Терейковський Ігор Анатолійович, кандидат технічних наук, доцент, доцент кафедри системного програмування та спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут».
E-mail: terejkowski@ukr.net
Терейковский Игорь Анатольевич, кандидат технических наук, доцент, доцент кафедры системного программирования и специализированых компьютерных систем Национального технического университета Украины «Киевский политехнический институт».
Tereykovskiy Igor, Ph.D., associate professor, assistant professor of specialized systems programming and computer systems of the National Technical University of Ukraine "Kiev Polytechnic Institute".

УДК 004.056.5

СТЕГАНОАНАЛИТИЧЕСКИЙ АЛГОРИТМ ДЛЯ ИЗОБРАЖЕНИЙ, ПОДВЕРГАВШИХСЯ ОПЕРАЦИИ СЖАТИЯ С ПОТЕРЯМИ

Владимир Рудницкий, Илья Узун

Легкость в применении, а также масса программных средств как платных, так и бесплатных, свободно распространяемых по сети, сделали стеганографию очень популярным инструментом, за счет которого можно обеспечить простой способ организации утечки ценной информации и неконтролируемый обмен информацией в противозаконных целях. Данные обстоятельства вынуждают активизировать усилия по разработке алгоритмов стеганоанализа. Одним из видов таких алгоритмов являются алгоритмы, базирующиеся на анализе пар цветов цифрового изображения. Большинство существующих подобных средств ориентировано на работу с изображениями, хранимыми в форматах без потерь, что значительно сужает область их применения. С учетом этого, в работе был предложен стеганоаналитический алгоритм определения наличия секретного сообщения, погруженного в цифровое изображение, хранимое в формате с потерями (JPEG), при помощи метода модификации наименьшего значащего бита. Полученные результаты показали, что принцип анализа пар цветов может быть успешно применен и для стеганоанализа цифровых изображений подвергавшихся операции сжатия с потерями.

***Ключевые слова**: стеганография, стеганоанализ, близкие пары цветов, уникальные цвета, сокрытие информации.*

Введение. Стремительное развитие информационных технологий и динамический рост форматов цифровых данных обеспечивают практически неограниченные возможности для сокрытия информации. Одной из наук, занимающихся скрытием данных, является стеганография [4, 5]. Отличительной особенностью стеганографии является то, что при передаче секретной информации в тайне здесь остается сам факт передачи. Преимущество стеганографии состоит в том, что она предоставляет возможность скрытно передать конфиденциальное сообщение – дополнительную информацию (ДИ) одновременно с открытой информацией – контейнером или основным сообщением (ОС), которое не является конфиденциальным. В качестве ОС может быть выбран любой мультимедиа объект – цифровое изображение (ЦИ), видео или аудио (в настоящей работе как контейнер используется ЦИ). В результате погружения ДИ в ОС не должно происходить заметных изменений контейнера. Данный процесс будем называть стеганопреобразованием (СП), а его результат – стеганосообщением (СС). Использование СП часто позволяет избежать прямых атак на ДИ, поскольку неизвестно, присутствует ли она в информационном потоке. ДИ, вносимая в контейнер, может быть предварительно зашифрована, чтобы усложнить задачу стегоаналитика [4]. Основная задача стеганоанализа (СА) [4, 5] – установление факта присутствия в контейнере скрытой информации.

Легкость в применении, а также масса программных средств как платных, так и бесплатных (Steganos, StegHide, S-tools и др.), свободно распространяемых по сети, сделали стеганографию очень популярным инструментом. Это простой способ для организации утечки ценной информации из компаний, неконтролируемого обмена информацией между подозреваемыми и правонарушителями и т.д. Посредством стеганографии между собой общаются как секретные государственные службы, шпионы [15], так и криминальные структуры, и террористы [1, 10, 12]. Поэтому развитие методов СА на сегодняшний день является задачей, актуальность которой трудно переоценить. Работа СА заключается в поиске и анализе определенных характеристик и признаков в исследуемом цифровом объекте, определение

факта наличия или отсутствия которых позволяет получить ответ на вопрос, является ли анализируемый объект СС или же он не подвергался СП.

