Міністерство освіти і науки України
Львівський національний університет імені Івана Франка
Факультет прикладної математики та інформатики
Кафедра обчислювальної математики

Курсова робота
Порівняння ефективності класів стеганографічних алгоритмів

Студента групи ПМп-31:
Грициндишина В. Л
Науковий керівник:
ст. в. Гарасим Я. С

Львів
2022

Зміст

Вступ

Захист інформації є однією з вічних проблем людства. Протягом історії способи її розв'язання визначались рівнем розвитку технологій. Криптологія вивчає методи шифрування та дешифрування інформації. Її можна поділити на два розділи: криптографію та криптоаналіз.

Зважаючи на розвиток обчислювальних машин, виникає потреба в передачі медіа файлів через Internet. Власне, з розвитком технологій передавання, отримання та обробки даних усе більш актуальною стає тема захисту інформації. Дослідження криптографії та стеганографії працюють на вирішення цього питання. Метою криптографії є захист змісту методами шифрування. Проте це привертає увагу зловмисників, котрі спрямовують всі свої сили на дешифрування інформації. Тому в деяких випадках доречніше приховати факт існування секретного файлу з цінним змістом, що, власне, і досліджує стеганографія. Слід зауважити, що вона доповнює криптографію: часто методи обох напрямків об'єднують з метою забезпечення більш ефективного захисту даних (наприклад, коли зміст повідомлення захищають певними методами шифрування криптографії, а факт передачі ключів методами стеганографії).

Мультимедійні файли - одне з найпопулярніших місць для зберігання інформації. Чим більший їхній розмір, тим більше інформації, якій потрібен захист, можна помістити у файл. Метою роботи є дослідження ефективності приховування інформації в різних мультимедійних структурах даних.

Слово «стеганографія» в перекладі з грецької дослівно означає «тайнопис». Відомо, що цей напрям створений значно швидше ніж, криптографія, проте з часом був витиснений. Перша згадка про неї була орієнтовно в 450 роках до н.е в одній з трактат Геродота, що описав випадок передачі повідомлення Демартом, який зіскоблював віск з дощечок, писав лист прямо на дереві, а потім знову покривав дощечки воском.

Використання смужок шовку, так званий книжковий шифр Енея, теж можна вважати древнім методом засекречення факту інсування іфнормації. Цікаво, що під час громадянської війни у 1779 році спеціальні агенти передавали звістки Джорджу Вашингтону, використовуючи безбарвні речовини в якості чорнила.

Перша книга, яка описувала різні методи приховування повідомлень, була написана у 1499 році ченцем, криптографом та стеганографом Трітеміусом.

Варто відзначити, що значний внесок в розвиток стеганографії зробили В.Хорошко, М.Шелест, Н.Кошкіна. Дослідження на стійкість проводили J.Fridrich, R.Popa, N.Johnson. Проблему приховування секретних відомостей та побудову алгоритмів захисту розглядали G.Simmons, J.Fridrich, R.Anderson, W.Bender.

В сучасному суспільстві методи стеганографії використовують для:

- захисту інформації від несанкційованого доступу;
- захисту авторських прав;
- маскування програмного забезпечення;
- протидії системам моніторингу мереж.

1 Постановка задачі

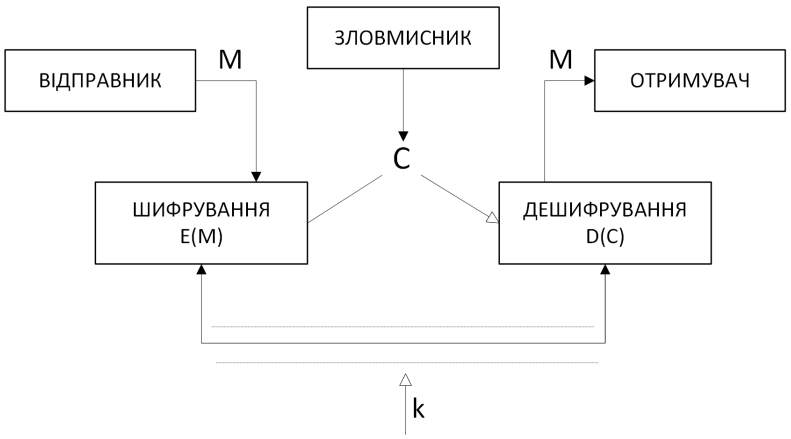
1.1 Принцип криптозахисту

Криптографія являє собою методи перетворення інформації таким чином, щоб для зловмисника вона не була корисною.
Такі методи дають змогу забезпечити:

- цілісність інформації - шляхом позбавлення зловмисника можливості змінити дані способом змінення вмісту або подати в канал зв'язку хибну інформацію;
- конфіденційність інформації - позбавлення зловмисника шансу отримати дані з каналу зв'язку.

Питання цілісності та конфіденційності взаємопов'язані між собою, тому для відповіді на одне з них використовують інше.

Забезпечує шифрування даних узагальнена схема криптосистеми (схема шифрування):



Відправник створює текст вихідного повідомлення M в незашифрованому вигляді, яке по незахищеному каналу потрібно передати отримувачу. Проте, з метою перехопити секретну інформацію за каналом стежить зловмисник, котрий має намір розкрити її. Для того, що він не зміг дізнатися вмсіт повідомлення M, відправник зашифрував його за допомогою перетворення E_k і отримує шифротекст $C = E_k(M)$, який і передає одержувачу.
Одержувач може дешифрувати вміст за допомогою зворотнього перетворення $D = E_k^{-1}$ і отримує повідомлення у первинному (незашифрованому) вигляді M.
Тобто,

$$D_k(C) = E_k^{-1}(E_k(M)) = M$$

Варто зауважити, що:

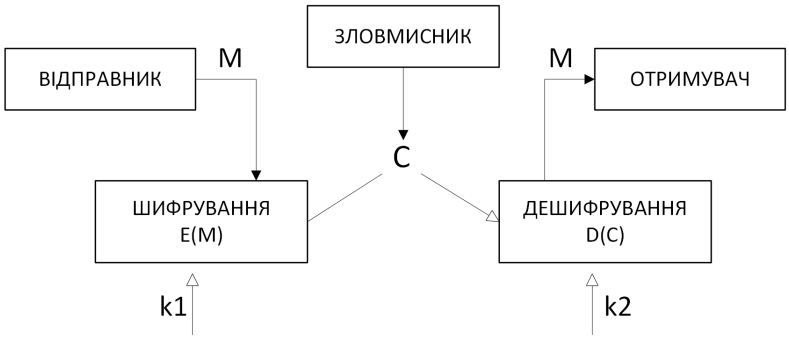
- криптографічний ключ k - це параметр, за допомогою якого вибирається окреме перетворення, яке використовувалося.
- криптографічна система - це однопараметричне сімейство $(E_k)_{k\in\overline{k}}$ оборотних перетворень

  $$E_k : \overline{M} \to \overline{C}$$

  з простору $\overline{M}$ повідомлень незашифрованого тексту в простір шифрованих. Ключ k вибирається з кінцевої множини $\overline{M}$, тобто з простору ключів.

