УДК 004.056

В.Ю. КОРОЛЬОВ, О.М. ХОДЗІНСЬКИЙ
Інститут кібернетики ім. В.М. Глушкова НАНУ, м. Київ

JPEG СТЕГАНОГРАФІЯ НА БАЗІ ТЕОРЕТИКО-ЧИСЕЛЬНИХ ПЕРЕТВОРЕНЬ

У роботі розглянуто задачу приховування і виявлення інформації у цифрових зображеннях методами стеганографії та стегоаналізу. Запропоновано новий метод стеганографії для зображень у JPEG-форматі на основі теоретико-чисельних перетворень. Розроблено структурну схему стегоаналітичних досліджень великих масивів мультимедійних файлів на основі хмарних технологій обробки даних.

Ключові слова: стеганографія, стегоаналіз, JPEG, теоретико-чисельні перетворення, хмарні технології.

V.YU. KOROLYOV, O.M. KHODZINKYI
Glushkov Institute of Cybernetic of Ukrainian National Academy of Sciences

JPEG STEGANOGRAPHY BASED ON THEORETIC NUMBER TRANSFORMATION

Abstract - The topics of article are problems of hiding and revealing information in digital image steganography techniques and stegoanalysis. A new method for image steganography in JPEG-format based on theoretical and numerical changes is given. The block diagram stegoanalityc studies of large volumes of multimedia files based on the cloud technology data is discussed.

Keywords: steganography, stegoanalysis, JPEG, theoretical and numerical transformation, cloud technology.

Вступ. За останні 30 років з розвитком цифрових систем передачі даних з'явилось багато досліджень в області цифрової стеганографії. Під цифровою стеганографією розуміють приховування інформації у потоках даних (цифрових сигналах). Комп'ютерна стеганографія [1–4] досліджує способи приховування інформації у даних, сформованих для використання у комп'ютеризованих системах, зокрема у файлах, потоках даних, пакетах протоколів. Сьогодні комп'ютерна стеганографія є самостійною теоретично-прикладною галуззю наукових досліджень, що вивчає методи і способи приховування повідомлень та стійкість стеганографічних систем до розкриття. Стеганографія є одним із напрямків досліджень більш широкої науково-практичної області – Information Hiding, яка вивчає способи і методи прихованої передачі повідомлень або зберігання даних.

На відміну від криптографії, яка розробляє методи шифрування даних для подальшого зберігання або передачі, які надсилаються по загально доступним каналам і тому можуть безпосередньо аналізуватись супротивником, метою стеганографії є прихована передача інформації у відкритих наборах або потоках даних у спосіб, що виключає виявлення прихованої складової серед множини інших носіїв.

Сучасний стан розвитку стеганографії

Поява стеганографічних програмних додатків пов'язана з тим, що в практиці інформаційної безпеки виникає задача не тільки зробити інформацію недоступною порушнику, але і приховати факт її передачі. Також не менш актуальними є інші задачі, які розв'язуються за допомогою методів стеганографії: вбудовування цифрових водяних знаків для захисту авторських прав і майнових прав на цифрову інформацію різного роду, вбудовування цифрових ідентифікаційних міток (тегів), призначених для маркування об'єктів у хмарних сховищах даних.

У дослідженнях стеганографія широко використовує методи суміжних наукових теорій: математичної статистики, теорії інформації, криптографії, цифрової обробки сигналів і зображень та ін. Для сучасних стеганографічних програм характерне комплексне використання алгоритмів стеганографії і криптографії та адаптація алгоритмів приховування даних до характеристик мультимедійного носія прихованого повідомлення. Зважаючи на високу обчислювальну складність стеганографічних чисельних експериментів актуальною науковою проблемою є створення загальної методології планування досліджень і накопичення статистики з використанням теорії дискретної оптимізації.

Класифікація стеганографічних алгоритмів для JPEG-формату

Стеганографічні алгоритми для зображень у JPEG-форматі ґрунтуються на варіаціях методу найменш значимого біту (НЗБ). Розрізняють два типи алгоритмів. Перший тип алгоритмів вносить зміни у спосіб квантування значень зображення. Другий – замінює значення блоків зображення після квантування так, щоб вони кодували задану послідовності біт. Крім стеганографічних алгоритмів також виділяють способи приховування інформації у службових полях даних зображень у JPEG-форматі. Аналіз цих способів [1] виходить за рамки теми даної роботи.

Сучасні стеганографічні алгоритми для зображень у JPEG-форматі переважно будують на основі другого типу алгоритмів. Це пояснюється тим, що другий тип алгоритмів дозволяє приховати більшу кількість даних без суттєвих змін стандартизованих способів квантування, які характерні для алгоритмів першого типу, і є очевидною для аналітика демаскуючою ознакою застосування стеганографічних перетворень. Слід зазначити також, що другий тип алгоритмів потребує застосування складнішого математичного забезпечення, більш трудомісткий для програмування та має вищу обчислювальну складність, а також часто потребує побудови математичної моделі процесів приховування даних.

