НОВІ ЗАСОБИ
КІБЕРНЕТИКИ, ІНФОРМАТИКИ,
ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ
ТА СИСТЕМНОГО АНАЛІЗУ

Н.В. КОШКІНА

УДК 004.056; 004.415.24

СТЕГАНОАНАЛІЗ J-UNIWARD

Анотація. Розглянуто проблему виявлення адаптивної стеганографії за методом J-UNIWARD стеганоаналітичними системами на базі машинного навчання. За допомогою порівняльного аналізу точності визначено, що найбільш чутливими до J-UNIWARD є статистичні моделі побудови характеристичних векторів, що формуються у просторовій зоні, — GFR, PHARM та DCTR. Запропоновано два способи підвищення точності стеганоаналізу з використанням цих моделей: аналіз найбільш імовірних місцеположень вкраплення; зважене голосування за трьома моделями. Показано, що без попередньої класифікації зображень згідно з їхніми параметрами точність стеганоаналізу суттєво знижується. Отримані результати можна використовувати для побудови ефективних систем стеганоаналізу зображень у форматі JPEG.

Ключові слова: інформаційна безпека, стеганографія, J-UNIWARD, стеганоаналіз, методи з навчанням та класифікацією, точність детектування.

ВСТУП

Стеганографія та стеганоаналіз взаємопов'язані — нові результати в одній із дисциплін стимулюють дослідників до покращення досягнень в іншій. На початку розвитку стеганографії було достатньо візуальної непомітності прихованих повідомлень, на сьогодні стеганографічне втручання зазвичай відслідковують за зміною статистичних показників. Чим складніше втручання, тим більше для його виявлення відстежують взаємозв'язків, задіюють різних показників та враховують особливостей. Проте стеганоаналіз як дисципліна, виникнення якої було спричинене наявністю стеганографії, загалом розвивається із певним відставанням. Постає питання: наскільки безпечним наразі є використання найсучасніших методів комп'ютерної стеганографії? Дослідимо цю проблему на прикладі стеганоаналізу методу J-UNIWARD (UNIWARD у застосуванні до JPEG-контейнерів).

ОПИС МЕТОДУ UNIWARD

UNIWARD є одним з нових контент-адаптивних методів комп'ютерної стеганографії, який запропоновано у 2013 р. [1]. Його можна використовувати як для просторового, так і частотного представлення зображення. Основна ідея UNIWARD полягає у прихованні даних у текстурованих і зашумлених ділянках (які складно змоделювати у будь-якому напрямку), оминаючи при цьому гладкі зони та чіткі краї. Для визначення придатної ділянки вводять універсальну функцію спотворення, що розраховується у вейвлет-зоні (звідки походить назва методу — Universal Wavelet Relative Distortion).

Зображення X розмірами $n_1 \times n_2$ фільтрується набором напрямлених лінійних інваріантних до зсуву фільтрів $B = \{K^{(1)}, K^{(2)}, K^{(3)}\}$, що використовують для оцінювання його гладкості у горизонтальному, вертикальному та діагональному напрямках, і обчислюють напрямлені залишки $W^{(k)} = K^{(k)} * X$, де $*$ — дзеркальна згортка. У загальному випадку фільтри можуть бути будь-які, але автори методу використовують вейвлет-декомпозицію на базі вейвлета Добеші

8-го порядку. Якщо це JPEG, то зображення спочатку розпаковується у просторову зону, а потім фільтрується. Щоб розрахувати універсальну функцію спотворення, вказані операції здійснюють для пари X — пустий контейнер та Y — той самий контейнер, але з прихованим повідомленням. Сама функція є сумою відносних змін усіх вейвлет-коефіцієнтів після стеганоперетворення зображення:

$$D(X,Y)=\sum_{k=1}^{3}\sum_{u=1}^{n_1}\sum_{v=1}^{n_2}\frac{|W_{uv}^{(k)}(X)-W_{uv}^{(k)}(Y)|}{\sigma+|W_{uv}^{(k)}(X)|}, \qquad (1)$$

де $\sigma>0$ — стабілізаційна константа, яка запобігає діленню на нуль.

Відношення (1) є найменшим у разі зміни великого за значенням вейвлет-коефіцієнта зображення, тобто в зонах, де з'являються текстура та краї. Водночас приховувати дані не рекомендують в ділянках, де значення $|W_{uv}^{(k)}(X)|$ є малим хоча б для одного з трьох k, бо це відповідає напрямку, вздовж якого контент може бути змодельованим (що допоможе у детектуванні стегановкладки).

