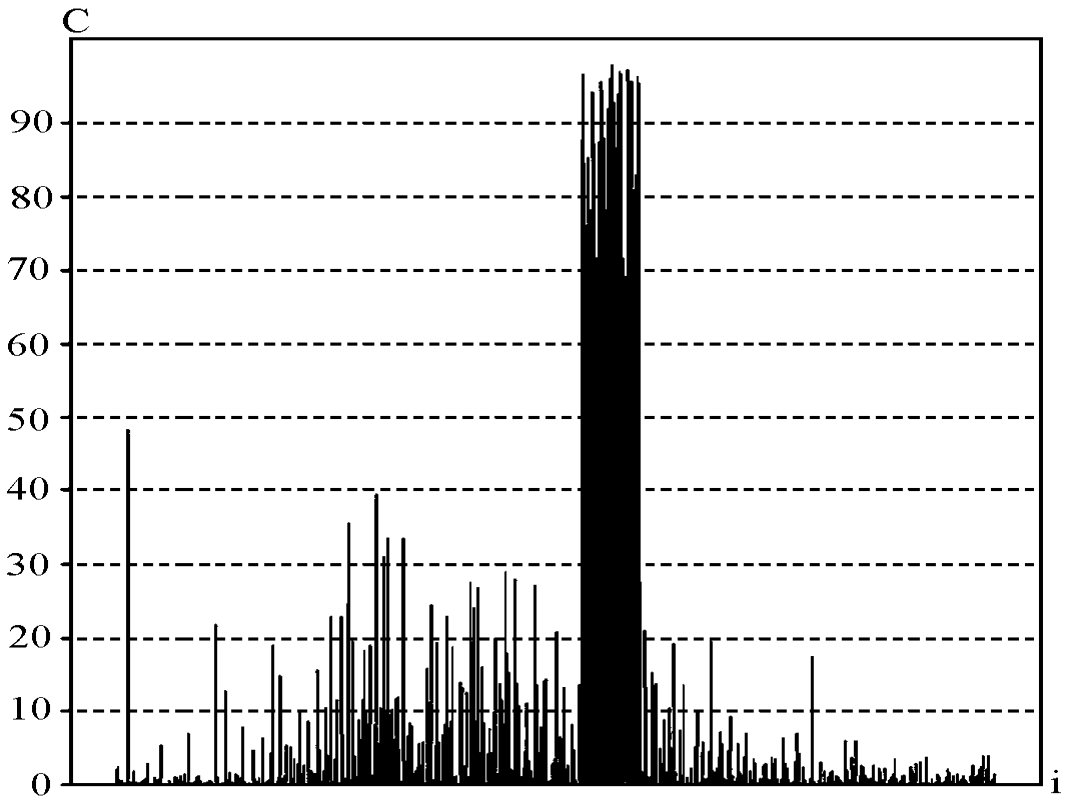


Рисунок 1.3 – Значення коефіцієнтів С заповненого контейнера

На рисунках 1.2, 1.3:

– за віссю абсцис розташовано індекси блоків;

– за віссю ординат розташовано значення C_i, що відповідають номерам блоків.

Для отримання меж області вбудовування укладається послідовність модулів різниць значень коефіцієнтів, де два найбільших значення мусять відповідати межам області вбудовування.

Однак за великих розмірів зображення виявлення схованого повідомлення малого розміру (а також, особливо, закодоване з малим значенням порога) може бути значно утруднене за рахунок випадкових пікових значень послідовності C_i. Це може відбуватися за рахунок:

– використання графічних форматів, що допускають стискання;

– розходжень у точності обчислення ДКП на пристроях кодування й аналізування;

– можливих спотворювань і втрат за передавання зображень.

Для того, щоб зменшити імовірність збоїв роботи алгоритму визначення області вбудовування через подібний "шуму", варто зробити попередній аналіз значень коефіцієнтів з урахуванням того факту, що убудоване повідомлення відповідає неперервній ділянці контрастних значень. Тобто, знайти найбільш довгу таку ділянку.

Таким чином:

– висновок про наявність убудованого повідомлення може бути зроблений;

– межі схованого повідомлення в значеннях індексів блоків можуть бути визначені.

Для обчислення граничного значення кодування варто знайти мінімальне зі значень C_i у виявленій області вбудовування. За даним значенням можна зробити видобування інформації.

Описані процедури аналізу мусять бути проведені для кожної з передбачуваних пар коефіцієнтів ДКП. Як правило, це наступні набори пар коефіцієнтів [8]:

– (3;4) і (4;3);

– (3;5) і (5;3);

– (4;5) і (5 ;4).

З отриманих результатів варто вибирати той, що відповідає найбільшому з виявлених значень порога кодування P.

Для оцінки алгоритму виявлення вбудованих повідомлень за методом Коха-Жао, визначалися кількості:

– коректно розпізнаних зображень, для яких алгоритм правильно оцінив розмір схованого повідомлення;

– некоректно розпізнаних, для яких алгоритм повернув неправильне, але відмінне від нуля значення розміру повідомлення;

– нерозпізнаних, для яких алгоритм ухвалив про відсутність вбудовування, у той час, як у зображенні дійсно сховано інформацію.

Результати обчислень наведено у таблиці 1.1.

Таблиця 1.1 – Результати аналізу приховування за методом Коха-Жао

Вид результату	Відносна кількість
Коректно розпізнані	73%
Некоректно розпізнані	16,5%
Не розпізнані	10,5%

Як видно з таблиці, більш, ніж у 70% випадків алгоритм зміг коректно визначити розмір убудованого повідомлення, а виходить, з великою часткою імовірності, його можна витягти. У найбільшому числі інших випадків розмір оцінений неправильно, що не дозволяє витягти сховані дані (принаймні, без ручного втручання), однак дозволяє зробити однозначний висновок про наявність приховування в частотній області.

Не розпізнаними зовсім залишилися 10,5% зображень, що може говорити про низький обраний поріг вбудовування в них чи інформації їхній високий зашумленності, що викликає сплески значень різниць коефіцієнтів матриці дискретного косинусного перетворення, що ускладнюють ідентифікацію області вбудовування.

Аналіз результатів тестування також показав, що в жодному з випадків стеганоаналіз Коха-Жао не виявив приховування в просторовій області зображення (у найменших значущих бітах), а методи Хі-квадрат і RS не могли показати результат, що дозволяє установити факт вбудовування в подавання зображення в частотній області.

2 АНАЛІЗ ТА ДОСЛІДЖЕННЯ МЕТОДІВ СТЕГАНОАНАЛІЗУ

2.1 Умови тестування ефективності методів

Тестування ефективності методів стеганоаналізу здійснювалося шляхом комплексного аналізу набору файлів зображень, у яких приховування інформації реалізовувалося з використанням методу найменших значущих розрядів у двох основних варіаціях: псевдовипадковій та послідовній. У межах дослідження формувалися як зображення-контейнери з вбудованим інформаційним повідомленням, так і контрольні зразки, у які жодної інформації не вбудовували. Такий підхід дозволив оцінити не лише здатність методів виявляти приховування, але й рівень їх помилкових спрацьовувань на порожніх контейнерах.

Для кожного графічного файлу аналіз проводився з використанням усіх реалізованих у роботі методів стеганоаналізу, а саме методу Хі-квадрат та регулярно-сингулярного (RS) методу. Це забезпечило можливість прямого порівняння результатів роботи методів за однакових вхідних умов і дозволило дослідити їх чутливість до різних способів вбудовування інформації. Аналіз виконувався як для зображень із послідовним розміщенням змінених найменших значущих розрядів, так і для зображень із псевдовипадковим розподілом модифікованих пікселів.

Результати роботи кожного з методів визначалися у вигляді оцінки довжини прихованого в зображенні інформаційного повідомлення. Для методу Хі-квадрат і RS-методу ці оцінки мали різну форму подання: у відносному вигляді для статистичних оцінок та в абсолютному або умовно-нормованому вигляді – для кількісних характеристик, отриманих у межах RS-аналізу. Отримані значення порівнювалися з реальними обсягами інформаційних повідомлень, вбудованих у відповідні зображення-контейнери.

Для узагальнення результатів тестування та кількісної оцінки точності кожного методу обчислювалося усереднене відхилення результатів аналізу. Як

міра похибки використовувалося середньоарифметичне значення модулів різниць між алгоритмічно визначеним обсягом прихованого повідомлення та його фактичною довжиною. Такий показник дозволив об’єктивно оцінити точність методів стеганоаналізу, а також виявити залежність їх ефективності від способу вбудовування та ступеня заповнення контейнера.

Застосована методика тестування забезпечила відтворюваність результатів і створила основу для подальшого порівняльного аналізу методів Хі-квадрат і RS, а також для формування висновків щодо доцільності їх спільного використання при виявленні прихованої інформації у графічних файлах.