Стан останніх досліджень і публікацій

Незважаючи на розробку нових алгоритмів і методів приховування даних, комерційно успішних

виключно стеганографічних програмних продуктів немає. Тому у технологіях захисту інформації стеганографія є допоміжним, другорядним засобом, що рідко використовується. Трапляються повідомлення, що стеганографія застосовувалась іноземними шпигунами та ісламськими терористичними групами.

Активно розвиваються також похідні від стеганографії методи приховування даних: цифрові водяні знаки, маркування, у меншій мірі проекти протоколи прихованої передачі даних на базі обміну мультимедійними носіями, найчастіше фотографіями.

Стеганоаналіз [1–4] розвивається в основному шляхом укрупнення і групування моделей виявлення прихованих даних шляхом використання методів інтелектуального аналізу даних за допомогою набору статистик, за множиною метрик та регресійних параметрів, які задає оператор. Аналіз виконується на високопродуктивних паралельних обчислювальних архітектурах. Характерно, що кількість наукових робіт присвячених стеганоаналізу, суттєво менша кількості робіт присвячених стеганографічним алгоритмам приховування.

Виділення невирішених частин проблеми. Формулювання цілей дослідження

В наших роботах [2–4] запропоновано нову дуальну методологію досліджень: оптимальне планування чисельних експериментів по вбудовуванню даних у мультимедійні носії або контейнери і інтелектуальний аналіз отриманих результатів за широким набором метрик, параметрів з подальшим узагальненням і трактуванням методами штучного інтелекту.

Схема стеганографічної передачі даних

Розглянемо загальну схему стеганографічної передачі даних, або *схему стеганографічної системи*, зображену на рис. 1, що використовується для передачі даних при наявності активного та пасивного супротивника. На стороні *відправника повідомлення* приховується у контейнері за допомогою стеганографічного перетворення. Таким чином *пустий контейнер* перетворюється на контейнер з повідомленням, тобто на *заповнений контейнер*. Потім контейнер відправляють по відкритим каналам *одержувачу*, де повідомлення виймають з контейнера, маючи для цього відповідні засоби.

Активний супротивник має можливість вносити зміни у контейнер, в процесі його руху по каналах зв'язку від відправника до одержувача. Вважається, що ні на стороні відправника, ні на стороні одержувача не знають, що контейнер міг бути змінений. Враховувати активного супротивника у схемі потрібно, коли розглядається стійкість стеганографічної системи до знищення повідомлень, їх модифікації або зниження пропускної здатності прихованого каналу передачі даних.

Рис. 1. Схема стеганографічної передачі даних

Задача *пасивного супротивника* полягає у визначенні факту наявності у контейнері прихованих даних. Припускається, що пасивний супротивник може перехопити усі відправлені контейнери і аналізувати їх як окремо, так і всі разом. У випадку, якщо пасивному супротивнику вдалось визначити наявність прихованого повідомлення у контейнері, він може спробувати витягти його з метою ознайомлення з його змістом. Оскільки, перед приховуванням повідомлення шифрується, то після витягання повідомлення потрібно буде дешифрувати.

Формальна модель стеганографічної системи

Для проведення оцінок комбінаторної складності задачі приховування інформації корисно мати формальну модель криптографічної системи. Пропонується описувати цю модель у такій формі.

Стеганографічною алгебраїчною системою S називають шістку параметрів

$$S = (H, I, M, D, D'),$$

де M – множина всіх повідомлень; K – множина всіх ключів; D – множина пустих контейнерів; D' – множина заповнених контейнерів; $H : M \times D \times K \to D'$ – відображення, яке перетворює пустий контейнер у заповнений контейнер; $I : D' \times K \to M$ – відображення, яке видобуває повідомлення з заповненого контейнера. Пару відображень (H, I) будемо називати *стеганографічним алгоритмом,* або *стеганографічним перетворенням.* Для вибраних ключа $k \in K$, повідомлення $m \in M$, пустого контейнера

$d \in D$, заповненого контейнера $d' \in D'$ стеганографічне перетворення має вид:

$$H(m,d,k)=d'$$
$$I(d',k)=m$$

Перетворення даних у JPEG формат

Відомо, що переважна більшість сучасних цифрових зображень зберігається у JPEG-форматі, який ґрунтується на косинусному перетворенні. Наведемо формули для одномірного випадку:

$$S[k]=\gamma(k)\sum_{n=0}^{N-1}s(n)\cos\left(\frac{\pi(2n+1)k}{2N}\right),k=\overline{1,N-1}$$

$$s(n)=\sum_{k=0}^{N-1}\gamma(k)S(k)\cos\left(\frac{\pi(2n+1)}{2N}k\right),n=\overline{0,N-1}$$

$$\gamma(k)=\begin{cases}\sqrt{\frac{1}{N}}, якщо\ k=0\\ \sqrt{\frac{2}{N}}, якщо\ k\neq 0.\end{cases}$$

JPEG – це формат стиснення зображень, тому виконати приховане вбудовування даних в зображення у ньому значно складніше. Оскільки косинусні коефіцієнти сильно зв'язані – малі зміни в коефіцієнтах суттєво впливають на вихідний результат і такі числові зміни достатньо просто виявити.