Зауважимо, що універсальна функція спотворення є неадитивною. Зміна пікселя в просторовій зоні вплине на окіл $s\times s$ вейвлет-коефіцієнтів, де s — кількість вагових коефіцієнтів фільтра. Зокрема, для вейвлета Добеші 8-го порядку маємо вплив на 16×16 коефіцієнтів. Для JPEG-стеганографії зміна коефіцієнта дискретного косинусного перетворення (ДКП) вплине на блок 8×8 пікселів та відповідно на $(8+s-1)\times(8+s-1)$ вейвлет-коефіцієнтів. У разі зміни сусідніх пікселів чи ДКП коефіцієнтів околи впливу перетинаються, тобто зміни у вейвлет-субсмугах будуть взаємодіяти.

Для вибору ділянок приховування використовують адитивну апроксимацію описаної функції спотворення. Оцінюють вартість спотворення кожного пікселя чи ДКП коефіцієнта зображення

$$\rho_{ij}(X,Y_{ij})=D(X,X_{\sim ij}Y_{ij}),$$

де $X_{\sim ij}Y_{ij}$ — зображення зі зміненим ij-м елементом.

Загалом, адитивна апроксимація має вигляд

$$D_A(X,Y)=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\rho_{ij}(X,Y_{ij})[X_{ij}\neq Y_{ij}].$$

Отож для методу UNIWARD передбачено, що елементи зображення використовуються як носії бітів повідомлення не послідовно чи, наприклад, за псевдовипадковим розподілом, а залежно від вартості їхнього спотворення. На рис. 1 показано, як саме розширюється ділянка приховування у разі збільшення довжини повідомлення.

Для зменшення спотворення у процедуру вкраплення можна долучити методи завадостійкого кодування. Зокрема, в роботі [1] йдеться про застосування синдромних гратчастих кодів (Syndrome Trellis Codes).

Нехай у контейнері-зображенні X визначено послідовність бітів-носіїв $x\in\{0,1\}^n$. Процедура вкраплення є двійковою, тобто в результаті вкраплення деякого повідомлення $m\in\{0,1\}^p$ послідовність x замінюється на $y\in\{0,1\}^n$. У разі застосування синдромного кодування вкраплення та вилучення повідомлення реалізують з використанням лінійного двійкового коду C

$$Embedding(x,m)=\arg\min_{y\in C(m)}D(x,y),\ \ Extraction(y)=Hy,$$

де всі операції є двійковими, $H\in\{0,1\}^{p\times n}$ — перевірна матриця коду C, $C(m)=\{z\in\{0,1\}^n\,|\,Hz=m\}$ — суміжний клас, що відповідає синдрому m.

У роботі [2] як перевірну запропоновано використовувати матрицю контролю парності спеціального вигляду, що дає змогу представити кожний розв'язок $Hy=m$ як шлях через гратку. Оптимальний y, найближчий до x, визначають за алгоритмом Вітербі [3].

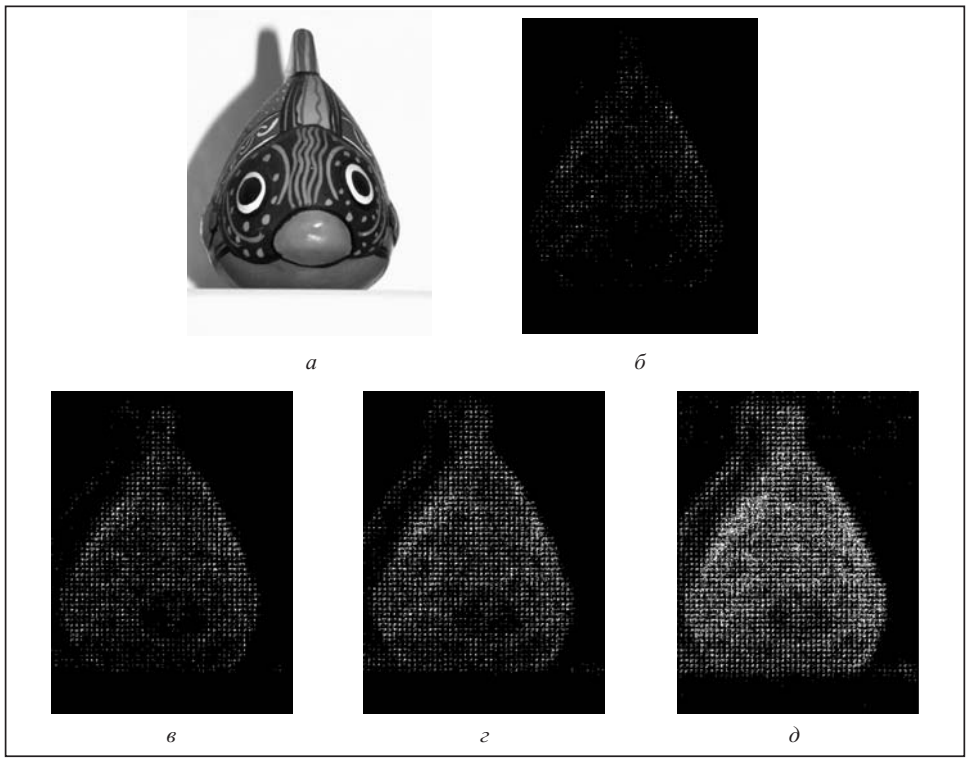

Рис. 1. Динаміка зміни ділянки приховування за умови збільшення стеганонавантаження: *а* — вихідне зображення; *б–д* — місцеположення змінених бітів у разі приховування 1, 3, 5 та 10 КБ даних відповідно