2.2 Програмна реалізація методів стеганоаналізу

Для відлагоджування методу Хі-квадрат використовувався наступний текст програми:

```
from PIL import Image
import numpy as np
from scipy.stats import chi2
def f_hi_kvadrat(mal_path, kanal='gray', d=0.05):
    """
    Аналіз зображення за критерієм хі-квадрат
    :param mal_path: шлях до зображення
    :param kanal: 'gray', 'r', 'g', 'b'
    :param d: рівень значущості
    :return: (value, p, z)
    """
    mal = mal.open(mal_path)
```

```
    # Вибір каналу
    if kanal == 'gray':
        mal1 = mal.convert('L')
        data = np.array(mal).flatten()
    else:
        mal1 = mal.convert('RGB')
        kanal_index = {'r': 0, 'g': 1, 'b': 2}[kanal]
        data = np.array(mal1)[:, :, kanal_index].flatten()
    mas = np.bincount(data, minlength=256)
    stat = 0.0
    pr = 0
    for i in range(0, 256, 2):
        n_0 = mas[i]
        n_1 = mas[i + 1]
        m = (n_0 + n_1) / 2
        if m > 0:
            stat += ((n_0 - m) ** 2) / m
            stat += ((n_1 - m) ** 2) / m
            pr += 1
    p = 1 - chi2.cdf(stat, pr)
    z = p > d
    return stat, p, z
if __name__ == "__main__":
    mal_path = "test_mal.png"
    value, p, z = f_hi_kvadrat(
        mal_path,
        kanal='gray',
        d=0.05
    )
    print(f"   = {value:.2f}")
```

```
    print(f"p = {p:.4f}")
    if z:
        print("+")
    else:
        print("-")
```

Для відлагоджування регулярно-сингулярного методу використовувався наступний текст програми:

```
from PIL import Image
import numpy as np
def gladk(grupa):
    """
    Функція гладкості
    """
    return np.sum(np.abs(np.diff(grupa)))
def obern_nzr(grupa, maska):
    """
    обернення відповідно до маски
    maska: [1, -1, 0]
    """
    obern = grupa.copy()
    for i, m in enumerate(maska):
        if m == 1:
            obern[i] = grupa[i] ^ 1       # обернення НЗР
        elif m == -1:
            obern[i] = grupa[i] ^ 1
    return obern
def f_grupa(grupa, maska):
    """
```

```
    Класифікація групи як R, S або U
    """
    pryam = gladk(grupa)
    obern = gladk(obern_nzr(grupa, maska))

    if obern > pryam:
        return 'R'
    elif obern < pryam:
        return 'S'
    else:
        return 'U'
def f_RS(mal_path, block_size=4):
    """
    RS-стеганоаналіз
    """
    mal = image.open(mal_path).convert('L')
    kr = np.array(mal).flatten()
    # Маски обернень
    maska_p = [1, 0, 1, 0]
    maska_m = [-1, 0, -1, 0]
    R_p = S_p = R_m = S_m = 0
    n = 0
    for i in range(0, len(kr) - block_size, block_size):
        grupa = kr[i:i + block_size]
        if len(grupa) < block_size:
            continue
        cls_p = f_grupa(grupa, maska_p)
        cls_m = f_grupa(grupa, maska_m)
        if cls_p == 'R':
            R_p += 1
```

```
            elif cls_p == 'S':
                S_p += 1
            if cls_m == 'R':
                R_m += 1
            elif cls_m == 'S':
                S_m += 1
            n += 1
        return {
            'R+= ': R_p,
            'S+= ': S_p,
            'R-= ': R_m,
            'S-= ': S_m,
            'Разом: ': n
        }
    if __name__ == "__main__":
        mal_path = "test_mal.png"
        result = f_RS(mal_path)
        print("RS-аналіз:")
        for k, v in result.items():
            print(f"{k}: {v}")
```

2.3 Тестування методів Хі-квадрат і RS

Отримані в ході тестування чисельні результати дозволяють зробити більш детальний аналіз особливостей роботи методів Хі-квадрат та регулярно-сингулярного методу залежно від способу вбудовування та ступеня заповнення графічного контейнера. Застосування однакової тестової вибірки зображень для обох методів забезпечило коректність порівняння та дозволило виключити

вплив сторонніх факторів, пов’язаних із властивостями конкретних файлів.

Аналіз результатів, наведених у таблиці 2.1, свідчить про суттєву відмінність у чутливості методів до характеру модифікації найменших значущих розрядів. Метод Хі-квадрат демонструє високу ефективність у випадку послідовного вбудовування інформації, що пояснюється значним вирівнюванням частот пар значень інтенсивностей, яке добре фіксується статистичним критерієм Пірсона. Водночас за псевдовипадкового вибору пікселів контейнера цей ефект проявляється значно слабше, що призводить до істотного зростання похибки оцінки обсягу прихованого повідомлення.

Таблиця 2.1 – Відхилення довжини інформаційного повідомлення, визначеної за допомоги методів Хі-квадрат і RS

Назва методу	Спосіб вбудовування інформації		
	без вбудовування, %	послідовний найменший значущий розряд, %	псевдовападковий найменший значущий розряд, %
Хі-квадрат	0,1	1,5	23
RS	5	23	4,5

Регулярно-сингулярний метод, навпаки, виявляє високу чутливість саме до псевдовипадкового вбудовування інформації. Це пов’язано з тим, що випадкова інверсія НЗР порушує локальну регулярність груп пікселів, що безпосередньо впливає на співвідношення кількості регулярних і сингулярних груп для прямої та інверсної масок. У випадку ж послідовного вбудовування зміни мають більш детермінований характер і значною мірою компенсуються при груповому аналізі, що зумовлює зростання похибки оцінювання. Додатковий аналіз залежностей, представлених на рисунках 2.1-2.6, дозволяє детально оцінити характер похибок методів Хі-квадрат та RS залежно від способу вбудовування інформації та ступеня заповнення графічного контейнера.

На рисунку 2.1 наведено помилку ΔL оцінки розміру інформаційного

повідомлення в залежності від ступеня заповнення контейнера ΔV для методу Хі-квадрат у випадку послідовного заповнення найменших значущих розрядів зображення. З наведеного графіка видно, що за такого способу вбудовування метод Хі-квадрат демонструє стабільно високу точність на більшій частині досліджуваного діапазону значень ΔV. Зі зростанням відносної довжини прихованого повідомлення похибка оцінки змінюється незначно та залишається на низькому рівні, що пояснюється характерним для послідовного НЗР-вбудовування вирівнюванням частот пар значень інтенсивностей пікселів. Найбільші відхилення спостерігаються в області середніх значень ступеня заповнення контейнера, де статистичні зміни частково компенсуються, однак навіть у цьому випадку абсолютна величина похибки не перевищує допустимих значень. Отримані результати свідчать про доцільність застосування методу Хі-квадрат для кількісної оцінки обсягу прихованого повідомлення у разі послідовного модифікування найменших значущих розрядів графічного контейнера.

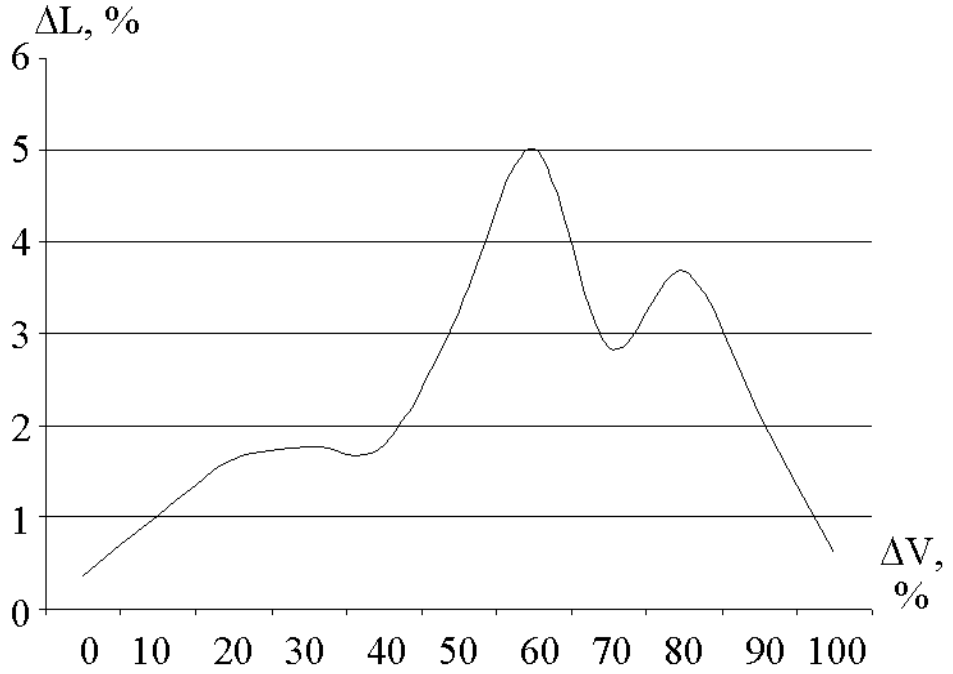


Рисунок 2.1 – Помилка оцінки розміру інформаційного повідомлення в залежності від ступеню заповнення контейнеру для методу Хі-квадрат у випадку послідовного заповнення найменших значущих розрядів

На рисунку 2.2 наведено помилку ΔL оцінки розміру інформаційного повідомлення в залежності від ступеня заповнення контейнера ΔV для методу Хі-квадрат у випадку псевдовипадкового заповнення найменших значущих розрядів зображення.