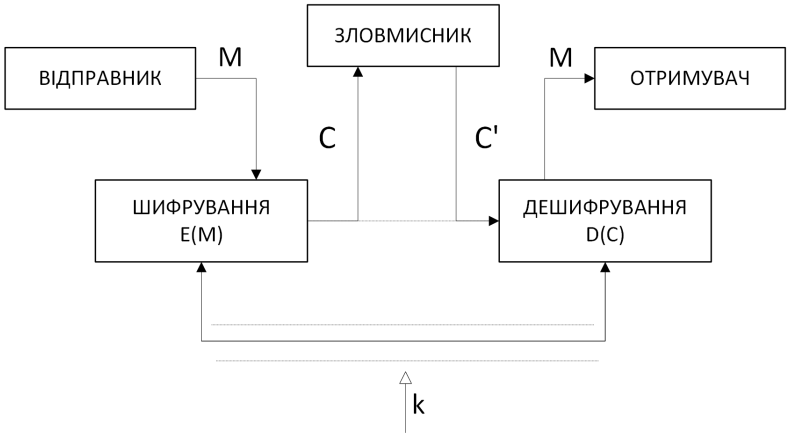
- перетворення шифрування буває симетричним та асиместричним щодо зміни дешифрування, тому визначають симетричні (одноключеві) та асиметричні (багатоключеві з відкритим ключем) класи криптосистем.

Раніше ми розглядали схему симетричної криптосистеми з одним секретним ключем та з однаковими секретними ключами в блоці шифрування та дешифрування. Зараз розглянемо узагальнену схему асиметричної криптосистеми з двома різними ключами k_1 та k_2:



Один з ключів у цій криптосистемі відкритий, а інший секретний. У випадку симетричної криптографічної системи ключ потрібно передавати захищеним каналом (на рис. штриховою лінією). Щодо асиметричної криптосистеми, варто пам'ятати, що по незахищеному каналу передають тільки відкритий ключ, а секретний зберігають на місці його генерації.

Часто буває так, що зловмисники активно намагаються не тільки перехопити важливі повідомлення, вираховуючи всі шифротексти, але і змінити вміст інформації, що передається каналом.



Спроба зі сторони зловмисника дешифрувати шифротекст C для одержання відкритого повідомлення M або зашифрувати власний текст M' для отримання підробленого шифротекста C, не маючи справжнього ключа, називається *криптоаналітичною атакою*. Якщо криптоаналітик не може, без справжнього ключа досягти своєї мети,а саме, вивести M з C або C' з M', то така криптосистема вважається *криптостійкою*.

Криптоаналіз - це наука, яка досліджує дешифрування вихідного тексту зашифрованого повідомлення без доступу до ключа.

Стеганографія забезпечує приховування факту існування повідомлення, тому зловмисник нічого про нього не знає про зв'язок між шифруванням і дешифруванням.

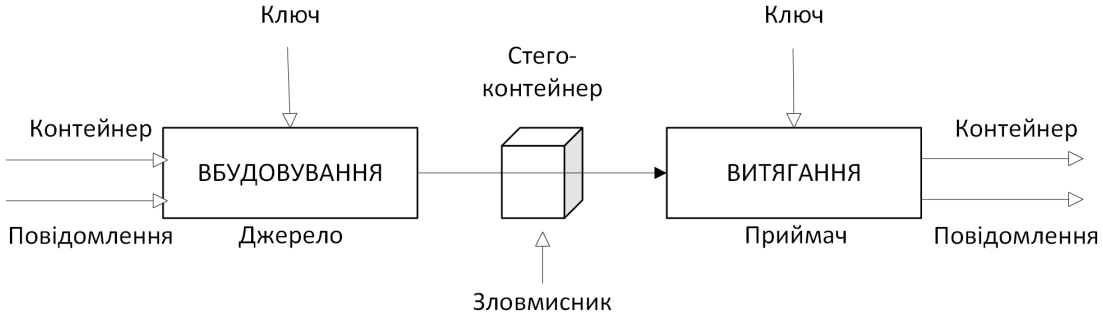
1.2 Задача та класифікація стеганографії

Стеганографія - це наука, яка досліджує методи та алгоритми захисту інформації приховуванням факту її існування, заснована на різних принципах.
В сучасності виділяють два напрямки стеганографії:

- технологічну (методи, які грунтуються на хімічних або фізичних властивостях матеріальних носівів);
- інформаційну (методи лінгвістичної та комп'ютерної стеганографії).

Зважаючи на розвиток обчислювальних машин та можливість передачі інформації у вигляді текстів, програм, звуку, фото чи відео, почала розвиватися стеганографя.

Опишемо загальну схему стеганографічної системи:



Її функціонування задовольняє наступні положення:

- для повідомлення, що вбудовується, контейнера й ключа стеганографічне перетворення означає стегограму;
- за наявності стегоключа зворотне, стегоперетворення дозволяє витягти приховане повідомлення;
- зловмисник (або стегоаналітик) не має апріорно точних відомостей про факт існування в контейнері прихованого повідомлення.

Мета стеганографічного аналізу - це моделювання систем та дослідження їх моделі, побудова методик виявлення, фчи існує інформація, захоплення, редагування чи видалення її вмісту. Стегосистема вважається зламаною, якщо факт існування прихованого повідомлення в перехопленому контейнері доведено.
За рівнем секретності стегоситеми поділяються на:

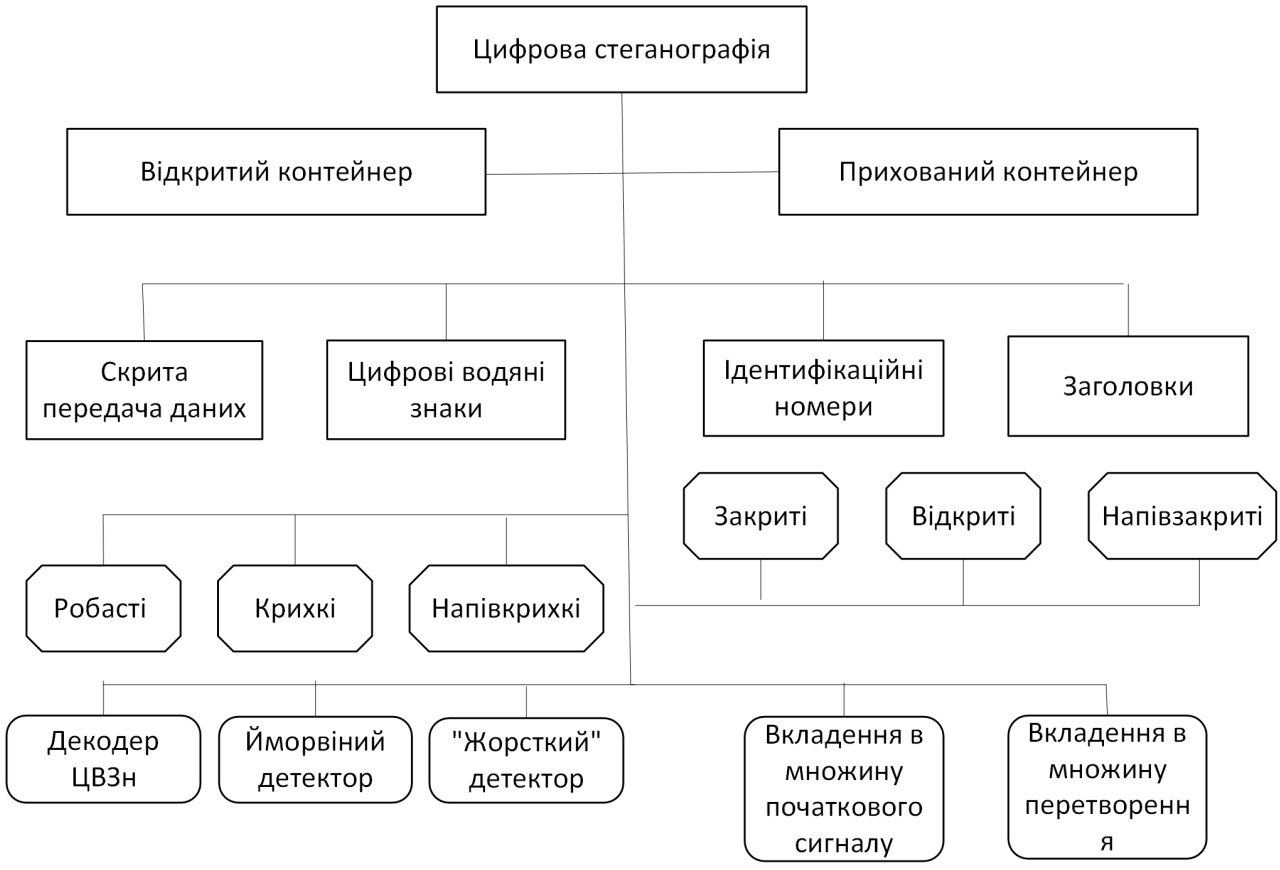
- теоретично стійкі (приховування інформації лише в елементах контейнерів, значення яких не перевищує рівень шуму або помилок квантування, і при тому доведено, що створити метод виявлення прихованої інформації неможливо);
- практично стійкі (здійснюють модифікацію елементів контейнера, зміни яких можуть бути виявлені, але у зловмисника поки не вистачає для цього ресурсів);
- нестійкі (приховують інформацію так, що її можна виявити наявними аналітичними засобами).

