УДК 004.056:519.246

ЕФЕКТИВНІСТЬ SPAM-МОДЕЛІ ДЛЯ РОЗПІЗНАВАННЯ ФАКТУ ВИКОРИСТАННЯ ДВОСТУПЕНЕВИХ МЕТОДІВ ВБУДОВУВАННЯ СТЕГОДАНИХ В ЦИФРОВІ ЗОБРАЖЕННЯ

Д. О. Панічева[1, а], Д. О. Прогонов[1, б], С. М. Кущ[1, в]

[1] *Національний технічний університет України «Київський політехнічний інститут»*

Анотація

У роботі досліджено ефективність використання сучасних статистичних моделей контейнера, зокрема SPAM-моделі цифрових зображень, для виявлення стеганограм, сформованих згідно двоступеневим методами приховання повідомлень в області перетворення контейнера (ОПК). Розглянуто випадок налаштування стегодетектора з використанням окремих каналів кольору чистих та заповнених контейнерів, що є найбільш близьким до реальних сценаріїв функціонування стегодетекторів. Встановлено, що використання комплексних статистичних моделей контейнера дозволяє з високою точністю розпізнавати наявність прихованих повідомлень навіть при слабкому заповненні контейнера стегоданими (менше 10%). Отримані результати можуть бути використані для виявлення стеганограм з даними, вбудованими в ОПК.

Ключові слова: стегоаналіз, статистичний аналіз, SPAM-модель

Вступ

Тенденцією розвитку існуючих стеганографічних систем (СС) є широке використання багатоетапних методів вбудовування стегоданих в області перетворення контейнера, зокрема цифрових зображень (ЦЗ). Це дає можливість суттєво підвищити стійкість (робастність) отримуваних стеганограм до відомих методів пасивного (хі-квадрат тест, RS-аналіз) та активного (фільтрація, стиснення ЦЗ) аналізу у порівнянні з класичними LSB-алгоритмами приховання повідомлень. Актуальною і важливою задачею є вдосконалення існуючих методів пасивного стегоаналізу для підвищення точності виявлення стеганограм при багатоетапному вбудовуванні стегоданих в ОПК.

Класичним підходом до виявлення стеганограм є використання статистичного стегоаналізу [1, 2]. Дані методи дозволяють з високою точністю розпізнавати наявність стегоданих, вбудованих в просторовій області ЦЗ, проте їх ефективність значно зменшується у випадку приховання повідомлень в ОПК. В літературі відсутні результати оцінки ефективності методів статистичного стегоаналізу при використанні стегоданих різних типів.

В роботі [3] було показано, що застосування комплексних статистичних моделей контейнера (КСМК) дозволяє підвищити точність розпізнавання стеганограм з даними, прихованими в ОПК ЦЗ з використанням одноетапних методів. Представляє інтерес дослідження ефективності застосування КСМК у випадку багатоетапного приховання повідомлень в ОПК ЦЗ.

[а] panicheva.d@gmail.com
[б] progonov@gmail.com
[в] gonorskaya@ukr.net

Метою роботи є визначення ефективності КСМК, на прикладі SPAM-моделі, для виявлення стеганограм, сформованих згідно з двоетапними методами приховання повідомлень, при вбудовуванні різних типів стегоданих.

1. Статистична модель SPAM цифрових зображень

Модель SPAM (Subtractive Pixel Adjacency Matrix) ЦЗ заснована на аналізі змін його шумових компонент при вбудовуванні повідомлень з використанням теорії марківських процесів [4]. У якості вектора ознак для стегоаналізатора використовується матриця ймовірностей переходу, отримана у процесі моделювання міжпіксельних залежностей як марківського ланцюга вищого порядку.

Для зменшення розмірності простору ознак при налаштуванні стегодетектора в SPAM-моделі проводиться оцінка різниці між показниками яскравості сусідніх пікселів зображення (аналог високочастотної фільтрації ЦЗ). Це не призводить до суттєвих втрат в точності розпізнавання стеганограм, оскільки зміна яскравості між суміжними пікселями $I_{i,j}$ та $I_{i,j+1}$ є незначною і не залежить від значення пікселя $I_{i,j}$ [4].