В настоящее время на рынке программных продуктов можно встретить достаточное количество стеганопрограмм, разработанных под некоторые форматы графических, видео и аудиофайлов, используемых в Интернете. В большинстве из них применяются различные модификации LSB-метода, основной идеей которого является использование одного или нескольких младших двоичных разрядов интенсивности цветовых компонент отдельных пикселей для внедрения ДИ. Популярность данного метода обусловлена его простотой и тем, что он позволяет скрывать в относительно небольших файлах достаточно большие объемы информации [7]. Визуально изображение при этом не изменяется, особенно если в качестве ОС выбрано многоцветное изображение с большим количеством деталей, то есть информационно нагруженное. Если, например, взять ЦИ цветовой модели RGB, на каждую компоненту цвета R, G и B которого отводится 8 бит, и изменить значения наименьших значащих бит (НЗБ) – то подобное искажение будет неуловимо для человеческого восприятия [3]. Это в значительной степени осложняет работу стегоаналитика, если он не обладает специальными средствами СА. В качестве таких средств могут выступать программы, реализующие методы и алгоритмы стегоанализа. К таким средствам СА относятся алгоритмы, основанные на анализе пар цветов, которые являются эффективными и широко используемыми, но (до настоящего момента) только для ЦИ-контейнеров, хранимых в форматах без потерь [8, 9, 11, 13, 14]. Специфика этих стеганоаналитических алгоритмов до настоящего момента не позволяла их использовать для СС, сформированных на основе ОС, сохраненных с потерями. Однако с ростом объемов используемой информации все чаще приходится прибегать к сжатию ЦИ. Стандартом де-факто в области сжатия ЦИ является алгоритм JPEG, основанный на дискретном косинусном преобразовании. Учитывая факты массового использования формата JPEG в целях сокрытия информации, *актуальным* является вопрос разработки стеганоаналитического алгоритма, основанного на анализе пар цветов, для контейнеров, хранимых в формате JPEG.

Цель статьи и постановка исследований. *Целью* статьи является разработка стеганоаналитического алгоритма (САА), основанного на анализе количества близких пар цветов и уникальных цветов для контейнеров, хранимых в формате с потерями (JPEG).

Данная работа является продолжением работы [6]. Как и в [6], для определения пороговых значений не проводится классификация анализируемых изображений по категориям, как это настоятельно рекомендуется и делается в [9, 13]. Для достижения поставленной цели необходимо решить следующие *задачи*:

1. Определить статистические характеристики, используемые для проверки ОС и СС, анализ которых, позволит отделить ЦИ, подвергавшиеся СП и ЦИ, не содержащие ДИ;

2. Выявить характерные особенности и отличия исходных ЦИ, не подвергавшихся СП, от СС – полученных после внедрения ДИ в ходе СП посредством модификации НЗБ;

3. Выявить характерные особенности и отличия изображений, уже подвергавшихся СП посредством модификации НЗБ, от СС – полученных после повторного СП. То есть, уже измененные битами случайного ЦИ, НЗБ анализируемого изображения, повторно модифицируются в ходе СП битами другого случайного ЦИ;

4. Исходя из результатов решения предыдущих задач, разработать САА для выявления СС, полученных при использовании метода LSB над изображениями в формате JPEG с потерями.

Основная часть. Введем необходимые обозначения и определения. Под цветом в дальнейшем будем понимать тройку компонент (R,G,B) или пиксель, который также подразумевается как триплет значений (R,G,B), где R – красная, G – зеленая и B – синяя компонента в цветовой модели RGB [3]. В качестве статистических характеристик для анализа выбраны коэффициенты близких пар цветов и уникальных цветов.

Пусть P – число близких пар цветов в изображении. Согласно определению, данному в [11] (для ЦИ, хранимых в форматах без потерь), под близкой парой понимают два цвета (R_1,G_1,B_1) и (R_2,G_2,B_2), если для них справедливо следующее соотношение:

$$\begin{cases} |R_1-R_2|\le 1, \\ |G_1-G_2|\le 1, \\ |B_1-B_2|\le 1 \end{cases} \Leftrightarrow \left(R_1-R_2\right)^2+\left(G_1-G_2\right)^2+\left(B_1-B_2\right)^2\le 3\,. \qquad (1)$$

При анализе изображений в формате BMP с использованием соотношения (1), не подвергавшихся сжатию и СП, четко наблюдается уменьшение количества близких пар цветов после СП [6]. В случае же с изображениями в формате JPEG последующее СП не вносит каких-либо заметных изменений. Разные картины в изменениях величин близких пар цветов в двух форматах объясняются тем, что в процессе сохранения ЦИ в JPEG (с потерями) происходит обнуление высокочастотных (и, возможно, некоторых среднечастотных) коэффициентов дискретного косинусного преобразования (ДКП) 8×8 – блоков, полученных после стандартного разбиения матрицы исходного изображения. Исключение высоких (и, возможно, средних) частот в JPEG ЦИ никак не восполнится при его пересохранении в формате TIF, поэтому матрица изображения, сохраненного в TIF первоначально и сохраненного в TIF после JPEG-сжатия должны качественно отличаться друг от друга по своим характеристикам [2]. Таким образом, соотношение (1), как проверено в ходе представительного вычислительного эксперимента, является «нерабочим» для контейнеров с потерями, поскольку в результате квантования частотных коэффициентов ЦИ, происходящего в процессе сжатия, количество цветов снижается (это является принципиальной проблемой, не позволявшей до сих пор использовать САА, основанный на анализе пар цветов для ОС с потерями [8, 9, 11, 13, 14]).