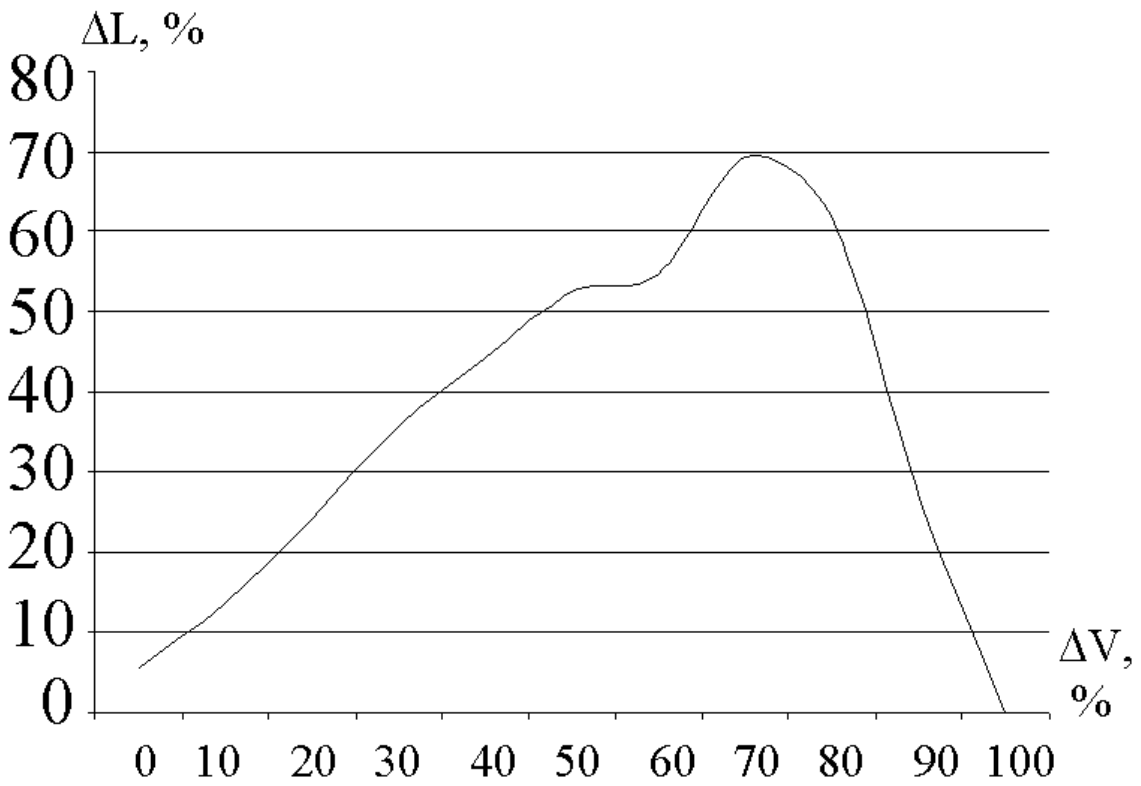


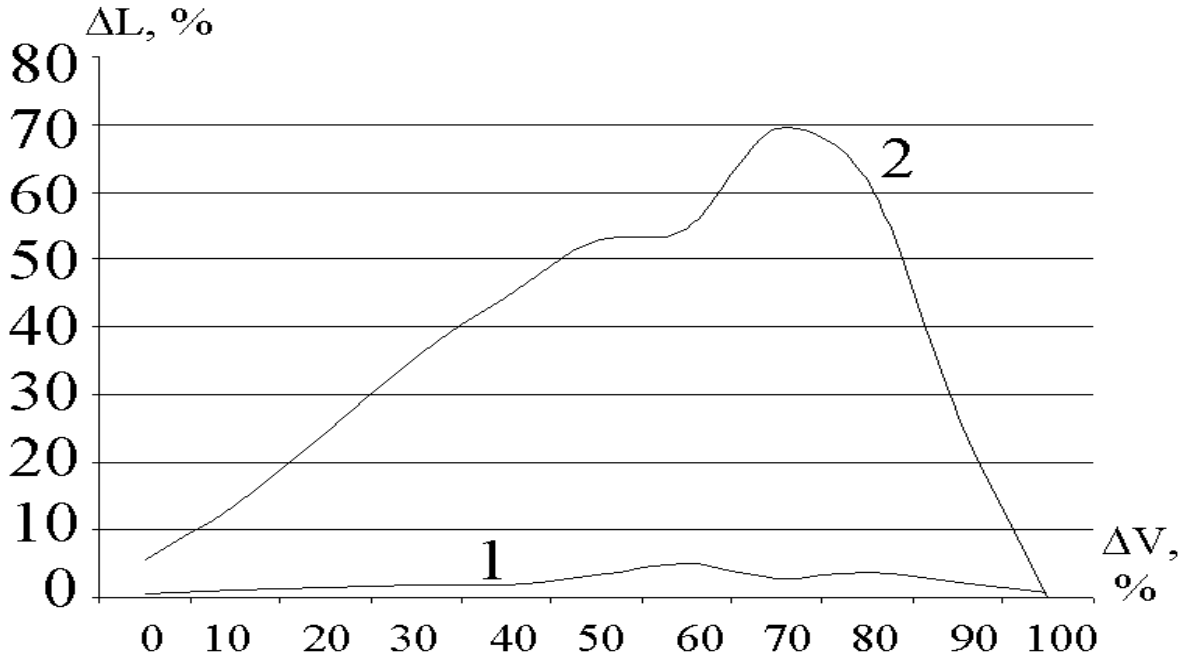
Рисунок 2.2 – Помилка оцінки розміру інформаційного повідомлення в залежності від ступеню заповнення контейнеру для методу Хі-квадрат у випадку псевдовипадкового заповнення найменших значущих розрядів

З рисунку видно, що помилка визначення обсягу інформаційного повідомлення, схованого у псевдовипадково розташованих байтах зображення, методом Хі-квадрат зростає практично прямопропорційно до збільшення довжини прихованого повідомлення. Найбільших значень, що досягають приблизно 70%, похибка набуває для файлів-контейнерів, заповнених на 70-80%, що зумовлено відсутністю характерного для послідовного вбудовування вирівнювання частот пар значень інтенсивностей пікселів. За псевдовипадкового вибору позицій для модифікації найменших значущих

розрядів статистичні зміни в контейнері мають менш виражений і нерівномірний характер, унаслідок чого ефективність критерію Хі-квадрат суттєво знижується.

Подальше зменшення похибки після заповнення понад трьох чвертей обсягу контейнера пояснюється майже повною інверсією найменших значущих розрядів у зображенні, що призводить до повторного вирівнювання частот пар значень інтенсивностей. У цьому випадку статистичні характеристики контейнера знову наближаються до теоретично очікуваних для повністю заповненого зображення, що частково відновлює чутливість методу Хі-квадрат. Водночас навіть за таких умов точність оцінки залишається істотно нижчою, ніж у випадку послідовного вбудовування, що свідчить про обмежену придатність методу Хі-квадрат для кількісного аналізу псевдовипадкового НЗР-приховування.

Для зручності порівняння ефективності різновидів методу Хі-квадрат на рисунку 2.3 наведено відповідні залежності помилки оцінки розміру прихованого повідомлення від ступеня заповнення контейнера для послідовного та псевдовипадкового способів вбудовування. Спільне подання цих залежностей дозволяє наочно проаналізувати вплив характеру модифікації найменших значущих розрядів на точність роботи методу Хі-квадрат та виявити його сильні й слабкі сторони. З аналізу наведених залежностей випливає, що точність визначення обсягу інформаційного повідомлення, вбудованого послідовним способом, залишається досить високою для методу Хі-квадрат практично на всьому досліджуваному діапазоні значень ступеня заповнення контейнера. Навіть за значних обсягів прихованого повідомлення похибка оцінки є невеликою, а найбільше відхилення, яке становить близько 5%, спостерігається для контейнерів, заповнених на 60-70%. Така поведінка пояснюється особливостями статистичного вирівнювання частот пар значень інтенсивностей пікселів, характерного для послідовного НЗР-вбудовування.



1 – випадок послідовного заповнення найменших значущих розрядів;

2 – випадок псевдовипадкового заповнення найменших значущих розрядів.

Рисунок 2.3 – Помилка оцінки розміру інформаційного повідомлення в залежності від ступеню заповнення контейнеру для методу Хі-квадрат

Водночас у випадку псевдовипадкового приховування інформації метод Хі-квадрат демонструє принципово інші результати. За повного заповнення контейнера, коли приховане повідомлення становить 100% його обсягу, метод дозволяє з украй високою точністю визначити розмір вбудованої інформації. При ступені заповнення дещо меншому за 100%, але більшому за 90%, метод Хі-квадрат здатний однозначно ідентифікувати факт вбудовування, однак похибка кількісної оцінки розміру повідомлення залишається значною.

Подальше зменшення ступеня заповнення контейнера до рівня нижче 90% призводить до різкого погіршення результатів. У цьому діапазоні метод Хі-квадрат не завжди здатний навіть виявити сам факт приховування, оскільки середнє відхилення оцінки в окремих випадках стає порівнянним з відносною довжиною прихованого повідомлення. Найгірші результати метод демонструє при ступені заповнення нижче 60%, де псевдовипадкове розташування

модифікованих пікселів практично нівелює статистичні ознаки, на яких базується критерій Хі-квадрат. У діапазоні заповнення від 60% до 90% метод, як правило, дозволяє зафіксувати наявність вбудовування, однак супроводжується надзвичайно великими помилками в оцінці розміру інформаційного повідомлення.

Таким чином, аналіз рисунку 2.3 підтверджує, що метод Хі-квадрат є ефективним інструментом для аналізу послідовного НЗР-вбудовування, проте його застосування для кількісної оцінки псевдовипадково прихованих повідомлень є суттєво обмеженим і потребує використання додаткових або альтернативних методів стеганоаналізу.

На рисунку 2.4 наведено помилку ΔL оцінки розміру інформаційного повідомлення в залежності від ступеню заповнення контейнеру ΔV для RS-методу у випадку послідовного заповнення найменших значущих розрядів зображення.

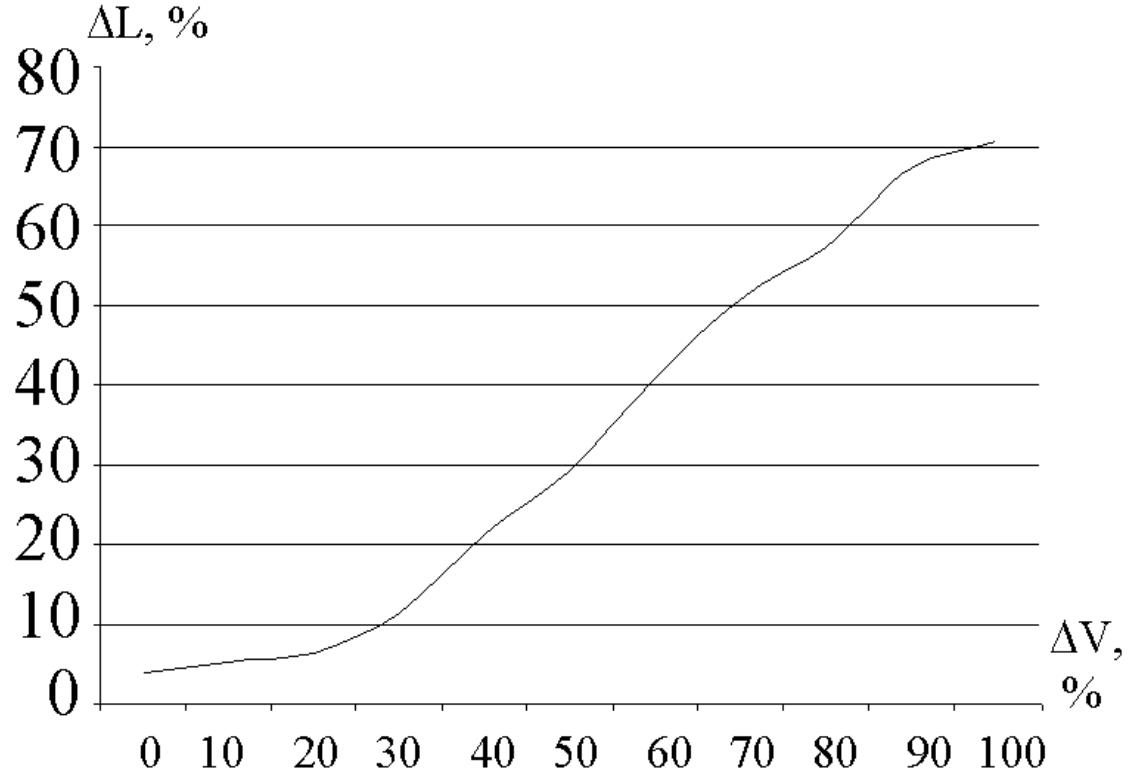


Рисунок 2.4 – Помилка оцінки розміру інформаційного повідомлення в залежності від ступеню заповнення контейнеру для RS-методу у випадку послідовного заповнення найменших значущих розрядів