Отже, UNIWARD вже за побудовою є безпечнішим за стенографію на основі найменшого значущого біта (НЗБ), що не враховує вміст зображення та не мінімізує кількість змінених бітів. Оскільки стеганографія може бути використана для вчинення протиправних дій, стеганоаналіз цього методу є актуальною та важливою проблемою інформаційної безпеки.

ОПИС СТЕГАНОАНАЛІЗУ ЗА МЕТОДАМИ З НАВЧАННЯМ ТА КЛАСИФІКАЦІЄЮ

Останнім часом значну увагу дослідників привернули стеганоаналітичні методи з навчанням і класифікацією, оскільки вони є універсальними та можуть покращуватися застосуванням новітніх здобутків теорії машинного навчання. Такі методи кожному об'єкту (контейнеру) ставлять у відповідність його характеристичний вектор, що чутливо реагує на стеганоперетворення і водночас не є залежним від вмісту контейнера. Цей вектор визначається набором різних статистичних показників досліджуваного зображення. Також він може доповнюватися статистикою каліброваної версії контейнера, де процедура калібрування полягає у потенційному знищенні прихованої інформації за допомогою незначних геометричних перетворень, що майже не впливають на статистичні показники.

За останні роки побудовано багато різних моделей характеристичних векторів. Оскільки тут розглядаємо стеганоаналіз J-UNIWARD, звернемо увагу на вісім наявних моделей, що можуть бути застосованими до зображень у форматі JPEG. Перша модель — це CHEN [4], яка використовує різницеві матриці та обчислені на їхній основі матриці ймовірностей переходу для фіксації внутрішньоблокових і міжблокових порушень статистики ДКП коефіцієнтів зображень. Друга — покращена декартовим калібруванням [5] версія попередньої моделі, що позначена авторами як CC-CHEN. Третя модель — це LIU, запропонована у [6], яка базується на тому, що стеганографічні приховування змінюють спільну щільність сусідніх елементів. Наступна — CC-PEV [7], яка включає значення глобальної гістограми розподілу ДКП коефіцієнтів, локальних гістограм розподілу значень перших п'яти AC (Alternating Current)-коефіцієнтів усереднe-

ної варіації та інші статистичні показники і доповнена декартовим калібруванням. Крім того, розглянемо модель CC-C300 [8], що ґрунтується на використанні векторів великої розмірності, елементи яких складають значення матриць спільної появи пар ДКП коефіцієнтів, та модель GFR [9], де характеристичні вектори побудовані як гістограми квантованих залишків, отриманих з використанням двовимірних фільтрів Габора з різними масштабами та орієнтаціями. Також до JPEG-зображень застосовна модель векторів із ДКП залишків DCTR [10], яка враховує фази; елементи характеристичних векторів у цій моделі формуються з гістограм залишків, розрахованих з використанням базових шаблонів ДКП. Восьма модель — PHARM [11], що базується на залишках, як дві попередні, але вони представлені з використанням статистики їхніх випадкових проєкцій першого порядку.

Стеганодетектором є класифікатор, на вхід якого подаються характеристичні вектори, обчислені за певною моделлю. Він формується за допомогою контрольованого навчання на контейнерах, для яких відома мітка класу — «пустий» чи «заповнений». Оскільки розмірності згаданих статистичних моделей досить великі (за зростанням: LIU — 216, CHEN — 486, CC-PEV — 548, CC-CHEN — 972, DCTR — 8000, PHARM — 12600, GFR — 17000, CC-C300 — 48600), найбільш універсальним вибором для процесів навчання та детектування є ансамблевий класифікатор, який добре масштабується за розмірністю характеристичного вектора і кількістю контейнерів у навчальній вибірці. Ансамблевий класифікатор працює за схемою, відображеною на рис. 2.