Щоб знайти нові підходи, пропонується розподілити зміну коефіцієнта на декілька сусідніх пікселів. Робити це потрібно так, щоб можна було декодувати приховані дані і внести мінімальні зміни в зображення. Таким інструментом можуть бути теоретико-чисельні перетворення (ТЧП).

ТЧП було розроблено як інструмент безпомилкових обчислень у цілих числах на основі модульної арифметики. Найбільш відомими варіантами побудови ТЧП є перетворення Гаусса, Ферма, Мерсена, Рейдера, а також на основі складених модулів. Загальний вигляд цих перетворень наступний:

$X(n)=\sum_{k=0}^{N-1}x(k)\alpha^{nk} \bmod F$ – пряме перетворення – вектор x у вектор X,

$x(k)=\frac{1}{N}\sum_{n=0}^{N-1}X(n)\alpha^{-nk} \bmod F$ – обернене перетворення – вектор X у вектор x.

тут α – *Первісний корінь:* $\alpha^N \equiv 1 \bmod F$.

Наведемо приклад перетворення Ферма для одномірного випадку.

Пряме і обернене перетворення для модуля F = 257, первісного кореня альфа α = 16 і довжині вектора, рівній N = 4, має наступні матриці:

матриця прямого перетворення

$$\begin{vmatrix}1 & 1 & 1 & 1\\ 1 & 16 & 256 & 241\\ 1 & 256 & 1 & 256\\ 1 & 241 & 256 & 16\end{vmatrix}$$

матриця оберненого перетворення

$$\begin{vmatrix}1 & 1 & 1 & 1\\ 1 & 241 & 256 & 16\\ 1 & 256 & 1 & 256\\ 1 & 16 & 256 & 241\end{vmatrix}$$

Крім параметрів перетворення або згладжуючих фільтрів також можна керувати параметрами оригінальних зображень, щоб досягти більш вигідних значень в околиці для приховування стеганографічних даних. Таким параметрами є квантування рівнів зображень, роздільної здатності, згладжування або підкреслення особливостей зображення.

Внесення змін при кодуванні у JPEG форматі

Оскільки пропонуються алгоритми, які ґрунтуються на кодуванні парних і непарних значень, проаналізуємо матрицю прямого перетворення, а також операції множення на вектор за модулем 2, що дозволить перейти від конкретних чисел до загальних математичних залежностей:

$$\begin{vmatrix}1 & 1 & 1 & 1\\ 1 & 16 & 256 & 241\\ 1 & 256 & 1 & 256\\ 1 & 241 & 256 & 16\end{vmatrix}_{\bmod 2}=\begin{vmatrix}1 & 1 & 1 & 1\\ 1 & 0 & 0 & 1\\ 1 & 0 & 1 & 0\\ 1 & 1 & 0 & 0\end{vmatrix}.$$

З точки зору передачі інформації чотири позиції, які кодуються парними і непарними числами дають чотири біти інформації – 2^4=16, тобто у десятковій системі дозволяють закодувати числа від 0 до 15.

Зазначимо, що отримана матриця перетворення має три рядки, лінійно незалежні за модулем 2. Це означає, що запропонований алгоритм не зможе однозначно закодувати всі чотири елемента вектора.

Знайдемо точну кількість елементів вектора, що може бути закодовано стеганографічним алгоритмом. Для цього представимо перетворення Ферма у символьному вигляді та зведемо у таблицю 2 парні і непарні позиції у вигляді нулів і одиниць.

x = (a+b+c+d) mod 2	x = (a+b+c+d) mod 2
y = (a+16b+256c+241d) mod 2	y = (a+d) mod 2
z = (a+256b+c+256d) mod 2	z = (a+c) mod 2
t = (a+241b+256c+16d) mod 2	t = (a+b) mod 2

Видно, що результат має обмежену кількість варіантів (половина значень) і симетричність відносно восьмого значення. Для 256 чисел у чотирьох позиціях результат буде повторюватись з періодом (циклічністю) 16. З табл. 1 також зрозуміло, що у чотирьох позиціях вектора можна закодувати три стегобіти. Покажемо, що зменшення кількості вихідних значень зберігається і для загального випадку.

Таблиця 1

Повний перебір вхідних і вихідних векторів у вигляді парних і непарних значень (нулів і одиниць) для стеганографічного алгоритму

a	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
b	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
c	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
d	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
x	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0
y	0	1	0	1	0	1	0	1	1	0	1	0	1	0	1	0
z	0	0	1	1	0	0	1	1	1	1	0	0	1	1	0	0
t	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0
	0	**12**	**10**	**6**	**9**	**5**	**3**	**15**	**15**	**3**	**5**	**9**	**6**	**10**	**12**	**0**

Нехай x – деяке число, що дорівнює нулю або одиниці, а число x’ обернене до x за модулем 2. Покажемо, що x’= x+1 mod 2. Для цього розглянемо табл. 2.