З рисунку видно, що помилка визначення обсягу інформаційного повідомлення, схованого у псевдовипадково розташованих байтах зображення, RS-методом невпинно зростає практично прямопропорційно до збільшення довжини прихованого тексту. Найбільшу похибку, що досягає 100%, метод демонструє на файлах-контейнерах, заповнених на весь обсяг, тобто за повного заповнення найменших значущих розрядів. Така поведінка RS-методу пояснюється особливостями його роботи, оскільки за повної інверсії НЗР порушення локальної регулярності груп пікселів досягає граничного рівня, що призводить до втрати коректного співвідношення між кількістю регулярних і сингулярних груп, на якому ґрунтується кількісна оцінка розміру повідомлення.

Для середніх значень ступеня заповнення контейнера похибка також залишається значною, що вказує на недостатню точність RS-методу при кількісному оцінюванні розміру повідомлення для даного способу вбудовування. Водночас навіть за великих значень похибки RS-метод, як правило, дозволяє зафіксувати сам факт стеганографічного вбудовування, оскільки характерні зміни співвідношення регулярних і сингулярних груп залишаються вираженими. Таким чином, результати, наведені на рисунку 2.4, свідчать про те, що у випадку послідовного вбудовування інформації RS-метод не забезпечує достатньої точності кількісного визначення обсягу прихованого повідомлення, особливо за середніх і великих значень ступеня заповнення контейнера. Водночас характерні зміни співвідношення регулярних і сингулярних груп дозволяють використовувати RS-метод як надійний індикатор факту наявності прихованої інформації в графічному контейнері..

На рисунку 2.5 наведено помилку ΔL оцінки розміру інформаційного повідомлення в залежності від ступеню заповнення контейнеру ΔV для RS-методу у випадку псевдовипадкового заповнення найменших значущих розрядів зображення. З рисунку видно, що помилка визначення обсягу інформаційного повідомлення, схованого у послідовно розташованих байтах зображення, RS-методом залишається відносно низькою для всього діапазону досліджуваних обсягів повідомлення. Незалежно від ступеня заповнення

контейнера, отримані значення похибки не перевищують помірних величин, що свідчить про достатню стійкість RS-методу до послідовного характеру модифікації найменших значущих розрядів. Найбільшу похибку, яка становить близько 12%, метод демонструє на файлах-контейнерах, заповнених на 100%, що може бути пов’язано з насиченням статистичних ознак та зменшенням контрасту між регулярними та сингулярними групами пікселів при повному заповненні контейнера.

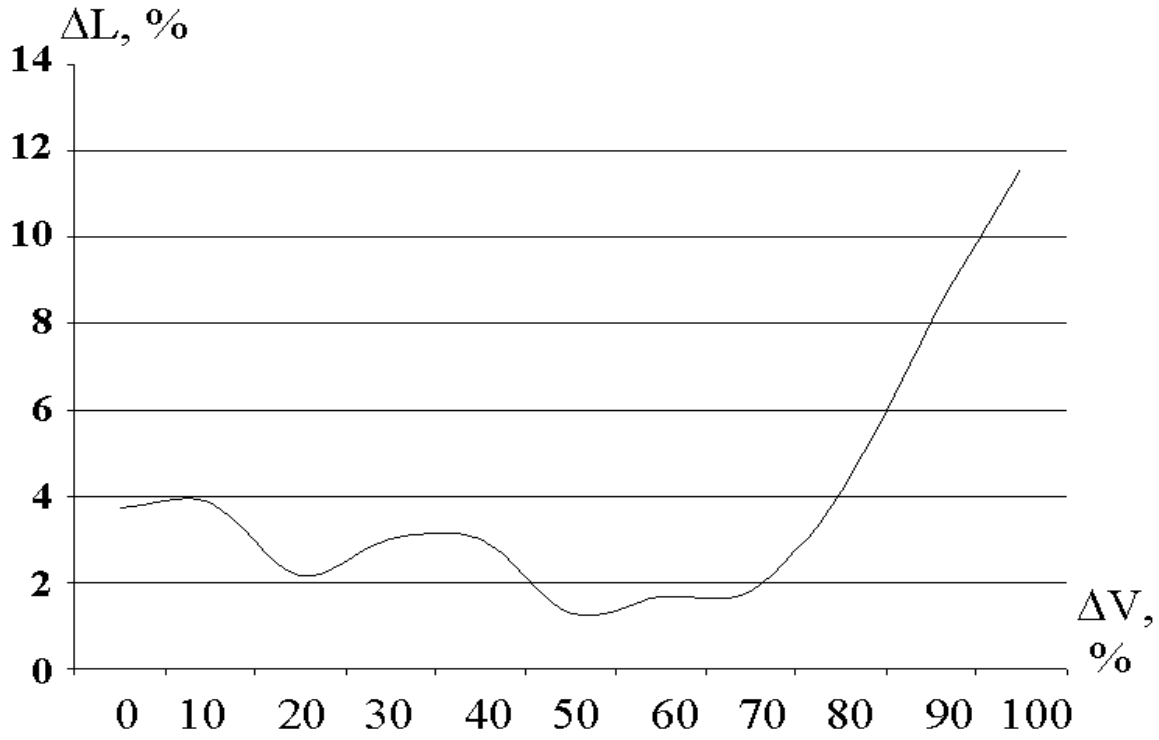


Рисунок 2.5 – Помилка оцінки розміру інформаційного повідомлення в залежності від ступеню заповнення контейнеру для RS-методу у випадку псевдовипадкового заповнення найменших значущих розрядів

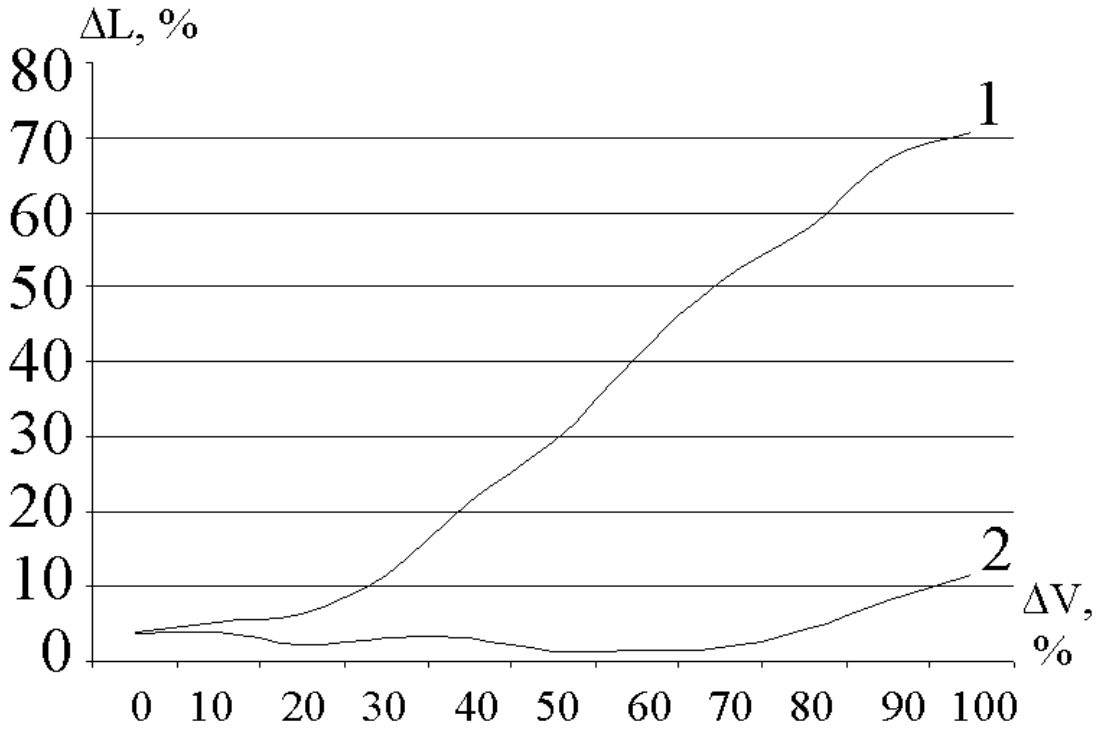
Для середніх і малих значень ступеня заповнення контейнера похибка оцінки є ще меншою, що зумовлено більш вираженим порушенням локальної регулярності груп пікселів у порівнянні з природною зашумленістю зображення. За умов псевдовипадкового вбудовування інформації RS-метод демонструє значно кращі результати, що проявляється у відносно низьких

значеннях похибки на більшій частині досліджуваного діапазону ступенів заповнення контейнера. Це дозволяє RS-методу коректно оцінювати кількісні характеристики прихованого повідомлення, зокрема навіть за відносно невеликих обсягів вбудованої інформації. Отримані результати підтверджують доцільність використання RS-методу для кількісної оцінки обсягу повідомлення саме у випадку псевдовипадкового вбудовування в найменші значущі розряди графічного контейнера.