В настоящей работе предлагается внести следующие изменения в базовые для разрабатываемого САА понятия.

Определение 1. Под близкой парой для ЦИ в формате с потерями (JPEG) понимаются два цвета (R_1,G_1,B_1) и (R_2,G_2,B_2), если для них справедливо следующее соотношение:

$$\begin{cases}|R_1-R_2|\le 2,\\ |G_1-G_2|\le 2,\\ |B_1-B_2|\le 2\end{cases} \Leftrightarrow \quad (R_1-R_2)^2+(G_1-G_2)^2+(B_1-B_2)^2\le 12\,. \tag{2}$$

В САА, основанном на анализе пар цветов, значимым является определение уникального цвета. Согласно определению [9, 13] уникального цвета для ЦИ в форматах без потерь, два цвета (R_1,G_1,B_1) и (R_2,G_2,B_2) называются уникальными, если выполняется, хотя бы одно из условий:

$$\begin{aligned}&|R_1-R_2|\le 1,\\ &|G_1-G_2|\le 1,\\ &|B_1-B_2|\le 1.\end{aligned}$$

Определение 2. Два цвета (R_1,G_1,B_1) и (R_2,G_2,B_2) называются уникальными в ЦИ, хранимом в формате с потерями, если выполняется, хотя бы одно из условий:

$$\begin{aligned}&|R_1-R_2|\le 2,\\ &|G_1-G_2|\le 2,\\ &|B_1-B_2|\le 2.\end{aligned} \tag{3}$$

Пусть R – отношение количества близких пар цветов P к количеству уникальных цветов U, определяемых согласно (2), (3) соответственно:

$$R=\frac{P}{U}. \tag{4}$$

Коэффициент R играет ключевую роль при отделении ОС от СС в разработанном САА.

Для исследования было подготовлено 200 изображений размером 300×200 пикселей, которые были конвертированы в формат JPEG из TIF. Посредством LSB-метода из данной базы JPEG-контейнеров было получено 200 СС, которые были сохранены в формате TIF, поскольку LSB-метод является неустойчивым к любого рода возмущающим воздействиям. В качестве ДИ использовалась случайно сформированная бинарная 300×200-матрица.

При организации вычислительного эксперимента на вход поступало ЦИ в формате JPEG, для которого рассчитывался согласно (4) коэффициент R. Затем ЦИ подвергалось СП методом модификации НЗБ. Для полученного СС по формуле (4) вычислялся коэффициент R'. Основой разработанного САА, выполняющего отделение ОС от СС, является анализ показателей R и R': вывод о наличии (отсутствии) ДИ в ЦИ делается на основании их сравнения с использованием порогового значения.

При детальном исследовании изменения величин близких пар цветов, уникальных цветов и их отношений была установлена целесообразность для повышения эффективности процесса детектирования наличия ДИ производить не одно, а n СП над исследуемым ЦИ (в данном исследовании использовалось $n=30$).

На рис. 1 для иллюстрации полученных результатов представлены примеры типичных графиков изменения коэффициентов R и R' при СП. Каждый из рисунков а, б, в и г содержит в себе слева график для ЦИ, не содержащего дополнительной информации, и справа – соответствующего стеганосообщения. Ось ординат – значения R и R', а ось абсцисс – это количество изменяемых НЗБ при СП.