Цифрова стеганографія - заснована на приховуванні або вбудовуванні додаткової інформації в цифрові об'єкти, викликаючи при цьому певні їх спотворення. Вважається, що об'єкти є мультимедійними і внесення спотворень, які нижче порога чутливості людини, не призводять до помітних змін.

Виділяють декілька напрямів цифрової стеганографії:

- вбудовування інформації з метою її прихованої передачі;
- вбудовування цифрових водяних знаків (watermarking) (застосовуються для захисту від копіювання та несанкційного використання, можуть бути видимі та невидимі, аналізуються спеціальним докером, що виносить висновок про їх коректність);
- вбудовування ідентифікаційних номерів (fingerprinting) (відмінність від watermarking полягає в тому, що в останньому кожна захищена копія має свій унікальний вбудований номер, який дає змогу виробнику в майбутньому відслідковувати свою роботу (працю));
- вбудовування заголовків (captioning) (зберігання різнорідно представленої інформації в єдиному цілому (напрямів, де в явному вигляді відсутній потенційний зловмисник)).

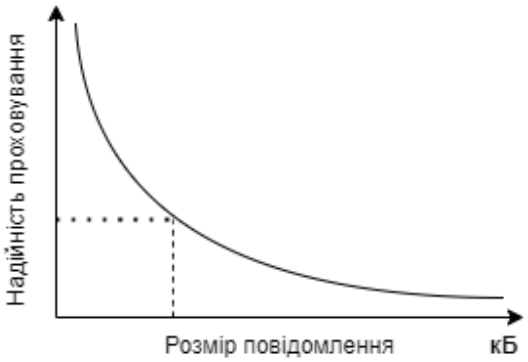
Розглянемо класифікацію систем цифрової стеганографії:



Їх можна класифікувати за:

- вибором контейнера;
- призначенням;
- стійкістю;
- наявністю ключа;
- декодером;
- принципом приховування (спектральні та безпосередньої заміни).

Різні методи стеганографії для створення контейнерів повинні мати фіксований розмір повідомлення та значення захищеності (надійності). Збільшуючи повідомлення, збільшиться розмір контейнера, тому файл, який буде виконувати роль контейнера може викликати підозри. Отже, визначають залежність захищеності приховування від розміру повідомлення:



Змінивши деякі якості контейнера, можна забезпечити високу надійність або великий розмір повідомлення, проте потрібно пам'ятати, що ріст одного призводить до зменшення іншого, тому рекомендовано притримуватись оптимального варіанту.

Отже, можна зауважити, що задача стеганографії - це вбудувати інформацію в контейнер таким чином, щоб зловмисник не міг знайти відмінність між оригінальним та з штатною інформацією контейнером.

1.3 Математична постановка задачі

Процес стеганографічного перетворення можна описати залежностями типу:

$$E : C \times M \to S; \tag{1}$$

$$D : S \to M, \tag{2}$$

де $S = \{(c_1, m_1), (c_2, m_2), ..., (c_q, m_q), \} = \{s_1, s_2, ..., s_q\}$ - множина стеганограм (заповнених контейнерів).

Залежність (1) - процес приховування інформації, а залежність (2) - витягування інформації. Обов'язковою умовою є $m_a \neq m_b$, тобто відсутність «перетину», тоді

$$E(c_a, m_a) \cap E(c_b, m_b) = \emptyset$$

де $m_a, m_b \in M$, $(c_a, m_a), (c_b, m_b) \in S$.

Також стеганосистему представляють у вигляді сукупності $\sum(C, M, S, E, D)$ - контейнерів, перетворень та повідомлень, що їх з'єднують. Контейнери c потрібно обирати так, щоб порожній контейнер майже не відрізнявся від заповненого.

Означення 1 *Нехай C - непорожня множина, тоді функція $sim(C) \to (-\infty, 1]$, є функцією подібності на множині C, якщо $\forall x, y \in C$ справедливо, що $sim(x, y) = 1$ у випадку $x = y$ і $sim(x, y) < 1$ при $x \neq y$.*

Якщо,

$$sim[c, E(c, m)] = 1, \forall m \in M, c \in C$$

(sim - функція подібності), то стеганосистема вважається надійною.

Контейнер можна обрати довільно (сурогатний метод) та в конкретному випадку підбором найпридатнішого контейнера (такого, що при перетворенні найменше зміниться). В другому випадку контейнер обирають спираючись на умову:

$$c = max(sim[x, E(x, m)]) \tag{3}$$

Зворотне та пряме перетворення (E та D) повинні відповідати одне одному та задовільняти умову того, що невелике викривлення контейнера (δ) не повинне допустити викривлення прихованої інформації.

$$E(c, m) \approx E(c + \delta, m)$$

або

$$D[E(c, m)] \approx D[E(c + \delta, m)]m \tag{4}$$

2 Структури мультимедійних даних

2.1 Приховування інформації у нерухомих зображеннях

Більшість засобів стеганографії орієнтуються на передачу даних повідомленями. Часто інформація передається у вигляді зображень. Обробляючи зображення, контейнери враховують формат файлу та методи стиснення. Складність процедур стеганографії залежить від формату контейнера, методу та об'єму стегосистеми.

Розглянемо BMP формат.
Розробники вирішили не прив'язувати BMP формат до якихось платформ. Файли такого формату складаються з декількох частин:

- заголовок (починається з «BM», після чого слідує довжина файлу, що виражена в байтах). Далі 4 байти зарезервовані для наступних розширень формату. Закінчує заголовок зсув від початку файла до збережених у ньому даних;
- інформаційний заголовок (містить власну довжину, розмір зображення, роздільну здатність й інщі параметри);
- палітра (таблиця кольорів) - 256 елементів по 4 байти;
- дані зображення (зображення, записане рядками зліва направо і знизу вгору).