Для зручності порівняння ефективності різновидів RS-методу на рисунку 2.6 наведено відповідні залежності помилки оцінки розміру прихованого повідомлення від ступеня заповнення контейнера для послідовного та псевдовипадкового способів вбудовування. Таке спільне подання результатів дозволяє наочно простежити вплив характеру модифікації найменших значущих розрядів на точність RS-методу та визначити межі його ефективного застосування.

Аналіз наведених на рисунку 2.6 залежностей свідчить, що RS-метод демонструє достатньо високу точність при виявленні псевдовипадкового вбудовування інформації. У діапазоні ступенів заповнення контейнера від 20% до 80% середнє відхилення оцінки не перевищує 3%, що дозволяє не лише впевнено ідентифікувати факт приховування, але й досить точно оцінити обсяг вбудованого повідомлення. Така поведінка пояснюється значним порушенням локальної регулярності груп пікселів при псевдовипадковій інверсії НЗР, що є ключовою ознакою, на якій ґрунтується RS-метод.

Водночас при малих значеннях ступеня заповнення контейнера, зокрема менших за 10%, ефективність RS-методу істотно знижується. У цьому випадку величина середнього відхилення оцінки (близько 3,7%) стає співмірною з відхиленням, характерним для порожніх контейнерів (приблизно 4,4%). За таких умов метод не завжди дозволяє однозначно зробити висновок про наявність стеганографічного вбудовування, оскільки кількість модифікованих пікселів є недостатньою для формування вираженого статистичного ефекту, що перевищує рівень природної зашумленості зображення.



1 – випадок послідовного заповнення найменших значущих розрядів;
2 – випадок псевдовипадкового заповнення найменших значущих розрядів.

Рисунок 2.6 – Помилка оцінки розміру інформаційного повідомлення в залежності від ступеню заповнення контейнеру для RS-методу

При подальшому зростанні ступеня заповнення понад 80% точність RS-методу знову зменшується. Це пов’язано з тим, що при великій кількості інверсій НЗР локальні порушення регулярності груп пікселів починають частково компенсуватися, а співвідношення регулярних і сингулярних груп наближається до граничних значень, що ускладнює коректну кількісну оцінку розміру прихованого повідомлення. У цьому діапазоні помилка RS-методу зростає тим більше, чим більшим є обсяг псевдовипадково вбудованої інформації, що чітко простежується на відповідній кривій рисунку 2.6. Найбільш помітні відхилення спостерігаються при ступені заповнення в інтервалі 10-20%, де статистичні зміни є ще недостатньо стабільними.

При аналізі послідовного вбудовування інформації RS-метод демонструє

значно гірші результати, що також наочно відображено на рисунку. Середнє відхилення оцінки у цьому випадку зростає практично монотонно разом зі збільшенням ступеня заповнення контейнера. Для обсягів прихованого повідомлення менших за 10% величина відхилення є порівнянною з відхиленням для порожніх контейнерів, що у більшості випадків не дозволяє однозначно ідентифікувати сам факт приховування. Це пов'язано з тим, що послідовна модифікація НЗР створює впорядковані зміни, які не призводять до суттєвого порушення локальної регулярності груп пікселів.

При ступені заповнення контейнера більшому за 10% RS-метод, як правило, дозволяє однозначно встановити факт наявності прихованої інформації, однак кількісна оцінка її розміру залишається вкрай неточною. У цьому діапазоні метод оцінює розмір схованого повідомлення в середньому на рівні приблизно 15-30% обсягу контейнера незалежно від фактичної довжини повідомлення. Таким чином, для послідовного вбудовування RS-метод має переважно індикативний характер і не може бути використаний як надійний засіб кількісної оцінки обсягу прихованої інформації.

Отримані результати підтверджують, що RS-метод є ефективним інструментом стеганоаналізу лише у певних сценаріях застосування, зокрема при псевдовипадковому вбудовуванні інформації з середнім ступенем заповнення контейнера. Це ще раз підкреслює доцільність поєднання RS-методу з іншими методами стеганоаналізу для підвищення загальної надійності виявлення та оцінки прихованих повідомлень.

2.4 Алгоритм поєднання методів для стеганоаналізу

Проведений у підрозділі 2.3 експериментальний аналіз показав, що метод Хі-квадрат і регулярно-сингулярний (RS) метод мають різні області максимальної ефективності, які визначаються як способом вбудовування

інформації в найменші значущі розряди пікселів, так і ступенем заповнення графічного контейнера. Метод Хі-квадрат демонструє високу точність у випадку послідовного вбудовування, тоді як RS-метод є більш ефективним для псевдовипадкового розташування модифікованих пікселів, особливо в діапазоні середніх значень ступеня заповнення контейнера. Водночас для малих обсягів прихованого повідомлення обидва методи можуть давати результати, співмірні з відхиленнями, характерними для порожніх контейнерів, що унеможливлює формування однозначного висновку на основі застосування лише одного критерію.

З огляду на це для підвищення надійності стеганоаналізу та зменшення похибки кількісної оцінки обсягу прихованої інформації запропоновано алгоритм поєднання методів Хі-квадрат і RS, який передбачає їх паралельне застосування до одного й того ж зображення-контейнера з подальшим узгодженим аналізом отриманих оцінок. Для кожного контейнера формуються дві незалежні оцінки відносного обсягу прихованого повідомлення: оцінка методом Хі-квадрат $p\chi 2$ та оцінка RS-методом RS, після чого прийняття рішення здійснюється на основі їх порівняння між собою та з пороговими значеннями, встановленими експериментально.

Якщо оцінка методом Хі-квадрат не перевищує 0,1%, а оцінка RS-методом є меншою за 4%, контейнер вважається порожнім, а стеганографічне вбудовування – відсутнім, оскільки такі значення відповідають рівню статистичних відхилень, характерних для зображень без прихованої інформації. У випадку, коли оцінка методом Хі-квадрат є близькою до 100% (практично перевищує 90%), контейнер розглядається як повністю або майже повністю заповнений. Подальше визначення способу вбудовування в цьому разі здійснюється за результатами RS-методу: значення RS на рівні близько 30% або нижче свідчить про послідовне вбудовування, тоді як значення, близькі до 80% і вище, вказують на псевдовипадковий характер приховування інформації.

Якщо оцінка RS-методом перевищує 30%, а оцінка методом Хі-квадрат не перевищує приблизно 30%, робиться висновок про псевдовипадкове

вбудовування повідомлення. У цьому випадку, за умови p_{RS}<80%, оцінку RS-методу можна вважати достатньо точною для визначення обсягу прихованої інформації, тоді як при перевищенні цього рівня доцільно вважати, що обсяг вбудованого повідомлення перевищує 80% місткості контейнера. Якщо ж значення RS є меншим за 30%, воно додатково порівнюється з оцінкою методу Хі-квадрат: у разі, коли p_χ^2 є значно меншою або близькою до нуля, виявляється псевдовипадкове вбудовування з обсягом повідомлення, що відповідає оцінці RS-методу, а у випадку, коли оцінка методом Хі-квадрат є порівнянною з p_{RS} або перевищує її, робиться висновок про послідовне приховування, і розмір схованих даних визначається за оцінкою методу Хі-квадрат.

Формалізовані правила визначення способу вбудовування повідомлення та оцінки ступеня наповнення контейнера, що застосовуються в межах запропонованого алгоритму поєднання методів, наведені у таблиці 2.2, де для різних комбінацій значень p_χ^2 та p_{RS} визначено відповідний тип стеганографічного вбудовування і спосіб оцінки обсягу прихованого повідомлення. Використання наведених у таблиці правил у зазначеній послідовності дозволяє охопити переважну більшість можливих поєднань результатів методів Хі-квадрат і RS та сформувати обґрунтований і однозначний висновок щодо наявності, характеру та кількісних параметрів прихованої інформації.

Отримання результатів, які не підпадають під описані у таблиці 2.2 правила, може бути зумовлене статистичними флуктуаціями, підвищеною зашумленістю зображення, малими розмірами контейнера або специфічними властивостями його структури, що вимагає додаткового аналізу або застосування інших методів стеганоаналізу. Таким чином, запропонований алгоритм поєднання методів Хі-квадрат і RS забезпечує більш надійне виявлення стеганографічного вбудовування та підвищує точність оцінки обсягу прихованої інформації порівняно з використанням кожного з методів окремо, що підтверджує доцільність його застосування в автоматизованих системах стеганоаналізу.

Таблиця 2.2 – Визначення способу вбудовування повідомлення та обсягу наповнення контейнера

Оцінка	p	Спосіб вбудовування повідомлення	Наповнення контейнера V
RS-методом	методом Хі-квадрат		
<4%	<0,1%	вбудовування відсутнє	порожній
<30%	>90%	послідовно	повний або майже повний
>80%	>90%	псевдовипадково	
>30%	<80%	псевдовипадково	$V=p_{RS}$ - за $p_{RS}<80\%$ $V>80\%$ - за $p_{RS}>80\%$
<30%	<10%	псевдовипадково	$V=p_{RS}$
<30%	$p_{xi2}>\approx p_{RS}$	послідовно	$V=p_{xi2}$

3 РОЗРОБКА АЛГОРИТМУ СТЕГАНОАНАЛІЗУ

На підставі проведеного дослідження, а також з урахуванням принципів роботи алгоритмів виявлення стеганографічних вбудовувань у найменші значущі розряди пікселів, можна зробити висновок, що ключовим фактором, який визначає стеганостійкість алгоритму приховування інформації, є спосіб вибору елементів контейнера для здійснення їх модифікації. Саме закономірності цього вибору безпосередньо впливають на статистичні характеристики заповненого контейнера та, відповідно, на ефективність методів стеганоаналізу. У зв'язку з цим з метою підвищення стеганостійкості методу доцільно запропонувати модифікацію алгоритму вбудовування, що ґрунтується на оптимальному виборі елементів контейнера для приховування інформаційного повідомлення.