Таблиця 2

Правила виконання арифметичних операцій за модулем 2

x'	x	1	1+x
0	1	1	0
1	0	1	1

Перейдемо до розгляду векторів. Позначимо **y**=(y1, y2, y3, y4)^T , а **y’** – обернений за модулем 2 вектор до **y**. Для того, щоб отримати залежність між прямим і оберненим за модулем 2 вектором розглянемо добуток за модулем 2 матриці перетворення **T** на одиничний вектор:

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{vmatrix} \times \begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \end{vmatrix} \bmod 2 = \begin{vmatrix} 0 \\ 0 \\ 0 \\ 0 \end{vmatrix}.$$

Оскільки, добуток за модулем 2 матриці перетворення на одиничний вектор дорівнює нульовому вектору, тоді вірно і наступне:

$$\mathbf{T}\cdot\mathbf{y'} \bmod 2 = \mathbf{T}\cdot(\mathbf{1} + \mathbf{y}) \bmod 2 = (\mathbf{T}\cdot\mathbf{1} + \mathbf{T}\cdot\mathbf{y}) \bmod 2 = \mathbf{T}\cdot\mathbf{y}.$$

Отже, за модулем 2 добуток матриці перетворення на вектор дорівнює добутку матриці на обернений вектор за модулем 2. Оскільки, в запропонованому стеганографічному алгоритмі кодування виконується з числами, отриманими після перетворення ТЧП, число і обернене цьому числу за модулем будуть кодувати одне вихідне значення. Таким чином, запропонований універсальний стеганографічний алгоритм має вдвічі меншу кількість варіантів для перебору і симетричність варіантів вибору чисел, але дозволяє приховати вдвічі менший об’єм даних у порівнянні з НЗБ стеганографічним алгоритмом. Для розглянутого прикладу кількість варіантів, яку потрібно перебрати для кодування одного значення дорівнює, (256/32) = 8, оскільки вибираємо один з 16 варіантів (табл. 1) і кожне пряме і обернене за модулем 2 число кодує однакове значення.

Максимальна інформаційна ємність прихованих даних I_H для запропонованого алгоритму розраховується за формулою:

$$I_{H\,макс} = 3xS/4 = 3x(MxN)/4,$$

де S – площа зображення, M – ширина зображення, N – довжина зображення. Коефіцієнт ? означає, що

вектор з чотирьох елементів кодує три стегобіти прихованого повідомлення.

Концепція стеганографічного алгоритму «згладжування – перетворення – балансування – вбудовування»

JPEG формат забезпечує найменший розмір файлу при найбільшому розрізненні для типових фотографій. Оскільки, переважна більшість зображень зберігається у JPEG форматі, тому розробку нових стеганографічних методів слід зосередити на цьому виді файлів, тому що використання інших видів файлів для передачі звичайних фотографій буде одразу викликати підозри у аналітика. Розробка стеганографічних методів для JPEG формату пов'язана з низкою труднощів. Формат передбачає виконання послідовних дій для стиску (зменшення психовізуальної і статистичної надлишковості) файлів: перетворення кольорового простору та зменшення кількості біт, які виділяють на кольори, виконання косинусного перетворення. В результаті дані зображення стають локально монотонними та функціонально зв'язаними. Відповідно проста модифікація найменш значимих біт або застосування похідних алгоритмів порівняно легко виявляється, оскільки додає ентропію у дані спектрально впорядкованих компонент, що характерно навіть для косинусних коефіцієнтів, які відповідають дрібним деталям сцени.

Пропонується вбудовувати дані в околицю 8x8 зображення у форматі JPEG, який базується на косинусному перетворенню. Спочатку над початковим зображенням – в області оригіналів – треба виконати перетворення, які збільшують нерегулярність (негладкість) образу, але не викликають візуальних спотворень для спостерігача. Наприклад, показники РС-аналізу, χ^2, дисперсії шуму або інші операції стандартні для зображень, які обробляють методами підготовки до друку. Завдяки цьому буде підвищено інтенсивність високочастотних складових спектру косинусного перетворення, що відповідають дрібним деталям.

Далі над зображеннями виконуються операцій дискретизації, ре-квантування і т.п., які також підвищують інтенсивність високочастотних компонент косинусного перетворення. Потім здійснюється підбір параметрів теоретико-чисельних перетворень (ТЧП) за критерієм найменшого відхилення JPEG-коефіцієнтів від початкових значень (рис. 2). Метою таких перетворень є отримання файлу зображення у JPEG форматі з високим ступенем природної для цього формату стиснення нерегулярності.