Важливою характеристикою є глибина кольору, яку вимірюють в бітах на піксель. Варто зауважити, що у файлах BMP формату не відбувається стиснення, але в заголовку файла зарезервовано поле, яке дорівнює одиниці, що означає, що дане зображення не стиснуте. Кількість байтів в пам'яті для зберігання зображення визначає роздільна здатність, яка вимірюється в пікселях (важливо для масштабування).

Розглянемо GIF формат.

Розробка фірми CompuServe призначена для визначення растрових кольорових зображень. Використовуючи цей формат, можна висвітлювати якісні зображення з кращою роздільною здатністю.
Розглянемо та опишемо загальний формат GIF файлу:

- Ідентифікатор GIF.
  На початку файлу повинен бути «підпис» з шести символів (GIF 87a). Він гарантує, що дані є потоком даних GIF формату. «87a» часто вказують на номер версії конкретного визначення GIF формату та використовуються як лінк на документ з відповідним описом GIF, залежно від номера версії.
- Дескриптор екрана.
  Описує параметри для наступних зображень GIF формату та визначає існування даних про палітру і «глибину» екрана, розміри зображення чи логічного екрана.
- Глобальна таблиця кольорів.
  Рекомендовано для зображень з точною передачею кольорів. Виявити наявність цієї таблиці допоможе поле «M» в п'ятому байті дескриптора екрану, що дорівнює нулю. Слід зауважити, що кількість елементів таблиці відповідає значенню 2^n, де n - кількість бітів або пікселів, причому всі елементи складаються з трьох байтів, значення яких визначає густину трьох кольорів: червоного, зеленого та синього.

- Дескриптор зображення.
  Описує локацію та розмірність наступного зображення в просторі, визначеним дескриптором екрану. Починається з роздільника зображень, який відповідає за синхронізацію при вході в дескриптор («$0 \times 2C$» або «,»). Символ, що перебуває між роздільником та кінцем попереднього зображення, ігнорується з метою допущення присутності декількох форматів і правильного ігнорування їхніми декодерами при наступних GIF модифікаціях.

- Локальна таблиця кольорів (повторюється від 1 до n-разів).
  Забезпечує переведення колірної таблиці та кількості бітів на піксель до тієї, що визначались після дескриптора екрану. Визначає розмір пікселя та кількість елементів в наступній колірній таблиці.

- Растрові дані.
  Формат зображення визначений як масив значень номерів пікселів, що утворють зображення. За замовчуванням пікселі запам'ятовуються зліва направо та зверху вниз.

- Термінатор GIF.
  Для фіксації процесу завершення файлу зображення GIF використовується один байт даних, який розглядається як останній файл символу. Його, значення завжди дорівнює 3Bh (шістнадцятирічний код символа ";"). Термінатор (завершитель) сприймається декодером GIF як сигнал закінчення процесу обробки зображення. Він повинен перебувати в кожному файлі GIF. За узгодженям декодування програми повинні робити паузу і чекати подальших дій, вказуючи, що користувач готовий до продовження роботи. Зазвичай декодируюча програма дає команду на перетворення графічного режиму і повертається до попереднього процесу.

- Розширений блок GIF.
  Визначення методу упаковки всередині потоку даних GIF. Зазвичай розміщується перед дескриптором зображення або перед термінатором. Кожен декодер повин розпізнати наявність розширеного блоку та читати його, якщо вони не можуть опрацювати адаптивний код. Це дозволить старим декодерам обробляти зображення без додаткових адаптивних ресурсів.

Розглянемо JPEG формат.

Формат JPEG набув популярності завдяки своєму алгоритму стиснення, який дозволяє значно стискати інформацію і при цьому зберігати основні кольори та яскравість. Файли в цьому форматі можуть мати невеликий розмір, але мати прийнятну якість зображень.Проте максимальний стиск може значно змінити вхідні дані.На основі формату JPEG були створені наступні розширення:JFIF, JPE і найпопулярнішим варіантом є JPG.
Опишемо структуру JPEG-перетворень:

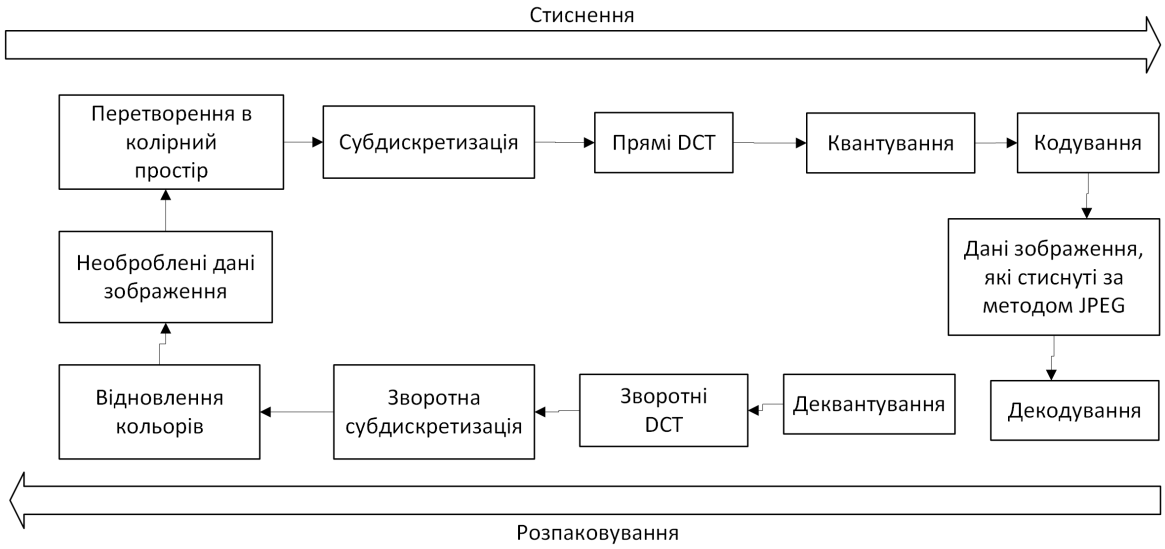
- модифікація зображення в прийнятний колірний простір;

- субдискретизація компонентів кольоровості усередненням груп пікселей;

- використання фрагментних косинус-перетворень з метою зменшення надмірності даних зображення;

- квантування всіх блоків коефіцієнтів DCT з використанням вагової функції, вдосконалених, з врахуванням візуального сприйняття людини;

- кодування даних зображення, застосовуючи алгоритм Хаффмена з метою знищення надмірності інформації.



Важливо зауважити, що розшифровування здійснюється в оберненому порядку.

Для приховування інформації всередині зображення використовують спеціальні методи. Найпоширенішими вважають: метод заміни найменш значущого біта, маскування та фільтації, перетворення зображення на обкладинці. Їх застосовують з різним ступенем успіху для різних типів файлів зображення.

Опишемо деякі з них:

- *Метод заміни найменш значущого біта.*
  Найпоширенішим способом приховування інформації є використання техніки заміни найменш значущого біта. Через відносно просту реалізацію цей метод є популярним, що робить його менш захищеним. Щоб приховати інформацію в зображенні, потрібно звернути увагу на обкладинку. Зважаючи на те, що метод використовує біти кожного пікселя, треба використати формат стиснення без втрат, інакше інформацію можна втратити в модифікаціях алгоритму.