Цей класифікатор з d_{all} елементів характеристичного вектора, обчисленого за певною моделлю, формує L випадково вибраних підмножин, кожна з яких має розмірність $d_{\text{sub}} < d_{\text{all}}$. Використовуючи сформовані підмножини, він навчає L базових класифікаторів розрізняти пусті зображення та стеганоконтейнери. Нехай далі $N_v(z)$ — кількість елементів ансамблю (базових класифікаторів), що голосують за належність зображення z до класу пустих. Фінальне рішення про мітку класу цього зображення визначають за таким правилом:

$$Rule(L, N_v) = \begin{cases} -1, & \text{якщо } N_v > L/2, \\ +1, & \text{якщо } N_v < L/2, \\ random\{-1, +1\} & \text{інакше.} \end{cases}$$

Зауважимо, що існують різні варіанти вибору базових класифікаторів ансамблю. За рекомендаціями авторів статті [12], де запропоновано використання ансамблевого класифікатора в стеганоаналітичних системах, зупинимося на лінійному дискримінанті Фішера (ЛДФ) з огляду на його швидке навчання та гарні результати, отримані під час розв'язання задач стеганоаналізу.

Далі для проведення чисельних експериментів використовувався пакет MatlabR2019, Matlab-реалізація описаних моделей характеристичних векторів та ансамблевого класифікатора на базі ЛДФ, що є у вільному доступі на ресурсі http://dde.binghamton.edu/. Були задіяні налаштування параметрів класифікатора за умовчанням та процедури автоматичного пошуку значень d_{sub} і L, оптимальних у сенсі мінімізації помилок детектування за прийнятний час розрахунків.

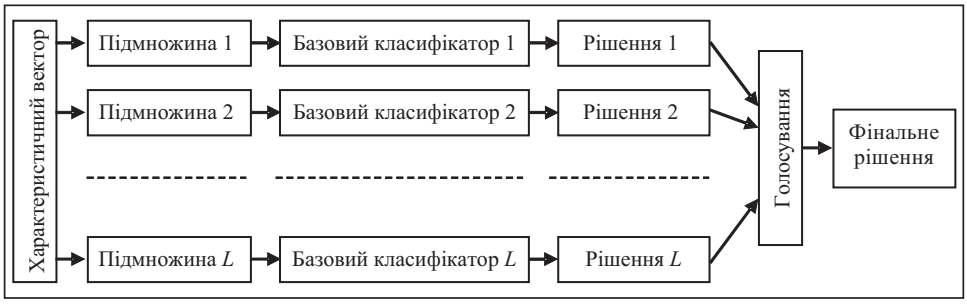

Рис. 2. Структурна схема функціонування ансамблевого класифікатора

ПОРІВНЯННЯ ТОЧНОСТІ КЛАСИФІКАЦІЇ ЗА РІЗНИМИ МОДЕЛЯМИ

Зважаючи на останні тенденції (коли під час проєктування стеганографічних систем з огляду на їхню безпеку дослідники обмежують корисне навантаження приблизно половиною придатних коефіцієнтів), будемо розглядати метод J-UNIWARD з діапазоном корисних навантажень від 50 до 5 %. Спочатку було порівняно точність усіх вказаних моделей у разі виявлення стеганоконтейнерів з 50 та 10 % корисним навантаженням, тобто коли для вкраплення використовували половину ненульових АС-коефіцієнтів ДКП та десяту частину відповідно.

Як тестовий набір вибрано 1114 кольорових JPEG-зображень розмірами 384×512 пікселів, стиснених з коефіцієнтом якості 75. За допомогою генератора випадкових чисел поділили набори пустих та заповнених контейнерів на дві половини — по 557 контейнерів. Першу половину використовували для навчання класифікатора, другу — як контрольну. Для отримання достатньо стабільних результатів у сенсі незалежності оцінок якості від поділу на навчальну та контрольну вибірки в подальшому повторювали експерименти по десять разів, змінюючи кожного разу стартове число генератора. Остаточні оцінки обчислювали як середнє значення в кожній такій серії тестів, а для оцінювання рівня стабільності рахували середньоквадратичне відхилення.

Зауважимо, що всі вісім моделей є ефективними для задач виявлення в частотній зоні НЗБ-стеганографії [13, 14]. У разі корисного навантаження 1 КБ (20–30 % для описаних контейнерів) вони з високою точністю виявляють стегановкладки програм Jsteg (98–100 %), Jphide (85–95 %) та Steganos Privacy Suite (93–99.8%). Результати виявлення цими моделями J-UNIWARD суттєво гірші (табл. 1).

З огляду на дані цієї таблиці можна дійти висновку, що J-UNIWARD з корисним навантаженням 10 % цілком безпечний, тобто він практично не розпізнається наявними статистичними методами на базі машинного навчання. Проте є три моделі, які все ж таки залишилися чутливими до змін, привнесених J-UNIWARD з малим навантаженням. За спаданням точності це: GFR — модель характеристичних векторів, побудованих як гістограми квантованих залишків, отриманих з використанням двовимірних фільтрів Габора (Gabor Filter Residual); PHARM — фазо-орієнтована проєкційна модель (PHase Aware pRojection Model); DCTR — модель характеристичних векторів із ДКП залишків (Discrete Cosine Transform Residual). Зазначимо, що між ними є спільні риси, зокрема, на відміну від усіх інших розглянутих моделей, ці три аналізують статистику не в частотному, а в просторовому представленні зображення (тобто не там, де безпосередньо здійснювалося приховування).