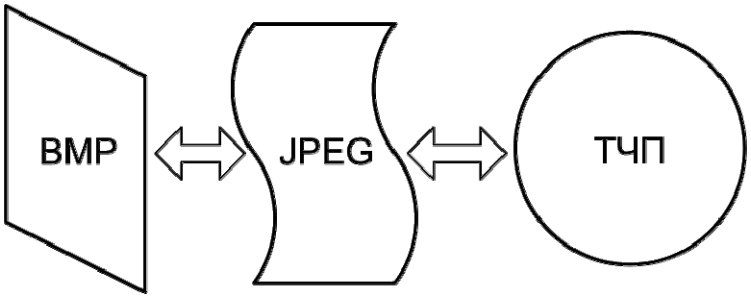

Рис. 2. Схематичне представлення процесів стеганографічного приховування з використанням ТЧП

Аналогічну схему роботи алгоритму можна запропонувати і для відео потоку (рис. 3)

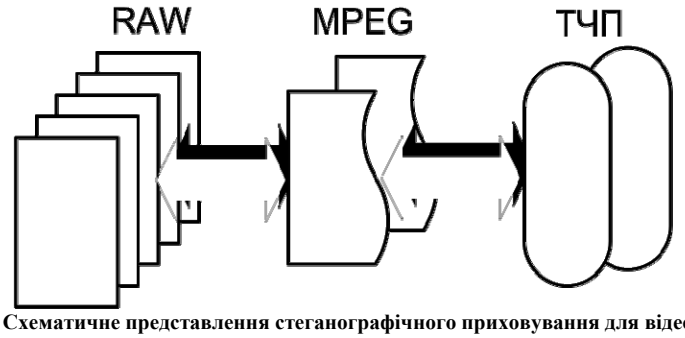

Рис. 3. Схематичне представлення стеганографічного приховування для відеоданих

Схема алгоритмів стеганографічного приховування інформації для JPEG контейнерів на базі ТЧП

На базі описаного ТЧП можна створювати алгоритми мтеганографічного приховування інформації. Загальна схема таких алгоритмів містить такі кроки.

1. Перед початком роботи відправник і одержувач виконують обмін ключами. Стеганографічний алгоритм і його параметри їм обом відомі.

2. Відправник виконує перебір парних і непарних значень, що відповідають значенням стегобітів в області ТЧП, щоб закодувати приховане повідомлення.

3. Одержувачу надсилається файл зображення з закодованими значеннями. Значення зчитуються, парні і непарні числа декодуються згідно до стеганографічного алгоритму у біти зі значеннями нуль та одиниця.

Схема конкретного алгоритму залежить, в основному, від того, яким чином деталізовано пункт 2

цієї загальної схеми. Для конкретного алгоритму треба вказати, зокрема такі параметри: тип ТЧП, значення модуля F та довжину вектора, спосіб вибору даних із зображення для вбудовування стеганографічної інформації, спосіб квантування рівнів зображення у JPEG-форматі. Найбільш відомими варіантами побудови ТЧП, або типами ТЧП, є перетворення Гаусса, Ферма, Мерсена, Рейдера, а також на основі складених модулів.

Найпростіший алгоритм стеганографічного кодування на базі ТЧП використовує вектори довжиною чотири, що є мінімальною розмірністю для більшості перетворень для яких є швидкий алгоритм. За побудовою ТЧП перетворення накладають жорсткий зв'язок на довжину вектора, максимальне значення чисел, над якими виконуються операції, та значення первісного кореня. Після виконання ТЧП у відповідності до значень бітів і кількості задіяних у приховуванні даних елементів вектора виконується підбір значень таким чином, щоб задовольнити вимогам парності та непарності, які відповідають нульовим і одиничним бітам. Вибір значень конкретного вектора обумовлюється близькістю значень (мінімумом відхилення) початкового вектора і вектора з прихованими даними. Проблема того, що модуль дорівнює 255 або 257, а максимальне значення байта дорівнює 255 вирішується відстежуванням максимальних значень у алгоритмі. Зрозуміло, що можна використати і більші значення модулів ТЧП, але це призведе лише до збільшення кількості арифметичних і логічних операцій при підборі значень, оскільки значення більше за 255 не дозволяють відображати більшість широко поширених форматів графічних файлів. В якості міри близькості векторів можуть виступати різні метрики та норми в залежності від типу зображення. Найпростішим варіантом є максимум співпадінь значень елементів векторів.

Кількість варіантів для впорядкованого вибору чотирьох довільних чисел від 0 до 255 складає 256^4 = 4 294 967 296 варіантів. Проаналізуємо математичні перетворення з метою зменшення комбінаторної складності алгоритму стеганографічного приховування.

Кодуючі фільтри (згладжуючі)

При внесенні змін навіть у найменші біти зображення при косинусному перетворенні, при оберненому перетворенні для візуалізації, зображення може мати вади, які візуально легко помітити.

Тому пропонується виконувати згладжування, що відбувається в області косинусних коефіцієнтів JPEG-перетворення; параметр згладжування можна витягти з околу пороговим методом (однозначна відповідність для вбудованої інформації). Найпростішим варіантом алгоритму є сума значень в околі, значення якої фіксується як парне або непарне, або кратне якомусь числу; або його значення перевищує поріг середнього значення (одиниця) або менше за середнє значення (нуль).