  Використовуючи 24-бітне колярове зображення можна використати частину компонент червоного, зеленого та синього кольору, тому, в кожному пікселі можна зберегти 3 біти. Тоді, зображення розміром 800×600 містить 144000 біт або 180000 байт прихованої інформації. Таку сітку розглядають як 3 пікселі 24-бітового кольорового зображення, використовуючи 9 байт пам'яті:

$$\begin{array}{lll} (00100111 & 11101001 & 11001000) \\ (00100111 & 11001000 & 11101001) \\ (11001000 & 00100111 & 11101001) \end{array}$$

  Коли вставляється символ A, двійкове значення якого дорівнює 10000001, виходить така сітка:

$$\begin{array}{lll} (00100111 & 11101000 & 11001000) \\ (00100110 & 11001000 & 11101000) \\ (11001000 & 00100111 & 11101001) \end{array}$$

  В даному випадку, щоб успішно вставити символ, було достатньо 3 біти. В середньому лише половину бітів зображення треба змінити, щоб засекретити дані,

використовуючи максимальний розмір обкладинки. Ці зміни є занадто малі,тому людина не зможе їх розпізнати, а повідомлення вважають прихованим.

Також як джерело обкладинки можна використовувати 8-розрядні зображення. В такому разі для представлення пікселів використовуватиметься один байт, але варто зауважити, що це значно змінить відтінок кольору.

Недоліками цього методу є вимога великого зображення та ризики втрати секретного повідомлення при стисканні.

- *Метод маскування та фільтрації.*
  Як правило, такі методи обмежують 24 бітами. Їх порівнюють до водяних знаків, створюючи позначки, трохи змінюючи яскравість на окремих частинах зображення.

  Щодо різних видів обробки, стиснення чи обрізання, методи є надійніші, ніж, скажімо стиль заміни найменш значущого біту. Дані не приховуються на рівні шуму, а перебувають у видимій частині зображення.

- *Метод перетворення.*
  Складнішим способом приховати секретні дані в зображенні використовуючи модифікації дискретних косинусних перетворень, які застосовуються алгоритмом стиснення JPEG, щоб перетворити послідовні блоки зображення розмірами 8×8 пікселів в 64 коефіцієнти дискретних косинусних перетворень. Такі коефіцієнти обчислюються за формулою:

$$F(u,v) = \frac{1}{4}C(u)C(v)\left[\sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y)\cdot\cos\frac{(2x+1)u\pi}{16}\cos\frac{(2y+1)v\pi}{16}\right],$$

  де $C(x) = 1/\sqrt{2}$, коли $x = 0$ та $C(x) = 1$ в інших випадках. Після обчислення коефіцієнтів виконують операцію квантування:

$$F^Q(u,v) = \left\lfloor\frac{F(u,v)}{Q(u,v)}\right\rfloor,$$

  де $Q(u,v)$ - 64 елементна таблиця квантування.

2.2 Приховування інформації у текстових файлах

До методів приховування інформації у текстових файлах або тих, що використовують текстові контейнери належать:

- синтаксичні методи (зміна стилізації, пунктуації та стуктури тексту);
- семантичні методи, які обумовлюють два синоніми, що відповідають значенням прихованих бітів (потребують таблиці синонімів);
- методи довільного інтервалу, що базуються на зміні інтервалу між реченнями, зміні пробілів вкінцях рядків, зміні пробілів між словами тексту.

Розглянемо деякі методи текстової стеганографії:

- *Форматування.*
  Суть методу: розусунути рядок збільшенням пробілів між словами та реченнями. Одному пробілі ставлять у відповідність 0 біт, а двом пробілам, наприклад, один

біт. Насправді, цей метод не є зручним, бо несе за собою багато підозр наявності стега, власне, через дивний та неохайний вид тексту. Проте правильний розподіл пробілів - довгі пробіли переносячи в кінець, а короткі - останнім словом рядка, зможе забезпечити акуратність та читабельність тексту.

- *Зміни порядку проходження маркерів кінця рядка.*
  CR/LF визначає байдужість пригніченої кількості засобів відображення текстових даних до порядку проходження символів перекладу рядка (CR) і повернення каретки (LF), що обмежують рядок тексту. Порядок проходження CR/LF відповідає 0, а обернений LF/CR означає 1.

- *Хвостових пропусків.*
  Метод полягає в дописуванні вкінці від 0 до 15 пробілів, які кодуватимуть значення напівбайта.

- *Знаків однакового накреслення.*
  Метод полягає в заміні (бітів 1) або відмову від заміни (бітів 0) символу кирилиці на латинський такого ж накреслення.

- *Двійкових нулів.*
  Є подібним до попереднього та полягає в заміні першого в групі із двох або більше внутрішніх пробілів двійковим нулем (бітів 1), або відмову від заміни (бітів 0).

Дослідження ефективності цих методів на щільність було проведене на одному з томів книги «Мертва вода» розміром понад 126729 тисяч байтів та налічує 2143 рядки, що вирівняли на 65-символьну, по 4-символьному абзацному відступі. Результати дослідження подано в таблиці:

Метод	**Знаків стега**	**Щільність,**%
Чергування маркерів кінця	267	0.21
Вирівнювання пробілами	411	0.32
Двійкові нулі	740	0.58
Хвостові пропуски	1071	0.85
Знаки одного накреслення	4065	3.21

Результати не є точними і залежать від властивостей контейнера та властивостей того, що міститься в стекові. Можна зауважити незвичайно високу ефективність методу знаків однакового накреслення.

2.3 Приховування інформації в аудіосигналах

Спочатку опишемо вимоги до стегоситеми, які забезпечать вбудовування даних в аудіофайли:

- секретне повідомлення має бути стійким до фільтрації, різних шумів, втрат зі стистом, аналогових та цифрових модифікацій;

- у випадку ЦВДЗ бажання позбутись секретного повідомлення може пошкодити контейнер;

- секретне повідомлення не має докладати в сигнал спотворення помітні людині;

- секретне повідомлення не має значно впливати на сатистику контейнера.

Загалом поділять три класи аудіосигналів:

- розмова телефонної якості (300 – 3400 Гц);
- широкосмугова мова (50 – 7000 Гц);
- широкосмугові аудіосигнали (20 – 20000 Гц).

Розглянемо декілька методів приховування інформації в аудіосигналах:

- Метод найменш значущого біта.
  Суть полягає в зміні найменш значущого біта у всіх точках вибірки, що представлена бітовою послідовністю. Техніка має слабку стійкість, тому є велика ймовірність знищення інформації через шуми в каналі, проте перевагою цього способу є можливість приховування великого об'єму інформації.

- Метод фазового кодування.
  Його суть - це зміна фази вихідного звукового сегмента на опорну фазу, характер зміни якої відображає собою інформацію, яку треба приховати. Щоб зберегти різницеву фазу, потрібно узгодити фази між сегментами. Заміна співвідношення періодів між кожними частотними складовими призводить до значного його розсіювання. Поки перетворення фази є відносно малим, приховування несприйнятне на слух.

- Метод розширення спектру.
  Для його реалізації потрібно сигнал даних помножити на сигнали несучої псевдовипадкової шумової послідовності, що характерезує широкий частотний спектр, який в свою чергу забезпечує роширення спектру даних. Далі послідовність розширених показників слабшає та приєднується до вихідного сигналу в ролі адаптивного випадкового шуму.