Таблиця 1

Оцінка якості	Точність класифікації за різними моделями векторів							
	CHEN	CC-CHEN	LIU	CC-PEV	CC-C300	GFR	DCTR	PHARM
Стеганоаналіз контейнерів з 50% корисним навантаженням								
Середня точність, %	55.5	61.1	61.8	61.7	58.0	85.2	79.2	81.4
Відхилення, %	0.7	0.9	0.8	1.1	1.2	0.9	0.9	1.0
Стеганоаналіз контейнерів з 10% корисним навантаженням								
Середня точність, %	49.7	50.6	50.4	50.5	50.3	54.2	52.4	52.6
Відхилення, %	0.4	0.7	0.7	0.5	1.3	0.5	0.9	0.5

ПІДВИЩЕННЯ ТОЧНОСТІ ДЕТЕКТУВАННЯ АНАЛІЗОМ НАЙБІЛЬШ ІМОВІРНО ЗАДІЯНИХ ЕЛЕМЕНТІВ ТА ЗЛИТТЯМ РЕЗУЛЬТАТІВ ГОЛОСУВАННЯ

Постає питання: чи можна підвищити отриману точність? Наведемо два способи та опишемо кроки їхньої реалізації.

1. Перший спосіб полягає у використанні контент-адаптивності J-UNIWARD. Тобто можна відстежувати статистику та аналізувати розбіжності не у всьому зображенні, а тільки у тих його ділянках, вкраплення в які є найбільш імовірним. Місце-

положення цих елементів будемо знаходити, використовуючи ту ж саму процедуру пошуку зашумлених та текстурованих ділянок, що закладена в J-UNIWARD.

Звісно, відтворити точно місцеположення вкраплення неможливо. Якщо це стеганозображення, то повторне застосування до нього J-UNIWARD призведе до іншого результату, та все ж він буде подібний до попереднього, бо візуально зображення залишилося таким, як є. Аналітик також не знає довжини прихованого повідомлення, проте він може розглядати певну верхню межу, яка визначається емпірично. Під час досліджень було визначено, що з огляду на досяжну точність для моделі DCTR є оптимальним розглядати довжину, що відповідає 100 % корисного навантаження, а для GFR та PHARM вищу точність забезпечують тести з довжиною, що близька до половини корисного навантаження (45 % в тестах).

Описаний підхід перевірено на стеганоконтейнерах J-UNIWARD з корисним навантаженням 50, 40, 30, 20, 10 та 5 %. Для цього порівняли точність детекту-

Таблиця 2

Модель	Точність класифікації у разі застосування моделей до всіх елементів та найбільш імовірно задіяних для корисного навантаження, %					
	50	40	30	20	10	5
GFR	85.2 (0.9)	77.8 (1.5)	69.4 (1.0)	60.9 (1.1)	54.2 (0.5)	51.5 (0.6)
modif_GFR	89.1 (0.7)	81.7 (0.8)	74.5 (0.7)	65.3 (0.9)	55.6 (0.9)	52.2 (0.7)
DCTR	79.2 (0.9)	71.3 (1.2)	63.8 (0.7)	57.1 (0.5)	52.4 (0.9)	50.6 (0.7)
modif_DCTR	79.7 (0.9)	73.1 (1.0)	65.1 (0.8)	59.5 (0.9)	55.6 (0.6)	54.8 (0.6)
PHARM	81.4 (1.0)	73.3 (1.3)	66.0 (0.8)	58.2 (0.6)	52.6 (0.5)	50.9 (0.5)
modif_PHARM	83.9 (0.9)	77.0 (0.7)	70.0 (0.8)	61.7 (0.6)	54.9 (0.8)	52.7 (0.7)