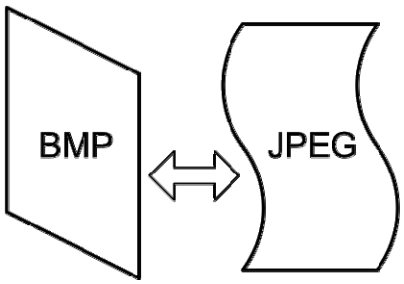

Рис. 4. Схематичне представлення ітеративних процесів стеганографічного приховування

Нехай C – оператор кодування, що перетворює значення пікселів контейнера в околі O за алгоритмом таким чином, що вони відповідали нульовим або одиничним бітам прихованого повідомлення:

$$C_1 O_i \xrightarrow{1} O_i^1 \;\; C_0 O_j \xrightarrow{0} O_j^0,$$

Позначимо D_{-1} оператор декодування, який за пороговим методом виймає з пікселів контейнера в околиці O стегобіти прихованого повідомлення I_H:

$$I_H = \begin{cases} 1, & D_{-1} O_k \geq P \\ 0, & D_{-1} O_k < P \end{cases}.$$

Прикладом кодуючих операторів можуть бути медіанні або згладжуючи фільтри на основі локально адаптивних алгоритмів цифрової фільтрації зображень.

Методи стеганографії вносять зміни у вхідні розподіли параметрів оригіналів для того, щоб приховати вбудовані дані. Оскільки, стегоінформація є дискретною, то внесені зміни полягають у зміні обраних параметрів даних зображень, які дозволяють витягти початкові дані. Простіше кажучи, відбувається введення порогу, за яким виконується витягання інформації.

Методи стеганографічного аналізу (СА, стегоаналізу) ґрунтуються на виявленні аномальних збурень у локальних параметрах частин зображень. На основі математичних розрахунків у виявлених даних виконується оцінка їх інформаційної ємності, щоб виміряти розмір прихованих повідомлень. При цьому, чим більше в зображенні вбудованих даних, тим сильніше буде спотворюється природні статистичні характеристики зображення і зростає імовірність їх правильного виявлення (прихованих даних) стегоаналітичними методами. На якість роботи методів стегоаналізу впливає зміст зображення (фон,

текстура, кількість стегобітів на одиницю площі і розташування дрібних деталей), а також умови його реєстрації, артефакти стиснення даних, шуми CCD-матриці та застосовані методи покращення візуальної якості.

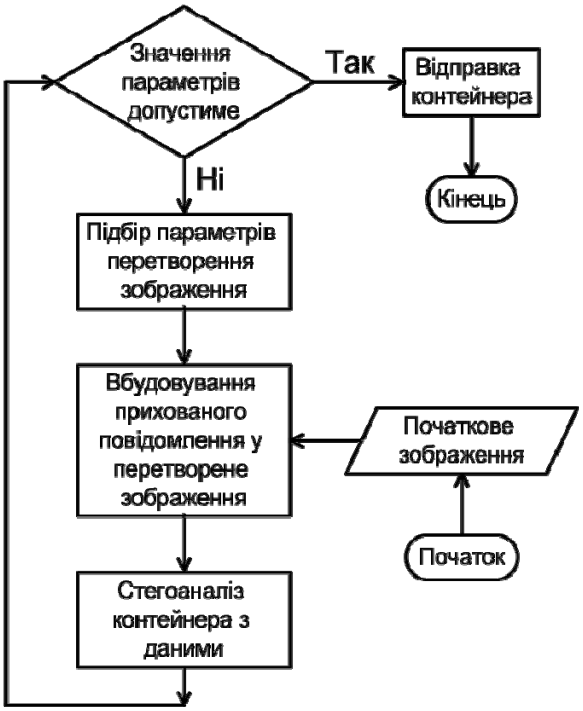

Рис. 5. Блок-схема ітеративного процесу стеганографічного приховування інформації

У роботі розглядається концепція керованої зміни статичних характеристик зображення на різних етапах його обробки та візуалізації, а також кодування стеганографічних даних у параметрах алгоритмів локально згладжуючих фільтрів. Тобто приховування даних буде здійснюватись у параметрах фільтрації, які можна витягти за значенням порогу в околиці.

У роботі пропонується перейти від прямого вбудовування в дані зображення та намагання підлаштуватись до природної статистики зображення до керованого згладжування (створення керованої статистики). Тоді приховування даних буде здійснюватись у параметрах фільтрації, які можна витягти за значенням порогу в околиці та до зміни природної статистики у межах, які є характерними для зображень в мережі Інтернет.

Вище викладене можна узагальнити наступним чином – обирається околиця і значення в околиці добираються таким чином, щоб вони були згладжені у спосіб, що дозволив би приховати в околиці дані.