На першому етапі побудови SPAM-моделі розраховуються ймовірності переходів у 8-ми напрямках відносно фіксованого пікселя: горизонтальному, вертикальному та діагональному [4]. Отримані залежності досліджуються як марківський процес першого F^{1st} та другого F^{2nd} порядків [4]. Наприклад, для горизонтального напрямку отримуємо:

$$D_{i,j} = I_{i,j} - I_{i,j+1}, i \in \{1, \dots, m\}, j \in \{1, \dots, n-1\}.$$

$$M^{\rightarrow}_{u,v} = P(D^{\rightarrow}_{i,j+1} = u | D^{\rightarrow}_{i,j} = v).$$

$$M^{\rightarrow}_{u,v,w} = P(D^{\rightarrow}_{i,j+2} = u | D^{\rightarrow}_{i,j+1} = v), D^{\rightarrow}_{i,j} = w,$$
$$u, v, w \in \{-T, \ldots, T\}.$$

На другому етапі побудови SPAM-моделі проводиться об'єднання результатів, отриманих для різних напрямків відносно заданого пікселя [4]:

$$F_{1,\ldots,kj} = \frac{1}{4}[M^{\rightarrow} + M^{\leftarrow} + M^{\downarrow} + M^{\uparrow}],$$
$$F_{k+1,\ldots,2kj} = \frac{1}{4}[M^{\searrow} + M^{\nwarrow} + M^{\swarrow} + M^{\nearrow}],$$

де $k = 2(2T+1)^2$ для параметрів SPAM-моделі першого порядку та $k = 2(2T+1)^3$ – для параметрів SPAM-моделі другого порядку.

Відповідно до рекомендацій [4], при налаштуванні SPAM-моделі було використано марківський ланцюг другого порядку, оскільки його використання дозволяє більш точно моделювати залежності різниць яскравостей суміжних пікселів ЦЗ. Загальна розмірність SPAM-моделі становила 686 параметрів (при $T = 3$).

2. Процедура налаштування стегодетектора

В роботі було використано ансамбль стегодетекторів [5], що складався з L базових класифікаторів B_l, $l = 1, \ldots, L$, які незалежно навчалися на вибірці чистих та заповнених контейнерів. У якості базових класифікаторів було обрано лінійні дискримінатори Фішера (ЛДФ) [6], що обумовлено їх низькою навчальною складністю.

Кожний базовий класифікатор навчався на вибірці $x_i^{D_1}, \overline{x}_i^{D_1} \ | i \in N_1^b$, де $D_1 \subseteq \{1, \ldots, d\}$, $|D_1| = d_{sub}$ – псевдовипадково обрана підмножина тестових ЦЗ. Для тестового ЦЗ l-ий базовий класифікатор приймає рішення шляхом порівняння визначених параметрів ЦЗ $y \in Y^{Tc}$ з попередньо встановленими пороговими значеннями. Кінцеве рішення стегодетектора формується на основі порівняння суми рішень всіх L базових класифікаторів з пороговим значенням [5].

Для оцінки похибки розпізнавання, на етапі налаштування стегодетектора, була використана стандартна out-of-bag (OOB)-оцінка, що визначається таким чином [7].

$$E^{(L)} = \frac{1}{2 \cdot N^{Tp}} \cdot \sum_{m=1}^{N^{EP}} (B^{(L)}(x^{(m)} + 1) - B^{(L)}(\overline{x}^{(m)})).$$

При налаштуванні стегодетектора проводиться корегування кількості базових класифікаторів L – для кожного значення d_{sub} кількість базових класифікаторів L покроково збільшується на одиницю та визначається відповідна OOB-оцінка. Оптимальне значення L_{opt} визначається за критерієм мінімуму OOB-оцінки похибки розпізнавання стеганограм на навчальній вибірці тестових ЦЗ.

3. Метод Джозефа

Даний метод вбудовування даних заснований на дворівневому дискретному вейвлет-перетворенні (ДДВП) та сингулярному розкладі (СР) ЦЗ. Формування стеганограм проводиться у декілька етапів. На першому етапі до зображення-контейнера (ЗК) застосовується дворівневе вейвлет-перетворення Хаара, а до стегоданих, представлених у вигляді ЦЗ – однорівневе вейвлет-перетворення Хаара [8]. Далі у HL та LH піддіапазонах контейнера (C) та стегоданих (D) відбувається СР, відповідно [8]:

$$HL_{C \setminus D} = U_{HL_{C \setminus D}} \times S_{HL_{C \setminus D}} \times V^T_{HL_{C \setminus D}},$$
$$LH_{C \setminus D} = U_{LH_{C \setminus D}} \times S_{LH_{C \setminus D}} \times V^T_{LH_{C \setminus D}},$$

де U, V – відповідно, матриці лівих та правих сингулярних векторів; S – діагональна матриця сингулярних чисел.