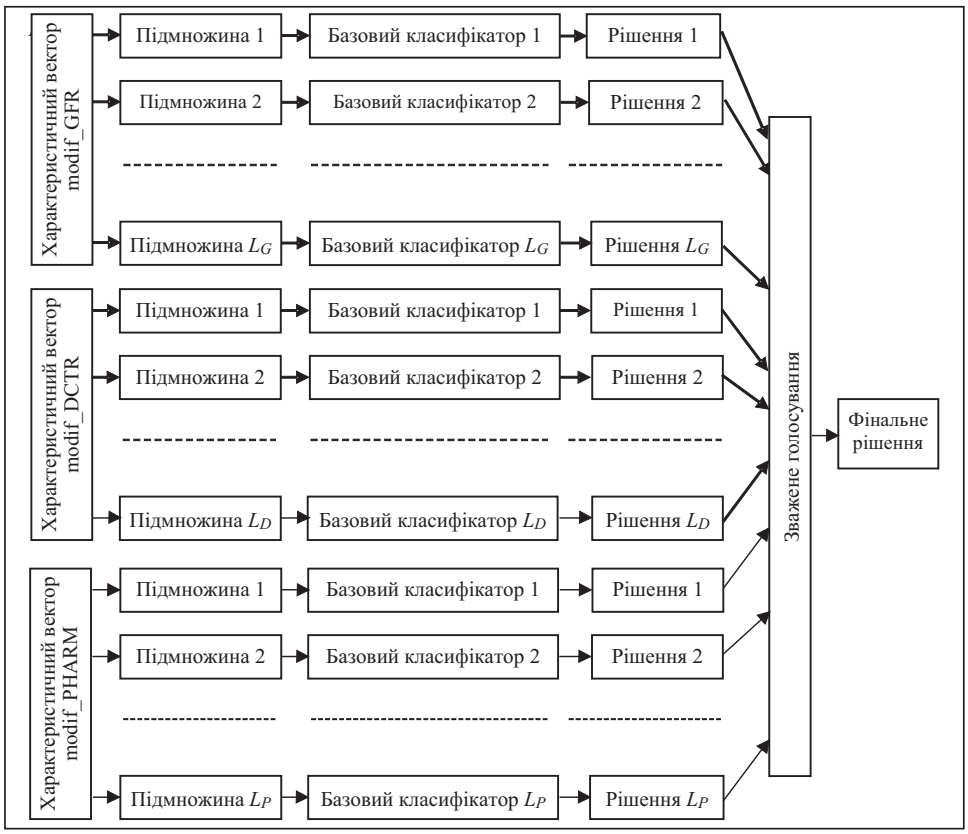

Рис. 3. Структурна схема злиття результатів трьох моделей

Таблиця 3

Точність класифікації у разі застосування зваженого голосування моделей для корисного навантаження, %					
50	40	30	20	10	5
89.6 (0.8)	82.3 (1.0)	74.8 (1.1)	65.7 (0.6)	57.4 (0.6)	54.3(0.9)

вання із застосуванням виділених на попередньому кроці моделей до всіх елементів контейнера та їх же у застосуванні до найбільш імовірних елементів-носіїв. Результати відповідних тестів наведено у табл. 2 (середня точність за десять експериментів, у дужках — середньоквадратичне відхилення).

Як видно, підвищення точності отримано для всіх варіантів корисного навантаження та всіх трьох моделей. У середньому внаслідок зазначеної модифікації точність стеганоаналізу підвищилась на 2.7 %.

2. Другий спосіб полягає у використанні для класифікації усіх трьох досліджуваних моделей. Для отримання фінального рішення про клас контейнера результати всіх трьох моделей об'єднуються згідно зі схемою на рис. 3. Перед підсумовуванням цих результатів вони зважуються з ваговим коефіцієнтом $1/L_G$ для моделі modif_GFR, $1/L_D$ — для modif_DCTR та $1/L_P$ — для modif_PHARM.

Результати такого зваженого голосування представлено у табл. 3. Згідно з ними цей підхід дав змогу підвищити точність у середньому ще на один відсоток у порівнянні з найбільш точним варіантом із розглянутих раніше (modif_GFR).

ДОСЛІДЖЕННЯ ВПЛИВУ ПАРАМЕТРІВ ТЕСТОВИХ КОНТЕЙНЕРІВ НА ТОЧНІСТЬ СТЕГАНОАНАЛІЗУ

Оцінки точності, зафіксовані у табл. 1–3, отримані для зображень, що мають однакові розміри та той самий коефіцієнт якості. Проте, якщо йдеться про практичний стеганоаналіз, де право вибору зображень не за стеганоаналітиком, а за протиборною стороною, потрібно попередньо класифікувати досліджувані зображення згідно з їхніми параметрами. Крім того, для спрощення класифікації можна ділити зображення на частини фіксованих розмірів. У подальшому для кожного класу зображень потрібно сформувати свою навчальну вибірку. Звісно, це ускладнює процес стеганоаналізу. Проте, якщо не здійснювати подібне попереднє оброблення, точність класифікації суттєво погіршується.

Для дослідження впливу параметрів контейнерів (їхніх розмірів та коефіцієнта якості) сформовано новий тестовий набір, з такою самою кількістю зображень, як і перший. Але коефіцієнт якості цих зображень варіюється в межах від 20 до 100 % (87 % у середньому), ширина зображень — 150–900 пікселів, а довжина — 103–800.