Планування чисельних експериментів і інтелектуальний аналіз даних для задачі стеганографії

З точки зору сучасної комп'ютерної техніки та інформаційних технологій обробки даних частина обчислювальних процесів стеганографічних досліджень будуються на основі хмарних технологій. Більшість розробників сучасних комерційно успішних програмних продуктів передбачають їх роботу на платформах віртуалізації для найповнішого використання комп'ютерних ресурсів, розподілених по територіально віддалених обчислювальних вузлах. При цьому вибір моделі приватної хмари, публічної або гібридної грунтується на економічних показниках, головним з яких є зменшення вартості володіння обчислювальною інфраструктурою. Аналіз масивів даних, створених алгоритмами приховування даних, можна віднести до описового типу задач аналізу даних, метою яких є поглиблення розуміння задачі стеганографії і стегоаналізу. Серед підзадач стеганографії, розв'язок яких можна покращити методами інтелектуального аналізу даних (ІАД) виділимо наступні:

1) Уточнена класифікація зображень на основі набору статистичних параметрів, семантичних ознак, що описують сцену, об'єкти та EXIF-теги – параметри фотографії. На основі розширеної класифікації можна сформувати набори характеристик, які впливають на результати стегоаналізу.

2) Кластеризація. Виявлення незалежних груп характеристик і параметрів, зібраних на етапі класифікації.

3) Побудова асоціативних правил – виявлення і опис залежностей між елементами у кластерах.

4) Відновлення функціональних залежностей та їх формалізація у вигляді формул, таблиць значень методами математичної статистики.

Розглянемо послідовність обробки інформації для стеганографічних досліджень за функціональною схемою на рис. 5. Задачею інтелектуальної обробки даних є отримання знань з множини даних, що грунтується на перетворенні неструктурованої інформації у структуровану. Процес обробки даних стеганографічних досліджень будується на основі технологій зберігання даних, проте має свої особливості. Накопичені масиви фотографій з джерел даних (ДД) після первинної обробки передаються у сховище даних,

де кожній фотографії ставиться у відповідність набір полів, які описують налаштування фотоапарату (EXIF-теги зображень), опис сцени, призначення колекції, походження і т.п. метадані контейнерів. Результатом обробки даних у сховищі є отримання з неструктурованої інформації слабо структурованої.

Наступним етапом є аналітична обробка даних, тобто отримання статистичних характеристик окремих зображень і їх масивів. Такими характеристиками є середні значення, медіани, дисперсії, оцінки шумів та інші статистики першого та другого порядків, результати різних методів стегоаналізу, а також застосування стандартних цифрових фільтрів та дизайнерських операцій візуалізації зображень і підготовки до друку та демонстрації у мережі Інтернет. У відомих системах підтримки прийняття рішень на цьому етапі часто ставляться вимоги обробки даних в реальному масштабі часу, проте для стеганографії такі системи наразі широко невідомі. На основі отриманої інформації виконується очищення зображень від даних з характеристиками (виділення в особливу групу), які сильно відхиляються від типових значень. Результатом даного етапу є реляційна база даних, на основі якої отримують структуровану інформацію (звіти, графіки, гістограми, статистики).

Розшифровка і пояснення позначень структурної схеми. Джерела даних (ДД) – тематичні колекції фотографій, масиви аматорських фотографій з мережі Інтернет. Сховище даних (СД) – первинна обробка даних (відбраковування, агрегація), створення зв'язаних описів фотографій (сцен, умов зйомки, налаштувань фотоапаратів). Неструктурована інформація – масиви фотографій без групування за темою. Слабо структурована інформація – до загального опису зображення додають результати чисельних експериментів з зображеннями (застосування фільтрів, додавання різної кількості прихованих даних у контейнер і т.п.). Структурована інформація – звіти (відповіді на запити), статистичні гістограми, графіки, виявлені закономірності (логічні правила, що дозволяють передбачати результати обробки).

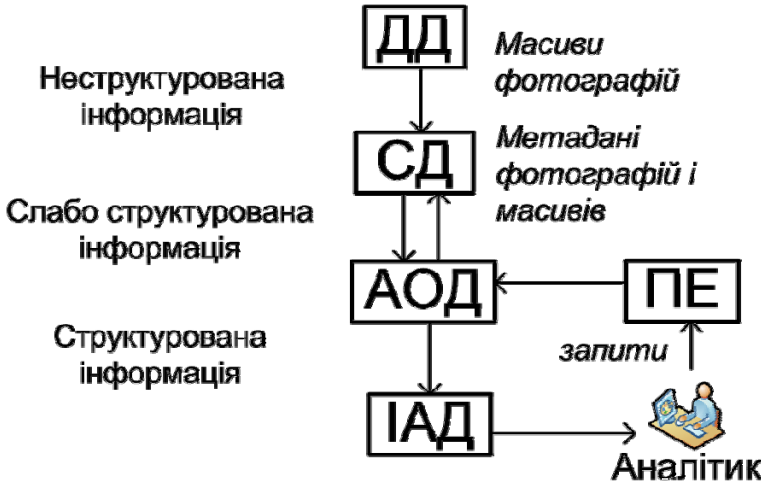

Рис. 6. Схема процесу обробки стеганографічних даних

Планування чисельних експериментів (ПЕ) – на відміну від інтелектуального аналізу даних або послідовних чисельних експериментів у стеганографічні дослідження вимагають поєднання обох підходів. Спочатку виконуються експерименти (досліди) і в залежності від результатів інтелектуального аналізу даних аналітиком (дослідником) формулюються нові гіпотези і запити на обробку (дослідження, аналіз даних).