- Метод приховування інформації використовуючи ехо-сигнал.
  Метою є введення в аудіо-файл ехо-сигналу. Інформація приховується при заміні початкової амплітуди, швидкості загасання і затримки. Через зменшення затримки між початковим і ехо-сигналом людина не зможе відчути різницю між ними. Також можна змінювати рівень первинної амплітуди чи згасання до рівня чутливості середньостатистичної людини. Сигнал розділяють на менші частини з метою закодовування в початковий сигнал більше ніж, одного біта. При цьому кожну частину розглядають як окремий сигнал.

- Метод приховування інформації за допомогою вставки тонів.
  Базується на поганій чутливості та нечітких низьких тонах при наявності компонент вищого спектра.

Підсумовуючи, деякі методи мають багато недоліків кореляційних приховуванню відлуння, поширенню спектру, чутливістю до шумів та надійністю. Фазове кодування зручне лише для малих файлів, бо потребує відносно багато часу на кодування. Найбезпечнішим та найефективнішим однаково вважається спосіб заміни найменш значущого біта, що дозволяє приховувати великий об'єм інформації.

2.4 Класифікація атак на стегосистеми

Стегосистему вважають зламаною, якщо зловмисник зміг довести, що в перехопленому контейнері наявна прихована інформація. Якщо йому це не вдалось, то кажуть, що система стійка.

Часто криптографію використовують для прихованого спілкування. Її алгоритми, безпека яких може доводитись або простежуватись до відомих непростих математичних проблем, широко доступні. Але на відміну від стеганографії, алгоритми криптографії створюють повідомлення, які розпізнаються як зашифровані, хоча їх вміст залишається секретним. Стеганографія вбудовує приховане повідомлення в інше більш широке, яке є носієм. Мета полягає в тому, щоб замінити носій непомітно, так, щоб він нічого не виявив: ні факту вбудовування повідомлення, ні самого повідомлення.

Виділяють декілька видів атак на стегосистему, причому перші три притаманні також і криптографії, а наступні лише стеганографії:

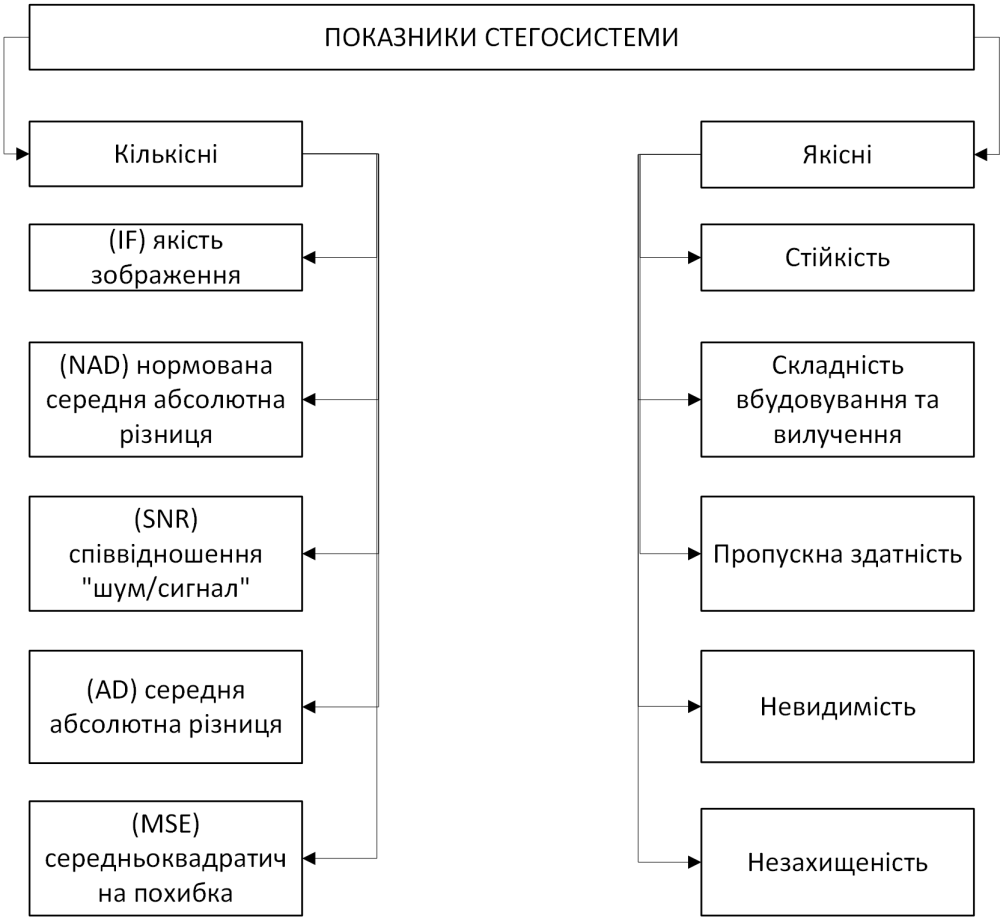
- атака, що грунтується на знанні заповненого контейнера (в зловмисника є щонайменше один заповнений контейнер, що можна використати для виявлення існування стегоканалу, визначення ключа чи добування сектретного повідомлення, що в свою чергу дозволить йому аналізувати інформацію для атаки на інші дані);
- атака, що грунтується на відомому повідомленні (властиво для цифрових водяних знаків, тобто для повідомлень із захистом авторської власності);
- атака, що грунтується на прихованих даних чи заповнених контейнерів, що обирались зловмисником;
- атака, що грунтується на відомому пустому контейнері (порівнявши пустий контейнер та підозрілий можна розкрити факт існування стегоканалу);
- атака, що грунтується на виборі порожнього контейнера зловмисником (він може замінити його на свій для стегоперетворень);
- атака, що грунтується на знанні математичної моделі контейнера (дані в ній є взаємозалежні, але відсутність залежності може бути сигналом про існування секретного повідомлення).

Зламування стеганографічкої системи відбувається декількома етапами:

- розкриття факту існування прихованої інформації;
- добування прихованих даних;
- перетворення секретних даних;
- заборона поширення скритого повідомлення.

2.5 Оцінка ефективності стеганоситеми

Для порівняння методів стеганографії виділяють критерії оцінки стегоситем:



Для оцінки ефективності цих методів використовують кількісні показники, які керують із зображеннями на рівні пікселів. Проте після деяких перетворень ці способи можна застосовувати і для інших мультимедійних представлень даних. Опишемо кожен з них:

- (IF - image fidelity) Якість зображення (визначає степінь відповідності між порожнім та заповненим контейнером).

$$IF = 1 - \frac{\sum_{x=1}^{n}\sum_{y=1}^{m}(C_{x,y} - S_{x,y})^2}{\sum_{x=1}^{n}\sum_{y=1}^{m}(C_{x,y})^2};$$

- (NAD - normalized average absolute difference) Нормована середня абсолютна різниця (вказує на рівень розбіжності між вихідним і з вбудованою інформацією контейнером).

$$NAD = \frac{\sum_{x=1}^{n}\sum_{y=1}^{m}|C_{x,y} - S_{x,y}|}{\sum_{x=1}^{n}\sum_{y=1}^{m}|C_{x,y}|};$$

- (SNR - signal noise ratio) Співвідношення "шум/сигнал"(рівень спотворення, що вноситься в контейнер при приховуванні даних). Визначається відношенням сигналу до потужності фонового шуму.

$$SNR = \frac{\sum_{x=1}^{n}\sum_{y=1}^{m}(C_{x,y})^2}{\sum_{x=1}^{n}\sum_{y=1}^{m}(C_{x,y} - S_{x,y})^2};$$

- (AD - absolute difference) Середня абсолютна різниця (середнє значення модулю різниці між пікселями пустого і заповненого контейнеру (чим більша середньоквадратична різниця, тим нижча якість забраження));

$$AD = \frac{1}{X \cdot Y}\sum_{x=1}^{n}\sum_{y=1}^{m}|C_{x,y} - S_{x,y}|$$

- (MSE - mean square error) Середньоквадратична похибка (визначає середньоквадратичний відхил вибіркового розподілу статистичних даних).

$$MSE = \frac{1}{X \cdot Y}\sum_{x=1}^{n}\sum_{y=1}^{m}(C_{x,y} - S_{x,y})^2$$

$C_{x,y}$ - значення пікселя порожнього контейнера;
$S_{x,y}$ - значення пікселя заповненого контейнера;
n - кількість рядків в масиві C;
m - кількість стовпцців в масиві C.