Загальна задача полягає у тому, щоб здійснювати вбудовування не у довільно вибрані елементи контейнера і не за фіксованим або псевдовипадковим правилом, а з урахуванням статистичних властивостей самих елементів контейнера. Для цього пропонується об'єднувати елементи контейнера у групи не за результатами однакових модифікацій, як це має місце, наприклад, у регулярно-сингулярному аналізі з використанням функцій обернення, а таким чином, щоб елементи кожної групи мали подібні властивості та описувалися однаковою або близькою функцією щільності розподілу $f_i(c_i)$ елементів i-тої групи.

Нехай контейнер поділено на m груп елементів, кожна з яких містить k_i елементів і характеризується власним законом розподілу. Позначимо через C_i область допустимих значень елементів контейнера i-тої групи. У такому випадку зміна одного елемента цієї групи дозволяє вбудувати інформацію в кількості:

$$\mathrm{qi} = \log_2 \mathrm{C_i}. \qquad (3.1)$$

Де q_i визначає максимально можливу кількість інформаційних розрядів, які можуть бути закодовані шляхом модифікації одного елемента відповідної групи. Таким чином, величина q_i задає потенційну інформаційну ємність групи і безпосередньо пов'язана з шириною області допустимих значень C_i.

Слід зауважити, що інформаційне повідомлення перед вбудовуванням, як правило, проходить етапи попередньої обробки, зокрема стискання з метою зменшення обсягу та шифрування для підвищення криптографічної стійкості. У результаті такого перетворення повідомлення набуває властивостей, близьких до випадкової послідовності, і характеризується високою ентропією. За цих умов вбудовування інформації призводить до змін статистичних характеристик контейнера, які можуть бути виявлені методами стеганоаналізу.

Функція щільності розподілу елементів iii-тої групи контейнера після вбудовування інформаційного повідомлення може бути подана у вигляді:

$$\bar{f}_i(c_i, x_i) = \frac{k_i - x_i}{k_i} \cdot f_i(c_i) + \frac{x_i}{k_i} \cdot \frac{1}{c_i}, \tag{3.1}$$

де x_i – кількість незмінених елементів у і-тій групі;

$f_i(c)$ – початкова функція щільності розподілу елементів контейнера.

Дана формула відображає той факт, що заповнений контейнер є сумішшю початкового розподілу та розподілу, нав'язаного прихованими даними.

Для оцінки можливості використання окремих областей контейнера для приховування інформаційного повідомлення вводиться імовірність того, що область S контейнера є придатною для вбудовування, яка визначається згідно з формулою (3.3). Ця імовірність характеризує ступінь збереження статистичних властивостей контейнера в межах розглядуваної області.

$$P(S) = \prod_{i=1}^{m} \prod_{j=1}^{k_i} \bar{f}_i(c_j^i). \tag{3.2}$$

У свою чергу, імовірність того, що в область SSS контейнера буде безпосередньо вбудовано інформаційне повідомлення, визначається формулою (3.4), яка доповнює попередню оцінку та дозволяє кількісно описати процес розподілу прихованих даних у контейнері.

$$P(S) = \prod_{i=1}^{m} \prod_{j=1}^{k_i} \bar{f}_i(c_j^i, x_i). \quad (3.3)$$

Очевидно, що стійкість стеганографічної системи безпосередньо пов’язана з величиною відхилення статистичних характеристик заповненого контейнера від початкових. У цьому зв’язку критерієм стеганостійкості виступає величина, визначена формулою (3.5). Чим меншою є ця величина, тим менш помітними є зміни контейнера і тим складніше виявити факт приховування інформації методами стеганоаналізу.

$$D(P|\bar{P}) = \sum_S P(S) \cdot log_2 \frac{P(S)}{\bar{P}(S)}. \quad (3.4)$$

Таким чином, задача оптимального розподілу інформаційного повідомлення у контейнері зводиться до знаходження такого вектора $\{x_i\}$, за якого мінімізується величина, визначена формулою (3.5), за умови відомості функції щільності розподілу $f_i(c)$. Досягнення цієї умови можливе, зокрема, шляхом розробки незмінного алгоритму вибору елементів контейнера для вбудовування, який дозволяє заздалегідь визначити області, що залишаються незмінними. Знаючи ці області, можна коректно обчислити їх функцію щільності розподілу та використовувати її для формування контейнера з мінімальними статистичними спотвореннями.

У крайньому випадку, коли збереження статистичних характеристик контейнера в межах допустимих відхилень є неможливим, виникає необхідність попередньої модифікації самого контейнера або вибору іншого контейнера, статистичні властивості якого є більш придатними для приховування

інформаційного повідомлення. Такий підхід дозволяє розглядати задачу стеганографічного вбудовування як задачу оптимізації, спрямовану на підвищення стеганостійкості та зменшення ефективності відомих методів стеганоаналізу.

ВИСНОВКИ

У ході виконання дипломної роботи було досліджено сучасні підходи до стеганографічного аналізу графічних файлів та розроблено алгоритмічні рішення, спрямовані на підвищення ефективності виявлення та приховування інформації. Отримані результати дозволяють зробити такі висновки.

У роботі детально досліджено принципи функціонування та практичну ефективність статистичних методів стеганоаналізу графічних файлів, зокрема методу Хі-квадрат і регулярно-сингулярного (RS) методу. Проведений експериментальний аналіз показав, що метод Хі-квадрат є найбільш ефективним для виявлення послідовного вбудовування інформації в найменші значущі розряди пікселів, тоді як RS-метод демонструє вищу чутливість до псевдовипадкового характеру модифікацій контейнера. Встановлено, що точність обох методів суттєво залежить від ступеня заповнення контейнера та способу вбудовування, а для малих обсягів прихованого повідомлення результати аналізу можуть бути співмірними зі статистичними відхиленнями, характерними для порожніх контейнерів.

На основі отриманих експериментальних даних запропоновано алгоритм поєднання методів Хі-квадрат і RS з метою підвищення достовірності стеганоаналізу. Запропонований алгоритм передбачає паралельне застосування обох методів до одного й того ж зображення-контейнера з подальшою інтерпретацією результатів на основі встановлених порогових значень і зон надійності кожного з методів. Такий підхід дозволяє визначати не лише факт наявності прихованої інформації, але й характер її вбудовування (послідовний або псевдовипадковий), а також отримувати більш точну оцінку обсягу прихованого повідомлення. Порівняння результатів показало, що поєднання методів забезпечує зменшення середньої похибки та підвищення стійкості до хибнопозитивних і хибнонегативних рішень порівняно з використанням кожного методу окремо.

У роботі проаналізовано можливості підвищення стеганостійкості алгоритмів приховування інформації в графічних файлах на основі врахування статистичних властивостей контейнера та принципів роботи відомих методів стеганоаналізу. Алгоритм ґрунтується на оптимальному виборі елементів контейнера для вбудовування інформації з урахуванням їх статистичних властивостей та функцій щільності розподілу. Показано, що задача приховування може бути сформульована як задача оптимального розподілу інформаційного повідомлення в контейнері з мінімізацією статистичних відхилень між початковим і заповненим контейнерами. Такий підхід дозволяє зменшити ймовірність виявлення факту приховування інформації методами Хі-квадрат і RS, а також підвищити загальну стеганостійкість системи.

У цілому результати роботи підтверджують доцільність комплексного підходу до стеганоаналізу та стеганографічного приховування інформації, що поєднує статистичний аналіз і оптимізаційні алгоритми. Отримані теоретичні та практичні результати можуть бути використані при розробці автоматизованих систем стеганоаналізу, а також для подальших досліджень у галузі захисту інформації та прихованого передавання даних у графічних контейнерах.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Provos N. Hide and seek: An introduction to steganography / N. Provos, P. Honeyman // IEEE Security Privacy. – 2003. – P. 32-44.

2. Reversible image hiding algorithm based on pixels difference / H. Ren, C. Chang, J. Zhang // In the IEEE International Conference on Automation & Logistics, ICAL'09, Shenyang. – 2009. – P. 847-850.

3. Кузнецов О.О. Стеганографія: навчальний посібник / О.О. Кузнецов, С.П. Євсеєв, О.Г. Король. – Х.: ХНЕУ, 2011. – 232 с.

4. F5 – A steganographic algorithm: High capacity despite better steganalysis / A. Westfeld // Proceedings of 4th International Workshop Information Hiding, Springer-Verlag. – 2001. – P. 289-302.

5. A steganographic method based upon JPEG and particle swarm optimisation algorithm / X. Li, J. Wang // Information Sciences. – 2007. – Vol. 177(3). – P. 99-109.

6. Robust steganography using bit plane complexity segmentation / S.T. Maya, M.N. Miyatake, R.V. Medina // 1st International Conference on Electrical and Electronics Engineering, Mexico. – 2004. – P. 40-43.

7. Free-form deformation of solid geometric models / T.W. Sederberg, S.R. Parry // Proceedings of ACM SIGGRAPH 1986, ACM Press. – 1986. – P 151-160.

8. Steganographic technique based on minimum deviation of fidelity (STMDF) / J.K. Mandal, M. Sengupta // 2nd International Conference on Emerging Applications of Information Technology (EAIT), Kolkata. – 2011. – P. 298-301.

9. A two-dimensional interpolation function for irregularlyspaced data / D. Shepard // Proceedings of the 1968 23rd ACM National Conference. – 1968. – P. 517-524.

10. Digital Watermarking Based on Neural Network Technology for Grayscale Images / J. Chen, Tung-Shou Chen, Keh-Jian Ma, Pin-Hsin Wang // Encyclopedia of Multimedia Technology and Networking. – 2005. – Vol. 29. – P. 204-212.

11. Digital watermarking based on neural networks for color images / Pao-Ta Yu, Hung-Hsu Tsai, Jyh-Shyan Lin // Signal Processing. – 2001. – Vol. 81(3). – P. 663-671.

12. A new steganography approach for image encryption exchange by using the least significant bit insertion / M.A.B. Younes, A. Jantan // International Journal of Computer Science and Network Security. – 2008. – P. 247-254.