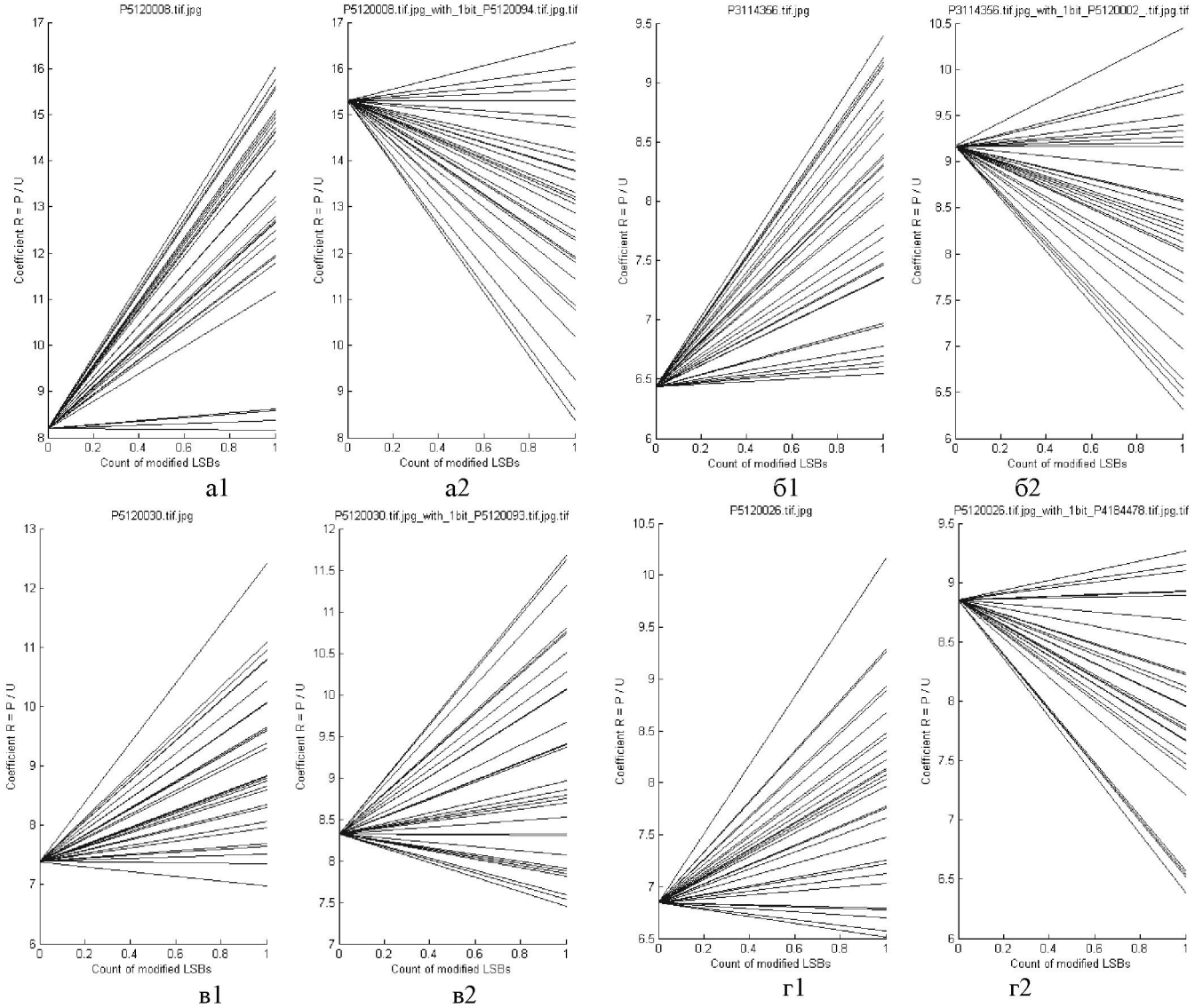


Рис. 1. Графики изменения коэффициентов R и R' при СП:
а1, б1, в1, г1 – в ЦИ, не содержащих ДИ; а2, б2, в2, г2 – в СС.

Из результатов вычислительного эксперимента вытекает, что в качестве разделителя СС и ОС возможно использовать:

– среднее значение из n коэффициентов R';

– минимальное значение из n коэффициентов R': в качестве оцениваемого значения берется величина $R - \min(R')$.

При анализе исходной базы ЦИ пороговое значение, в случае использования среднего значения по R', составило $T_1 = -1.4$; при использовании минимума по R' величина порога $T_2 = 0.64$. Эти значения были определены посредством анализа зависимости величин ошибок первого и второго рода от изменения порогового значения (рис. 2).