На наступному етапі визначається сума коефіцієнтів розкладу ЗК та стегоданих [8]:

$$S_{WM} = S_C + \alpha \cdot S_D,$$

де α – ваговий коефіцієнт вбудовування стеганограми, мінімальне значення якого відповідає випадку, при якому вбудовані дані неможливо вилучити на приймальній стороні СС, а максимальне – появі помітних візуальних спотворень контейнера.

Для формування стеганограми в просторовій області проводиться множення модифікованої матриці сингулярних чисел та матриці лівих та правих власних векторів з подальшим застосуванням зворотного ДДВП [8].

Використання HL та LH піддіапазонів зумовлено тим, що вбудовування в інших піддіапазонах призводить до погіршення маскування стегоданих (приховання в діапазоні НЧ) або зниження стійкості прихованих повідомлень до активних атак (вбудовування в діапазоні ВЧ) [8].

4. Результати роботи

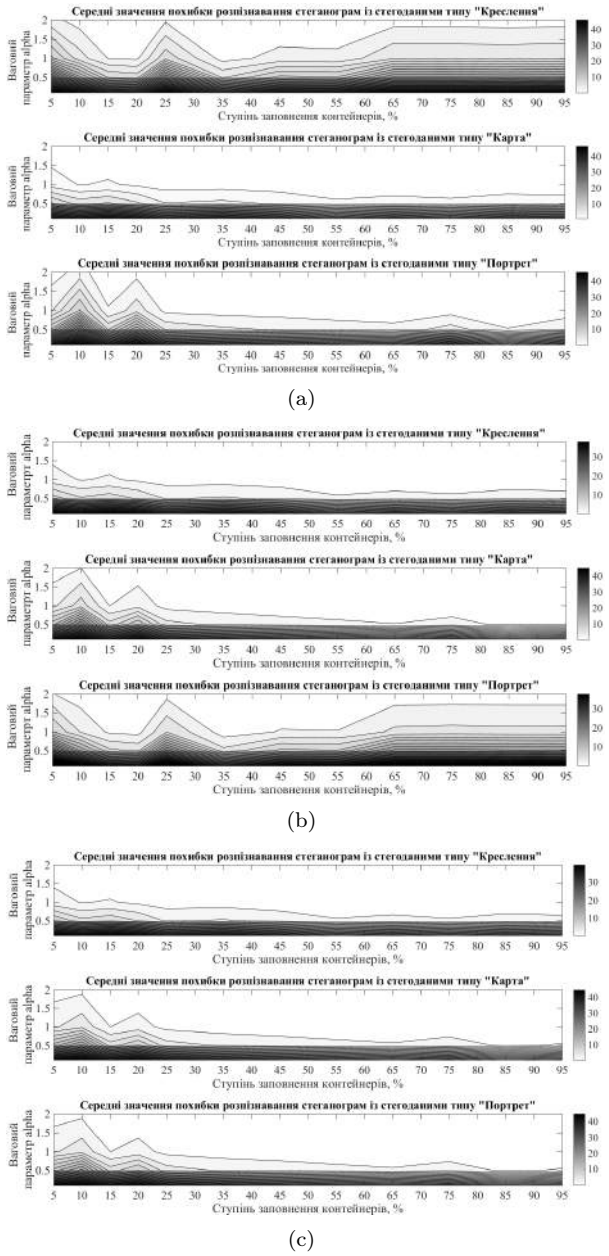
При проведенні досліджень у якості тестового пакету зображень було використано псевдовипадкову вибірку з 9000 ЦЗ із пакету MIRFlickr-25к. Тестові ЦЗ були масштабовані до розміру 512×512 пікселів та поділені на навчальну та контрольну вибірки однакового розміру. Рівень заповнення тестових зображень стегоданими (частка змінених коефіцієнтів СР ЦЗ) варіювався від 5% до 25%, з кроком 5%, та від 25% до 95%, з кроком 10%.

У якості стегоданих було використано кольорові зображення трьох типів: "Креслення", "Карта" та "Портрет". Вбудовування стегоданих у контейнер відбувалося при варіації вагового коефіцієнта α в інтервалі $\{0.5;1;2;5\}$.