Для оцінки ефективності стеганографічних методів також використовують якісні показники:

- стійкість - цілісність інформації після певних дій з вихідним контейнером;
- складність вбудовування та вилучення - визначається кількістю дій, що потрібні для вбудовування та виявлення секретного повідомлення.(складність обчислення стеганографічної системи);
- пропускна здатність - розмір інформації в бітах, що передається певним методом в контейнері певного розміру;
- невидимість - забезпечення неспроможності виявлення інформації середньостатистичною людиною без стороніх засобів;
- захищеність - інформація не повинна змінюватись чи видалятись через атаки, що грунтуються на відомих алгоритмах вбудування/вилучення та відомості бодай про один носій прихованого повідомлення.

3 Приклади застовування

3.1 Приховування інформації в BMP форматі методом заміни найменш значущого біта

Існує багато різних алгоритмів, що використовують найменш значущий біт. Найпростіше, починаючи з верхнього лівого і до нижнього правого кута, вписувати секретні дані у всі пікселі зображення. Проте такий метод є не надійним і піддається атакам, але якщо приховані дані розмістити відповідно до якоїсь, залежної від ключа, послідовності, то зламати систему буде складніше. Також можна записувати секретні дані в пікселі порожніх контейнерів, які менш помітні людині. Такий метод дозволить використовувати декілька найменш значущих бітів для одного байта кольору. Поєднання цих методів може покращити стійкість стегосистеми.

Розглянемо загальний алгоритм приховування даних в BMP форматі зображення:

- вікриття файлу контейнера для прихованих даних;
- вибір кольору палітри контейнера (червона, синя, зелена);
- відкриття файлу з даними, які треба приховати;
- конвертування даних в двійковий формат для засекречення в масиви по 8 біт (всі характеристики зображення та інтенсивність кольорів пікселів);
- заміна найменш значущого біта пікселя контейнера бітом прихованих даних відповідно до величини ключа;
- зберігання заповненого контейнера, тобто зміненого файлу, у форматі BMP відповідно до правил зберігання.

Для видобування даних потрібно:

- знати ключ, з яким в контейнері розміщувались дані;
- відкрити файл з секретною інформацію, щоб отримати значення кольорів пікселів повного контейнера;
- використовуючи ключ, вибрати найменш значущі біти кольорів пікселів та згенерувати з них секретні дані.

В ролі контейнера потрібно обрати зображення BMP формату, яке містить масиви червоного, зелененого та синього кольорів, кожен з яких визначається своєю інтенсивністю від 0 до 255.

Дані, які потрібно приховати, зберігаємо в текстовому файлі та обчислюємо кількість символів. Щоб знати кількість символів при видобовуванні, це число записують на почаку файлу.

Також потрібно врахувати вирівнюючі байти, які потрібні для читання та запису BMP формату. Через те, що зображення обробляються в 24-бітному форматі довжини всіх рядків пікселів повинна бути кратна чотирьом байтам, відтак, можлива потреба в додаткових порожніх байтах. Це роблять за формулою

$$b = 4 - (H \times 3)\%4, \tag{1}$$

де b - порожні байти, H - ширина зображення в пікселях, % - остача від цілочисельного ділення.

Вибравши якийсь колір, після вбудовування певної кількості символів секретних даних треба вбудувати їхній двійковий формат в найменш значущі біти пустого контейнера. Почавши з лівого верхнього кута зображення, який буде ключем, по рядках вбудовуємо значення даних.

Сформувавши всі рядки заповненого контейнера, вирівнюючі байти треба зчитати та записати у записати у повний контейнер.

Для видобування даних спочатку потрібно зчитати дані про BMP-файл заповненого контейнера та обчислити кількість дадаткових бітів рядка, після чого потрібні байти кольорів в найменш значущих бітах, у яких є секретні дані. Відомо, що приховувались біти зліва направо і зверху вниз, тому читання та видобування потрібно здійснювати в такому ж порядку.
Також відомо, що на початку прихованого тексту записувалось число кількості символів, тому спочатку потрібно визначити його за формулою:

$$t = \sum_{i=0}^{7} LSB_i \times 2^i, \tag{2}$$

Перевівши значення t, в символ, відповідно до ASCII-таблиці кодування, отримуємо десяткове число.
Кожен отриманий символ записують у файл.

Для більшої ефективності секретне повідомлення рекомендують зашифрувати методами криптографії.

3.2 Приховування інформації в JPEG форматі методом Коха–Жао

Досить поширеним є метод приховування даних в частотній множині зображень, суть якого полягає в зміні величин коефіцієнтів дискретно косинусного перетворення. Нагадаємо, що цей метод використовують для стиснення JPEG.

Метод Коха-Жао можна описати наступним алгоритмом:
Крок 1.
Спочатку початкове зображення розділюють на блоки розміром 8×8 пікселів до яких застосовують дискретно косинусні перетворення, причому в одному блоці приховують один біт даних.
Крок 2.
Застосувати дискретно-косинусне перетворення до кожного блоку. В результаті отримаємо набір матриць коефіцієнтів D_i розмірності 8.
Крок 3.
Обираємо блоки для вбудовування та записуємо в кожен блок по одному біту даних.
Крок 4.
У всіх блоках потрібно вибрати два коефіцієнти перетворення, які симетричні головній діагоналі.
Крок 5.
Щоб передати біт «0», необхідно, щоб різниця абсолютних значень коефіцієнтів дискретно-косину
перетворень була більша ніж якась порогова величина. Щоб передати біт «1», необхідно, щоб різниця була меншою ніж якась порогова величинаю. Тому, наприклад, для

вбудовування нульового біту збільшуть перший, але зменшують другий коефіцієнт на однакову величину та навпаки.
Крок 6.
Четвертий та 5 пункт потрібно виконати для всіх блоків
Крок 7.
Виконати обернене дискретно косинусне перетворння для всіх блоків.

Варто зауважити, що внесення змін в коефіцієнти спотворює початкове зображення за умови, що відносна величина не відповідає засекреченому біту. Зі збільшенням цього значення стегосистема стає стійкішою, проте меншою стає якість зображення.

Для розшифровування інформації у декодері так само, тобто перші чотири кроки такі самі, вибираються коефіцієнти, а рішення про переданий біт приймаються згідно з правилом:
Крок 4.
Потрібно виділити матриці колірних компонент зображення:
Визначимо розмітність матриці і позначимо розмірність блоків розбиття:

- $X := rows\{B\},\ X = 128$ - рядки
- $Y := cols\{B\},\ Y = 128$ - стовпці
- $N = 8$ пікселів - розмірність блоків.
- $Nc = X - Y/N^2,\ Nc = 256$ - кількість блоків розбиття контейнера.