Отримана інформація надходить до програмних додатків інтелектуального аналізу даних, де виконується більш глибокий пошук закономірностей на основі побудованих аналітиком та стандартизованих розробником програмних комплексів моделей даних. Результатом роботи є логічна модель, яка на основі співвідношень характеристик дозволяє виявляти приховані дані з більшою імовірністю. Для стеганографічних досліджень важливим є визначення як статистичні характеристики зображень змінюються при різному об'ємі вбудованих у контейнер даних для різних стеганографічних алгоритмів. На основі стеганографічних чисельних експериментів отримують порогові значення для діапазонів параметрів, які є визначальними для логічних моделей інтелектуального аналізу даних. Як було показано у попередніх публікаціях [2–4] оптимальні послідовності дій дозволяє побудувати теорія планування експериментів.

Планування чисельних експериментів

Планування експериментів не може покращити створену модель даних, проте може покращити статистичні характеристики моделі. Відомо, що при виконанні складних досліджень велика частина факторів залишається недоступною для безпосереднього спостереження. Рандомізація даних дозволяє зменшити систематичні помилки. Відсутність рандомізації при неоднорідних даних призводить до зміщення оцінок і впливає на якість вихідного результату.

Початкові етапи ПЕ для стеганографічних досліджень було розглянуто у попередніх роботах [2–4], де було описано відсіювання, рандомізацію (зменшення впливу початкового зміщення даних),

відбраковування зображень та створення тематичних груп.

Оптимізація ПЕ можлива як для дисперсій (найбільш розроблені теоретичні плани), так і для кількості чисельних експериментів. Тому вибір схеми плану залежить від мети аналітика та моделі, що побудована або перевіряється. Оскільки, для чисельних експериментів немає багатьох обмежень, які характерні для вимірювання фізичних величин або параметрів хімічних реакцій, то для них можна побудувати настільки оптимальний план, наскільки дозволяє досліджувана модель.

ВИСНОВКИ

В рамках продовження досліджень задачі приховування інформації в зображеннях JPEG-формату одержано результати в таких напрямках:

- Запропоновано загальну схему алгоритмів стеганографічного приховування даних на базі теоретико-чисельних перетворень.
- Розроблено нову концепцію стеганографічних алгоритмів на основі кодуючих перетворень контейнерів для приховування інформації.
- Побудовано загальну структуру інтелектуального аналізу даних для стеганографічних досліджень на основі перетворення неструктурованої інформації у структуровану. Для одержання даних запропоновано план проведення чисельних експериментів.

Література

1. Стеганография, цифровые водяные знаки и стеганоанализ / [А.В. Аргановский, А.В. Балакин, В.Г. Грибунин, С.А. Сапожников]. – М. : Вузовская книга, 2009. – 220 с.

2. Корольов В.Ю. Визначення можливостей RS-стегоаналізу для дослідження статистичних властивостей зображень / В.В. Поліновський, В.А. Герасименко // Вісник Хмельницького національного університету. – 2010. – № 4. – С. 102–110.

3. Королёв В.Ю. Стеганография по методу наименее значимого бита на базе персонализированных флеш-накопителей / В.В. Полиновский, В.А. Герасименко // Управляющие системы и машины. – № 1 (231). – 2011. – С. 79–87.

4. Корольов В.Ю. Планування досліджень методів стеганографії і стегоаналізу / В.В. Поліновський, В.А. Герасименко, М.Л. Горінштейн // Вісник Хмельницького національного університету. – 2011. – № 4. – С. 187–195.

References

1. Arganovskiy A. V., Balakin A. V., Gribunin V. G., Sapozhnikov S. A. Steganografiya, tsifrovyye vodyanyye znaki i steganoanaliz. Moscow. Vuzovskaya kniga, – 2009. – 220 p. [in Russian]

2. Korolyov V.Yu., Polinovskyi V.V., Gerasimenko V.A. Viznachennya mozhlivostey RS-stegoanalizu dlya doslidzhennya statistichnikh vlastivostey zobrazhen // Visnyk Khmelnytskoho natsionalnoho universytetu. Technical science. Khmelnitsky. – 2010. – № 4. – pp. 102 – 110. [in Ukrainian]

3. Korolyov V.Yu., Polinovskyi V.V., Gerasimenko V.A. Steganografiya po metodu naimeneye znachimogo bita na baze personalizirovannykh flesh-nakopiteley // Upravlyayushchiye sistemy i mashiny. – 2011. – № 1. – pp. 79 – 87. [in Russian]

4. Korolyov V.Yu., Polinovskyi V.V., Gerasimenko V.A., Gorinshteyn M.L. Planuvannya doslidzhen' metodiv steganografii i stegoanalizu // Visnyk Khmelnytskoho natsionalnoho universytetu. Technical science. Khmelnitsky. – 2011. – № 4. – pp. 187 – 195. [in Ukrainian]

Рецензія/Peer review : 14.1.2014 р. Надрукована/Printed :6.2.2014 р.
Рецензент: д.т.н., проф., зав. від. Інституту кібернетики ім. В.М. Глушкова НАН України Гуляницький Л.Ф.