13. Азаров О.Д. Основи комп'ютерної стеганографії: навчальний посібник. / О.Д. Азаров, В.О. Хорошко, М.Є. Шелест, Ю.Є. Яремчук. – Вінниця: ВДТУ, 2003. – 143 с.

14. Katz J. Introduction to Modern Cryptography, 2nd ed. / J. Katz, Y. Lindell. – CRC Press, Boca Raton. – 2014.

15. Tutuncu K. New Approach in E-mail Based Text Steganography / K. Tutuncu, A.A. Hassan. – IJISAE, 3(2). – 2015. – P. 54-56.

16. Xiang L. A linguistic steganography based on word indexing compression and candidate selection / L. Xiang, W. Wu, X. Li, et al. // Multimedia Tools and Applications, 77. – 2018. – P. 28969-28989.

17. Конахович Г.Ф. Комп'ютерна стеганографічна обробка й аналіз мультимедійних даних: підручник / Г.Ф. Конахович, Д.О. Прогонов, А.Ю. Пузиренко. – К.: ЦУЛ, 2018. – 555 с.

18. Основи комп'ютерної стеганографії: навч. посібн. для студентів і аспірантів / В.О. Хорошко, О.Д. Азаров, М.В. Шелест та ін. – Вінниця: ВДТУ, 2003. – 143 с.

19. Юдін О. Аналіз стеганографічних методів приховування інформаційних потоків у контейнерах різних форматів / О. Юдін, Р. Зюбіна, О. Фролов // Радиоэлектроника и информатика. – 2015, № 3. – С. 13-21.

20. Fridrich J. Reliable Detection of LSB Steganography in Color and Grayscale Images Binghamton University / J. Fridrich, M. Goljan, R. Du. – New York, USA. – 2001.

ДОДАТОК А

Презентація

Національний університет "Запорізька політехніка"
Кафедра інформаційної безпеки та наноелектроніки

Дипломна робота

Розроблення та аналіз алгоритму приховування інформації в графічних файлах із підвищеною стійкістю до стеганоаналізу

Виконав: ст. гр. БКз-814м С. А. Нападайло

Для захисту інформації існує кілька підходів:
1. Блокування несанкціонованого доступу до інформації.
2. Шифрування інформації.
3. Приховування інформації так, аби її неможливо було знайти.

Метою стеганографічних методів захисту є вбудовування інформації до даних так, щоб не можна було, навіть, запідозрити існування підтексту.

2

Перетворення набору літер для стеганографічного вбудовування:

Символ ASCII	Основа	
	10	2
S	83	1010011
T	84	1010100
E	69	1000101
G	71	1000111
O	79	1001111

"STEGO" =
= "1010011101010010001011000111100111 1"

[3]

Основні методи графічної стеганографії:
1. Заміна найменших значущих розрядів кольорів пікселів зображення.
2. Заміна палітри.

Основні методи стеганоатаки на метод заміни найменших значущих розрядів:
1. Метод Хі-квадрат.
2. Регулярно-сингулярний (RS-метод).

[4]

Метою дипломної роботи є розробка нового стеганостійкого алгоритму приховування інформації у графічних файлах.

Для досягнення поставленої мети, необхідно розв'язати наступні задачі:

- дослідження основних методів стеганоаналізу;
- розробка нового стеганостійкого алгоритму.

5

1. Стеганоаналіз методом Хі-квадрат

У методі використовуються:

- аналіз гістограми, отриманої за елементами зображення;
- оцінки розподілу пар значень цієї гістограми.

Джерело формування пар:

- для *.bmp - значення пікселів зображення;
- для *.jpeg - коефіцієнти ДКП.

У порожньому контейнері частоти елементів зі значеннями 2N й 2N+1 - переважно різні, а у наповненому – зближуються чи стають рівними.

6

1. Стеганоаналіз методом Хі-квадрат

Переваги. Підходить для аналізу зображень з наповненням за будь-яким методом приховування.

Недоліки. Результати роботи методу значною мірою залежать від способу приховування даних:

- за послідовного запису в елементи контейнера метод забезпечує гарні результати;
- за псевдовипадкового вибору елементів й розсіювання повідомлення усією довжиною контейнера ефективність методу істотно знижується. [7]

2. Стеганоаналіз RS-методом

1. Зображення розбивається на групи $G(g_i, g_{i+1}, \ldots, g_{i+n})$, де n - парне.
2. Для групи пікселів визначається довільна функція регулярності f(G) їхніх значень g_i:

$$f(G) = f(g_1, g_2, \ldots, g_n) = \sum_{i=1}^{n-1} |g_{i+1} - g_i| \quad (1)$$

3. Для всіх груп зображення обчислюються дві функції обернення над функціями f(G):

$F_1(x)$ - інвертування молодшого розряду аргументу;

$$F_{-1}(x) = F_1(x+1) - 1, \quad (2)$$

де $x = f(G)$, $F(F(x)) = x$. [8]

2. Стеганоаналіз RS-методом

4. Поділимо всі групи пікселів на класи:
- регулярні групи: $G \in R$: $f(F(G)) > f(G)$;
- сингулярні групи: $G \in S$: $f(F(G)) < f(G)$;

отримавши 4 класи:

$$R_M, S_M, R_{-M}, S_{-M}, \quad (3)$$

де M й -M означають застосування F_1 й F_{-1}, відповідно.

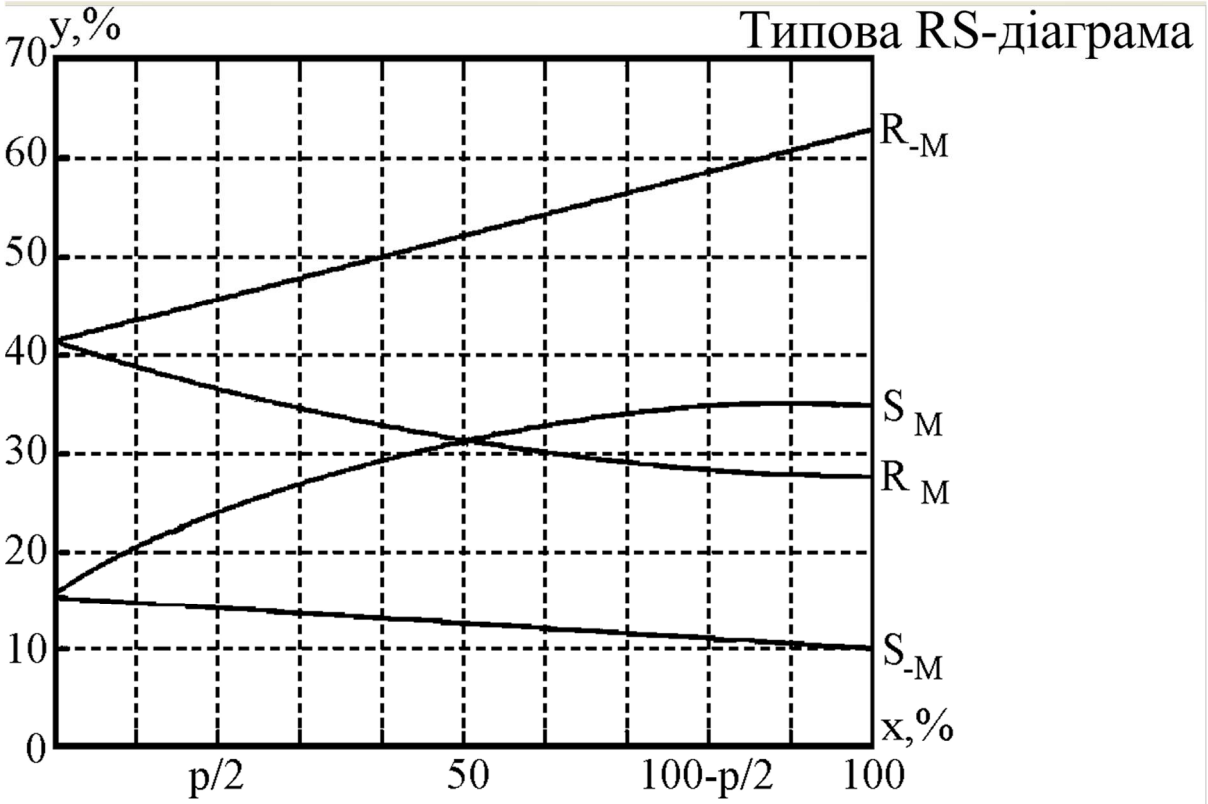
Для порожнього контейнера:

$$R_M = R_{-M}, S_M = S_{-M}. \quad (4)$$

Для 100%-заповненого контейнера:

$$R_M - S_M = 0, R_{-M} - S_{-M} \sim L, \quad (5)$$

де L - довжина повідомлення. 9



Крапки на прямих: $R_{-M}(p/2)$, $R_{-M}(1-p/2)$, $S_{-M}(p/2)$, $S_{-M}(1-p/2)$,
Крапки на параболах: $R_M(p/2)$, $R_M(1-p/2)$, $S_M(p/2)$, $S_M(1-p/2)$. 10

2. Стеганоаналіз RS-методом

Розв'язання системи

$$2(d_1 + d_0)\,x^2 + (d_{-0} - d_{-1} - d_1 - 3d_0)\,x + d_0 - d_{-0} = 0 \quad (6)$$

де

$$d_0 = R_M(p/2) - S_M(p/2),$$

$$d_{-0} = R_{-M}(p/2) - S_{-M}(p/2),$$

$$d_1 = R_M(1-p/2) - S_M(1-p/2),$$

$$d_{-1} = R_{-M}(1-p/2) - S_{-M}(1-p/2).$$

дозволяє знайти оцінку довжини повідомлення p:

$$p = \frac{x}{x - \frac{1}{2}} \quad (7)$$

[11]

Дослідження ефективності методів

Умови проведення: шляхом аналізу файлів зображень, у яких одним із двох способів:

- псевдовипадкової варіації методу НЗР;
- послідовної варіації методу НЗР

було:

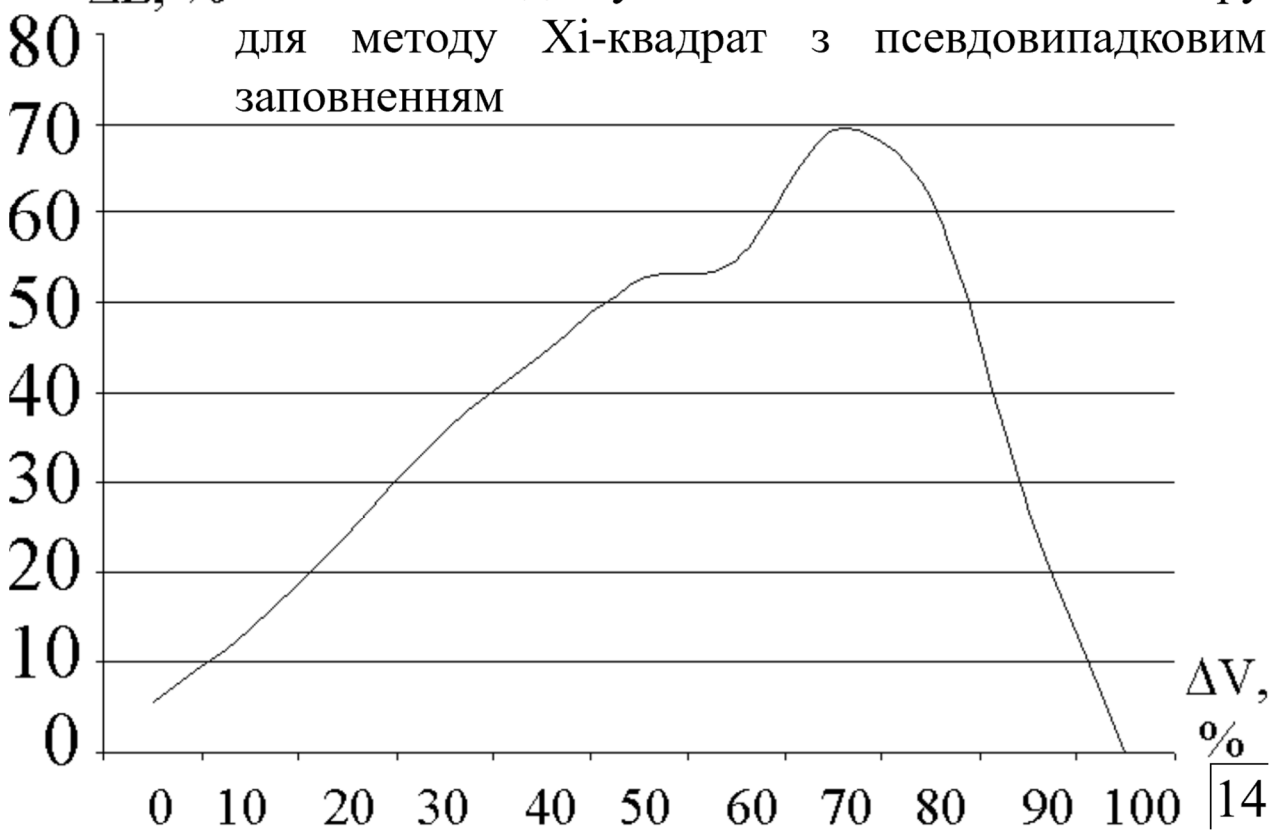
- або сховане інформаційне повідомлення;
- або жодної інформації не вбудовували.

Для кожного файлу аналіз здійснювся за допомоги кожного із двох реалізованих методів:

- методом Хі-квадрат;
- RS-методом.

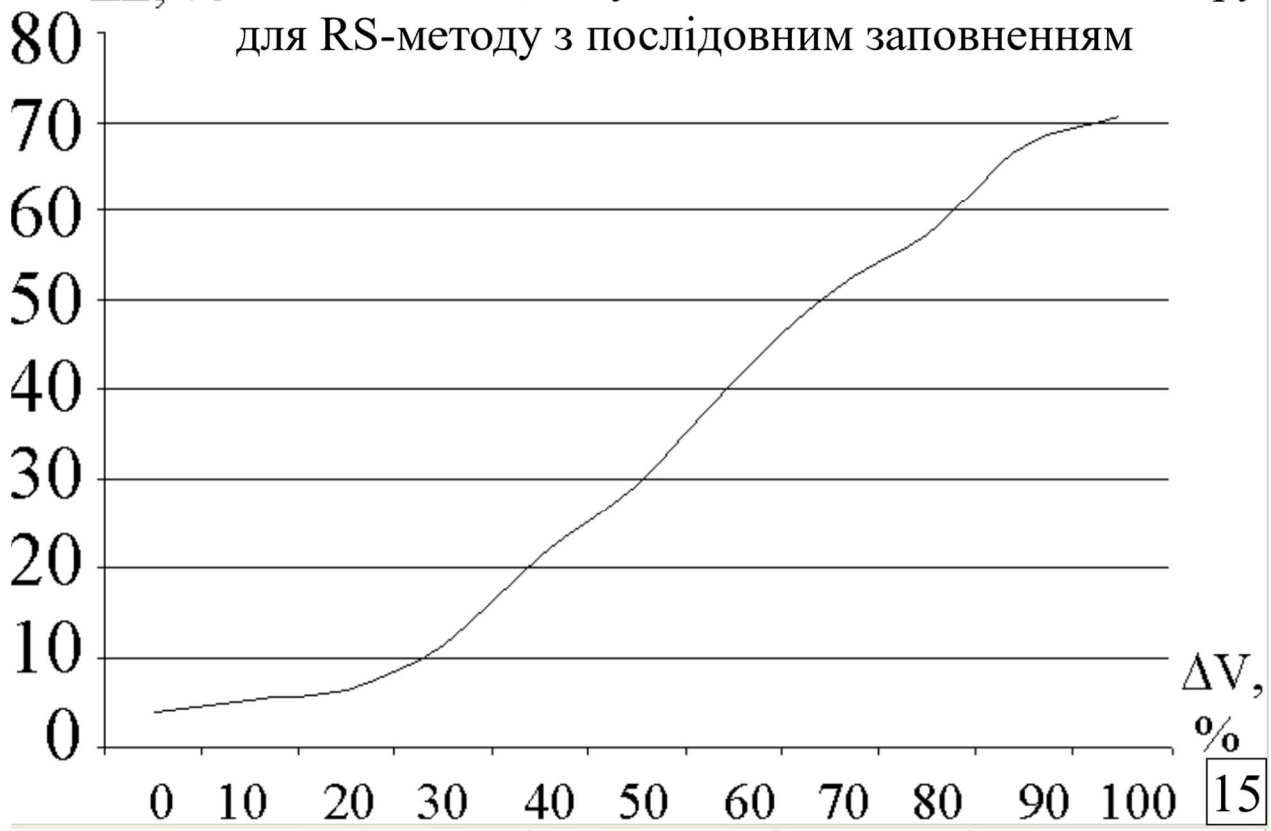
[12]

Похибка оцінки розміру повідомлення в залежності від ступеню заповнення контейнеру для методу Хі-квадрат з послідовним заповненням 13



Похибка оцінки розміру повідомлення в залежності від ступеню заповнення контейнеру для методу Хі-квадрат з псевдовипадковим заповненням 14

Похибка оцінки розміру повідомлення в залежності від ступеню заповнення контейнеру для RS-методу з послідовним заповненням



15

Похибка оцінки розміру повідомлення в залежності від ступеню заповнення контейнеру для RS-методу з псевдовипадковим заповненням

16

Оцінка p		Спосіб вбудовування повідомлення	Наповнення контейнера V
RS-методом	методом хі-квадрат		
<4%	<0,1%	вбудовування відсутнє	порожній
<30%	>90%	послідовно	повний або майже повний
>80%	>90%	псевдовипадково	
>30%	<80%	псевдовипадково	$V=p_{RS}$ - за $p_{RS}<80\%$ $V>80\%$ - за $p_{RS}>80\%$
<30%	<10%	псевдовипадково	$V=p_{RS}$
<30%	$p_{xi2}>=p_{RS}$	послідовно	$V=p_{xi2}$

17

Алгоритм

Функція щільності розподілу елементів і-групи контейнеру з вбудованим повідомленням:

$$\bar{f}_i(c_i,x_i)=\frac{k_i-x_i}{k_i}\cdot f_i(c_i)+\frac{x_i}{k_i}\cdot\frac{1}{c_i} \qquad (8)$$

де $f_i(c_i)$ - щільність розподілу елементів і-групи,

x_i - кількість незмінених елементів,

k_i - кількість елементів і-групи,

c_i - область допустимих значень елементів контейнеру і-групи

18

Алгоритм

Імовірність того, що область S контейнеру є придатною для вбудовування інформаційного повідомлення:

$$P(S) = \prod_{i=1}^{m} \prod_{j=1}^{k_i} \bar{f}_i(c_j^i) \qquad (9)$$

Імовірність того, що в область S контейнеру буде вбудовано інформаційне повідомлення:

$$P(S) = \prod_{i=1}^{m} \prod_{j=1}^{k_i} \bar{f}_i(c_j^i, x_i) \qquad (10)$$

19

Алгоритм

Стійкість стеганографічної системи тим вище, чим менша величина:

$$D(P \mid \bar{P}) = \sum_{S} P(S) \cdot \log_2 \frac{P(S)}{\bar{P}(S)} \qquad (11)$$

Задача придатного розподілу повідомлення у контейнері зводиться до знаходження такого вектору $\{x_i\}$, за якого мінімізується величина $D(P \mid \bar{P})$. Ця задача розв'язується за умови відомості функції $f_i(c_i)$.

20

Висновки

1. Досліджено роботу стеганографічних методів аналізу графічних файлів методами хі-квадрат й регулярно-сингулярним.
2. Запропоновано алгоритм їхнього об'єднання з метою підвищення якості стеганоаналізу.
3. Запропоновано новий алгоритм приховування інформації в графічних файлах.