Крок 5.
Виконуємо розбиття матриці на блоки Ic:
Всі блоки призначені для засекречення одного біту інформації M, тому важливо перевірити достатність кількості блоків.
Припустимо, що інформація виглядає наступним чином: $M :=$ *«Назаренко А.Ю., 2005 р.»*. Кількість бітів в ньому: $8\ strl(M) = 200\ bit < N_c = 256$ - допустимий розмір
Крок 6.
До кожного блоку застосовуємо дискретно-косинусне перетворення.
Перша частина перетворення обчислює значення коефіцієнтів, друга обчислює спектральні коефіцієнти для всіх блоків. Результатом є матриця, елементами лівого верхнього кута якої, відповідають коефіцієнтам, що містять інформацію про яскравість блоку (DC-коефіцієнт). Варто зауважити, що коефіцієнти низькочастотного шуму перебувають біля лівого верхнього кута, а високочастотного - біля правого нижнього.

Для атак на даний алгоритм потрібно визначити пари коефіцієнтів дискретно косинусного перетворення та порогові значення.

Висновок

Стеганографія, особливо в поєднанні з криптографією, є потужним захистом інформації. Вона дозволяє людям спілкуватися без можливих підслуховувачів, навіть знаючи, що існує певна форма комунікації. Методи, які використовуються в стеганографії, значно вдосконалилися протягом останніх років, особливо з розвитком обчислювальних машин.

В роботі було описано принципи криптозахисту та класифікацію стеганографії, сформульовано її завдання та математичну постановку, різні алгоритми захисту інформаці в мультимедійних даних, порівняно їхню ефективність, описано класифікацію атак та критерії їх оцінки на стегосистеми. Детальніше проаналізовані методи приховування інформації в BMP форматі методом найменш значущого біта та в JPEG форматі методом Коха-Жао.

Метою стеганографічних алгоритмів є забезпечити приховати факт наявності інформації, що потребує захисту. Вони дають змогу вирішити багато завдань із захисту інформації в обчислювальних системах та мережах телекомунікацій.

Зробивши загальний аналіз методів стеганографії можна поговорити про їхні переваги та недоліки.
До переваг відносяться:

- проста реалізація;
- стійкість до атак;
- унікальність до й після модифікацій;
- доступ до програмного забезпечення для реалізації алгоритмів.

До недоліків :

- чутливіть до спотворень контейнера;
- прояв помилок при розшифровуванні;
- складність вбудовування таємної інформації.

Підсумуємо ефективність алгоритмів захисту даних в різних мультимедійних даних.

Розглядаючи алгоритми приховування даних в нерухових зображеннях можна зробити висновок, що більш надійними є алгоритми маскування, що використовують видимі аспекти зображення, ніж, скажімо, модифікація модетів заміни найменш значущого біта щодо різних видів обробки зображень. Інформація не прихована на рівні «шуму», а перебуває всередині видимої частини зображення, що робить її більш придатною, ніж модифікації модетів заміни найменш значущого біта, якщо використовується алгоритм стиснення з втратами, наприклад JPEG.

Що стосується текстових файлів, то найбільш надійним є метод знаків одного накреслення (він полягає в заміні символів кирилиці на символи латини однакового накреслення), найменш ефективним - метод чергування маркерів кінця (визначає байдужість пригніченої кількості засобів відображення текстових даних до порядку проходження символів перекладу рядка).

Щодо аудіофайлів, деякі методи мають проблеми з чутливістю до шумів та надійністю. Найефективніши вважають метод найменш значущого біта, що дозволяє працювати з великим об'ємом даних та фазове кодування, проте лише для невеликих файлів.

Також варто зауважити, що відеофайли - це зазвичай набір зображень і звуків, відтак багато методів, що представлялись для зображень та аудіо можна застосувати до відеофайлів. Великою перевагою відео є великий та рухомий об'єм даних, які можна приховати всередині.Також відео - рухомий потік зображень і звуків, відтак невеликі, проте помітні в іншому випадку викривлення можуть бути непоміченими людьми через безперервний потік інформації.

Досліджуючи метод заміни найменш значущого біта для приховування даних в BMP форматі, видно, що даний алгоритм є простий і при тому має можливість приховувати великий об'єм інформації. Проте він є слабким до різних атак.

Що стосується методу Коха-Жао для приховування інформації в JPEG форматі, він стійким до більшості відомих стегоатак, зокрема до афінних, геометричних та атак стисненням, але, має низьку пропускну здатність.

Щодо майбутнього стеганографії, найширше використання стегометодів буде в області цифрових водяних знаків. Постачальники інтелектуального матеріалу хочуть захистити свої дані авторськими правами від незаконних поширень. Також базою для розвитку стеганографії є сучасні методи в галузі штучного інтелекту та комп'ютерних технологій.

Література

[1] *О.О. Кузнецов, С.П. Євсеєв, О.Г. Король.* Стеганография : навч. посібник // Вид. ХНЕУ, 2011. – 232 с.

[2] *О.К. Юдін, Р.В. Зюбіна, О.В. Фролов.* Аналіз стеганографічних методів приховування інформаційних потоків в контейнери різних форматів // Радіоелектроніка та інформатика, 2015. – №3. – С.13-21.

[3] *Романцев Ю.В, Тимофеев П.А, Шаньгин В.Ф.,* Защита информации в компьютерных системах и сетях // Под. ред. В.Ф. Шаньгина. -2-е изд., перераб. и доп. - М.: Радио и связь, 2001. - 376 с.: ил.

[4] *В.Г. Грибунин, І.Н. Оков, І. В. Туринцев.* Цифровая стеганография – М.: СОЛОН-ПРЕСС, 2009. // 272с.

[5] *Конахович Г. Ф., Прогонов Д. О., Пузиренко О. Ю..* омп'ютерна стеганографічна обробка й аналіз мультимедійних даних [підручник]. // «Центр навчальної літератури», 2018. — 558 с.

[6] *Bilal A. Shaw..* Quantum steganography and quantum error-correction // University of Southern California. – 2010. – P.137.

[7] *Б. В. Дурняк, Д. В. Музика, В. І. Сабат.,* Стеганографічні методи захисту документів // — Львів : Укр. акад. друкарства, 2014. — 159 с. : іл., портр. ; 21 см. — На паліт.: Інформ. технології. — Частина тексту парал. укр., англ. — Бібліогр.: с. 149—159 (118 назв). — 300 пр.

[8] *N.N. El-Emam,.* Hiding a large amount of data with high security using steganography algorithm // Journal of Computer Science 3 (2007) 223-232.

[9] *M. Kharrazi, H. T. Sencar, and N. Memon,.* Image Steganography: Concepts and Practice // April 22, 2004 1:49 WSPC/Lecture Notes Series: 9in x 6in

[10] *N.J. Hopper, J. Langford, and L. von Ahn* "Provably secure steganography," Advances in Cryptology: CRYPTO // August, 2002.

[11] *Гарасим Я.С, Романенко А.В., Хапко Р.С.,* . LaTeX: створення математичних документів: Навч. посібник. // Львів: Видавничий цент ЛНУ імені Івана Франка 2002. 140 с.