Тестування стегодетектора було проведено для випадку аналізу окремих каналів кольору ЦЗ, що відповідає найбільш наближеній до реальної ситуації сте-

Табл. 1. Значення похибки розпізнавання стеганограм для різних типів стегоданих

	Канал кольору		
Тип стегоданих	Червоний	Зелений	Синій
"Креслення"	0,41;0,003	0,39;0,003	0,41;0,003
"Карта"	0,48;0,002	0,47;0,002	0,46;0,002
"Портрет"	0,48;0,001	0,48;0,002	0,46;0,002



(a)

(b)

(c)

Рис. 1. Середні значення похибки розпізнавання стеганограм для трьох типів стегоданих та використанні: (a) – червоного каналу кольору ЦЗ; (b) – зеленого каналу кольору ЦЗ; (c) – синього каналу кольору ЦЗ.

гоаналізу ЦЗ. Результати тестування стегодетектора (максимальні\мінімальні значення ООВ-оцінки похибки розпізнавання) представлені у таблиці (таб. 1).

Найбільші значення похибки розпізнавання стеганограм були отримані при $\alpha = 0,5$, тобто для випадку, при якому неможливим є вилучення вбудованих даних на приймальній стороні СС. При підвищенні енергії стегоданих значення похибки розпізнавання стеганограм зменшується і не перевищує значення 0,13% (α= 1), що свідчить про високу ефективність використання SPAM-моделі для розпізнавання стеганограм.

На рис. 1 представлені середні значення похибки розпізнавання стеганограм для трьох типів стегоданих при аналізі окремих каналів кольору ЦЗ в залежності від ступеня заповнення контейнера.

Незалежно від типу стегоданих найбільші значенні ООВ-оцінки похибки спостерігаються для стеганограм, сформованих із ваговим коефіцієнтом $\alpha = 0.5$, що відповідає найменшим спотворенням контейнера, та, як наслідок, найменшій зміні статистичних залежностей шумових компонент контейнера. Також характерним для всіх стегоданих виявилося підвищення значення ООВ-оцінки похибки при малих ступенях заповнення контейнера (до 25%), але треба відмітити, що у випадку слабого заповнення контейнера ймовірність визначення стеганограм іншими відомими методами є дуже низькою.

Висновки

В роботі встановлено, що застосування SPAM-моделі ЦЗ дає можливість з високою точністю виявляти стеганограми з даними, прихованими в ОПК, з використанням двоступеневих методів.

Експериментально підтверджено, що найбільшу складність при стегоаналізі становлять стеганограми, сформовані при малих значеннях вагового коефіцієнта вбудовування стегоданих та малих ступенях заповнення контейнера (до 25%). Проте, навіть за таких умов, похибка розпізнавання стеганограм при використанні SPAM-моделі не перевищує 0.13%.

Перелік використаних джерел

1. Fridrich J. Steganography in Digital Media: Principles, Algorithms and Applications. — Cambridge University Press,New York, USA. — 2010. — 437 p.
2. Панічева Д. О., Прогонов Д. О., Кущ С. М. Статистичні характеристики стеганограм при вбудовуванні повідомлень у просторовій області контейнеру. — XII наукова конференція “Теоретичні і прикладні проблеми фізики, математики та інформатики”. — К.: “Політехніка”, 2014. — с. 201-202.
3. Прогонов Д. О., Кущ С. М. Статистичний аналіз стеганограм з даними, прихованими в області перетворення контейнеру. — Збірник матеріалів науково-практичної конференції “Актуальні проблеми управління інформаційною безпекою держави”. — 19 березня 2015 року, м.Київ. — К.: Центр навч., наук. та період. видань НА СБ України, 2015. — с. 329-332;
4. T. Pevny Steganalysis by Subtractive Pixel Adjacency Matrix. \ T. Pevny, P. Bas and J. Fridrich \\ IEEE Trans. on Info. Forensics and Security. — 2010 — Vol. 5, Issue 2.— pp. 215-224.
5. J. Kodovsky Ensemble classifiers for steganalysis of digital media. \ J. Kodovsky, J. Fridrich, and V. Holub \\ IEEE Transactions on Information Forensics and Security. — 2012.
6. Duda R. O., Hart P. E. and Stork D. G. Pattern Classification. — 2^{nd} edition. — New York: John Wiley&Sons, Inc. — 2001.
7. Breiman Bagging predictors. Machine Learning.— 24:123–140. August 1996.
8. Joseph A. Robust watermarking based on DWT-SVD. \ Joseph A., Anusudha K. \\ International Journal on Signal&Image Security. — Vol. 1, Issue. 1, Oct 2013.