Здійснивши стеганоаналіз з тими самими параметрами, що й для першого набору для J-UNIWARD стеганоконтейнерів з 50 % корисним навантаженням, отримали точність, вказану в другому рядку табл. 4. У порівнянні з першим набором точність впала на 25 % (modif_GFR), 16 % (modif_DCTR) та 23 % (modif_PHARM і варіант об'єднання моделей). Зауважимо, що для визначення кроку квантування q у моделі modif_GFR використано лінійне наближення $q = \sigma(107 - K)/8$, запропоноване в [15], де σ — коефіцієнт масштабування, K — коефіцієнт якості зображення. Також виконано перехресні тести, коли класифікатор навчався за вибіркою, сформованою з одного набору, а для контролю використовувалися контейнери з іншого. Результати цих тестів наведено у третьому та четвертому рядках табл. 4.

Експерименти підтвердили, що вибір контейнерів з різними розмірами та коефіцієнтом якості для прихованого передавання даних ускладнює роботу стеганоаналітика.

Таблиця 4

Навчальний набір	Контрольний набір	Точність класифікації пустих та J-UNIWARD стеганоконтейнерів з 50 % корисним навантаженням			
		modif_GFR	modif_DCTR	modif_PHARM	Fusion
1	1	89.1 (0.7)	79.7 (0.9)	84.0 (0.7)	89.6 (0.8)
2	2	63.9 (0.9)	63.8 (0.9)	60.8 (0.9)	66.2 (0.9)
1	2	52.9 (0.7)	55.4 (0.6)	53.3 (0.4)	53.7 (0.6)
2	1	65.1 (4.3)	62.0 (1.5)	59.5 (2.0)	65.1 (3.9)

ВИСНОВКИ

У реальних умовах у користувача J-UNIWARD немає обмежень у виборі вихідних зображень, що будуть носіями прихованих повідомлень. Тому без попередньої класифікації досліджуваних контейнерів більш наближеними до реалій є оцінки, отримані в останніх серіях експериментів. Це становить 54–66 % у разі виявлення стеганоконтейнерів з 50 % навантаженням. Безперечно, ще меншою буде точність у разі зменшення корисного навантаження. До того ж, коли класифікатор навчається на контейнерах з різними можливими параметрами (коефіцієнтом якості та розмірами), а в подальшому використовується для розпізнавання зображень однакових параметрів, фіксується вища точність стеганоаналізу, ніж для інверсного перехресного тестування. Водночас для цього варіанту отримуємо й більше, ніж зазвичай, середньоквадратичне відхилення, тобто вміст навчальної вибірки сильніше впливає на результат стеганоаналізу.

Отже, сліпе виявлення J-UNIWARD контейнерів з корисним навантаженням 50 % та меншим на практиці є складною задачею, що потребує більш детального дослідження. Для проєктування дієвих стеганоаналітичних систем є доцільним, зокрема, введення до розгляду більших на порядок тестових наборів, дослідження питання попередньої класифікації зображень та її критеріїв, аналіз задачі формування навчальної вибірки для подальшої класифікації контейнерів з широким діапазоном корисних навантажень тощо.

СПИСОК ЛІТЕРАТУРИ

1. Holub V., Fridrich J., Denemark T. Universal distortion function for steganography in an arbitrary domain. *EURASIP Journal on Information Security*. 2014. N 1. P. 1–13. https://doi.org/10.1186/1687-417X-2014-1.
2. Filler T., Judas J., Fridrich J. Minimizing embedding impact in steganography using trellis-coded quantization. *Proc. SPIE, Electronic Imaging, Media Forensics and Security II* (27 Jan 2010, San Jose, USA). San Jose, USA, 2010. Vol. 7541. P. 1–14. https://doi.org/10.1117/12.838002.
3. Sidorenko V., Zyablov V. Decoding of convolutional codes using a syndrome trellis. *IEEE Trans. on Information Theory*. 1994. Vol. 40, N 5. P. 1663–1666. https://doi.org/10.1109/18.333887.
4. Chen C., Shi Y.Q. JPEG image steganalysis utilizing both intrablock and interblock correlations. *IEEE ISCAS, Intern. Symp. on Circuits and Systems* (18–21 May 2008, Seattle, USA). Seattle, USA: IEEE, 2008. P. 3029–3032. https://doi.org/10.1109/ISCAS.2008.4542096.
5. Kodovsky J., Fridrich J. Calibration revisited. *Proc. of the 11th ACM Multimedia and Security Workshop* (Sept 2009, Princeton, USA). New York: ACM, 2009. P. 63–74. https://doi.org/10.1145/1597817.1597830.
6. Liu Q. Steganalysis of DCT-embedding based adaptive steganography and YASS. *Proc. of the 13th ACM Multimedia & Security Workshop* (Sept 2011, Buffalo, USA). New York: ACM, 2011. P. 77–86. https://doi.org/10.1145/2037252.2037267.
7. Pevny T., Fridrich J. Merging Markov and DCT features for multiclass JPEG steganalysisю *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX* (2 March 2007, San Jose, USA). San Jose, USA, 2007. Vol. 6505. P. 301–314. https://doi.org/10.1117/12.696774.
8. Kodovsky J., Fridrich J. Steganalysis in high dimensions: fusing classifiers built on random subspaces. *8th SPIE Electronic Imaging, Media, Watermarking, Security and Forensics* (23–27 Jan 2011, San Francisco, USA). San Francisco, USA, 2011. Vol. 7880, P. 1–13. https://doi.org/10.1117/12.872279.
9. Song X., Liu F., Yang C., Luo X., Zhang Y. Steganalysis of adaptive JPEG steganography using 2D Gabor filters. *Proc. of the 3rd ACM Workshop on Information Hiding and Multimedia Security. ACM* (June 2015, Portland, USA). New York: ACM, 2015. P. 15–23. https://doi.org/10.1145/2756601.2756608.

10. Holub V., Fridrich J. Low complexity features for JPEG steganalysis using undecimated DCT. *IEEE Trans. on Information Forensics and Security.* 2015. Vol. 10, N 2. P. 219–228. https://doi.org/10.1109/TIFS.2014.2364918.
11. Holub V., Fridrich J. Phase-aware projection model for steganalysis of JPEG images. *Proc. SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XVII* (8–10 Feb 2015, San Francisco, USA). San Francisco, USA, 2015. Vol. 9409. https://doi.org/10.1117/12.2075239.
12. Kodovský J., Fridrich J., Holub V. Ensemble classifiers for steganalysis of digital media. *IEEE Trans. on Information Forensics and Security*. 2012. Vol. 7, N 2. P. 432–444. https://doi.org/10.1109/TIFS.2011.2175919.
13. Koshkina N.V. Comparison of efficiency of statistical models used for formation of feature vectors by JPEG images steganalysis. *Theoretical and Applied Cybersecurity.* 2020. Vol. 2, N 1. P. 22–28. https://doi.org/10.20535/tacs.2664-29132020.1.
14. Кошкіна Н.В. Дослідження основних компонентів систем JPEG-стеганоаналізу на базі машинного навчання. *Захист інформації.* 2020. Т. 22, № 2. С. 97–108. https://doi.org/10.18372/2410-7840.22.14801.
15. Song X., Liu F., Chen L., Yang C., Luo X. Optimal Gabor filters for steganalysis of content-adaptive JPEG steganography. *KSII Trans. on Internet and Information Systems*. 2017. Vol. 11, N 1. P. 552–569. https://doi.org/10.3837/tiis.2017.01.029.

Надійшла до редакції 31.08.2020

Н.В. Кошкина
СТЕГАНОАНАЛИЗ J-UNIWARD

Аннотация. Рассмотрена проблема выявления адаптивной стеганографии по методу J-UNIWARD стеганоаналитическими системами на базе машинного обучения. С помощью сравнительного анализа точности определено, что наиболее чувствительны к J-UNIWARD статистические модели построения характеристических векторов, формируемых в пространственной зоне, — GFR, PHARM и DCTR. Предложены два способа повышения точности стеганоанализа с использованием этих моделей: анализ наиболее вероятных местоположений внедрения; взвешенное голосование по трем моделям. Показано, что без предварительной классификации изображений согласно их параметрам точность стеганоанализа существенно понижается. Полученные результаты можно использовать для построения эффективных систем стеганоанализа изображений в формате JPEG.

Ключевые слова: информационная безопасность, стеганография, J-UNIWARD, стеганоанализ, методы с обучением и классификацией, точность детектирования.

N.V. Koshkina
J-UNIWARD STEGANOANALYSIS

Abstract. The author analyzes the problem of detecting adaptive steganography by the J-UNIWARD method by steganoanalytical systems based on machine learning. A comparative analysis of the accuracy has determined that statistical models of constructing characteristic vectors that are calculated in the spatial domain, such as GFR, PHARM and DCTR, are most sensitive to J-UNIWARD. Two ways to improve the accuracy of steganoanalysis based on these models are proposed: via the analysis of the most probable embedding locations and via the balanced vote on the three models. Significant degradation of the accuracy of steganoanalysis without preliminary classification of images according to their parameters is demonstrated. The obtained results can be used to generate efficient steganoanalysis systems for JPEG images.

Keywords: information security, steganography, J-UNIWARD, steganalysis, machine learning methods, detection accuracy.

Кошкіна Наталія Василівна,
докторка техн. наук, старша наукова співробітниця Інституту кібернетики імені В.М. Глушкова НАН України, Київ, e-mail: nata.koshkina@gmail.com.