На основании вышесказанного основные шаги разработанного САА выглядят следующим образом:

1. Для анализируемого ЦИ вычислить:

– количество близких пар цветов P в соответствии с (2);

– уникальных цветов U в соответствии с (3);

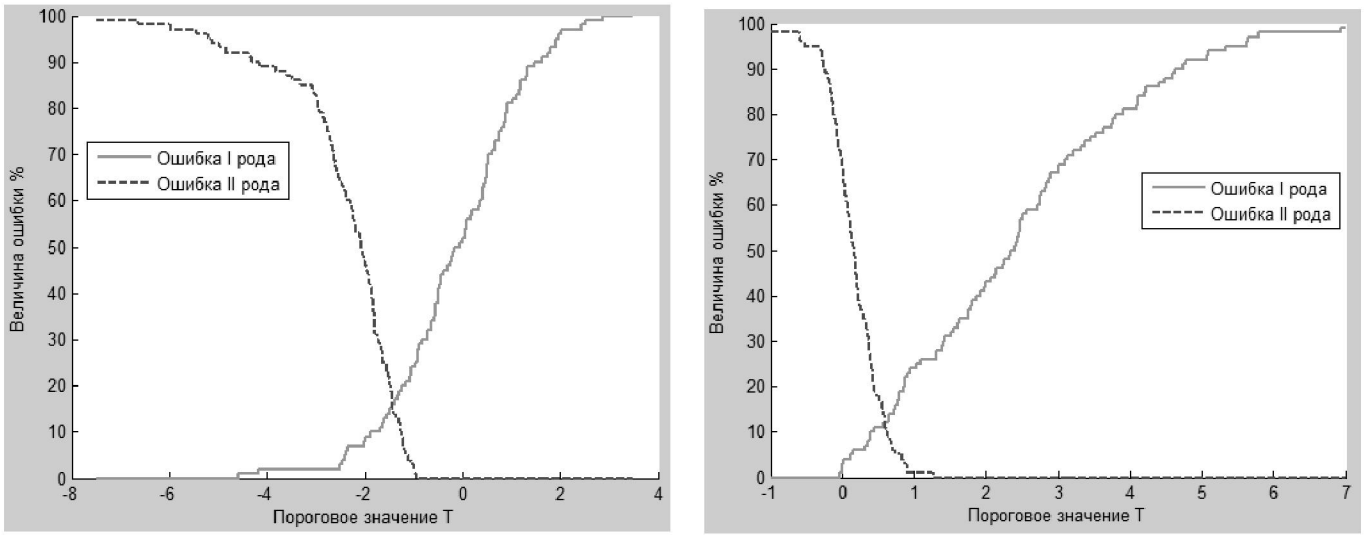
– коэффициент R по формуле (4).

2. Используя LSB-метод, внедрить ДИ – случайно сформированную бинарную матрицу в анализируемое ЦИ. Произвести подсчет

коэффициентов P', U' и R' для полученного стеганосообщения. Повторить данный шаг n раз.

3. Определить минимальное (среднее) значение в наборе из n значений R', полученных на шаге 2;

4. Если величина $R - \min(R') < T_2$ (среднее значение R' больше T_1), то анализируемое ЦИ не подвергалось стеганопреобразованию, иначе – анализируемое изображение является СС.



а б

Рис. 2. Графики зависимости ошибок I и II рода от порогового значения:
а – для T_1; б – для T_2

Заключение. Результаты вычислительного эксперимента, проведенного с целью проверки эффективности работы нового САА, представлены в таблице 1. На основании полученных результатов можно сделать вывод о предпочтительности использования порога T_2, обеспечивающего большую по сравнению с T_1 эффективность алгоритма.

Таблица 1

Результаты работы разработанного стеганоаналитического алгоритма

Порог	Ошибки I рода	Ошибки II рода
T_1	16	14
T_2	12	9

В данный момент усилия авторов направлены на дальнейшее повышение эффективности разработанного алгоритма, главным образом, на уменьшение ошибок первого рода, которое может быть достигнуто за счет изменения ширины эксперимента на шаге 2, а также уточнения пороговых значений.

ЛИТЕРАТУРА

[1]. Бобок И.И. Стеганоанализ, как частный случай анализа информационной системы / Бобок И.И., Кобозева А.А. // Сучасна спеціальна техніка. – 2011. – №1. – С. 25-36.

[2]. Бобок И.И. Стеганоаналитический алгоритм для основного сообщения, хранимого в форматах с потерями / И.И.Бобок // Вісник національного технічного ун-ту «ХПІ». – 2012. – № 29. – С.41-49.

[3]. Гонсалес Р., Вудс Р. Цифровая обработка изображений.- М.: Техносфера, 2005.– 1072 с.

[4]. Грибунин В.Г. Цифровая стеганография. / Грибунин В.Г., Оков И.Н., Туринцев И.В. – М.: Солон-Пресс, 2002. – 272 с.

[5]. Конахович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография. Теория и практика. – МК-Пресс, 2006.

[6]. Рудницкий В.Н. Стеганоаналитический алгоритм, основанный на анализе пар цветов / Рудницкий В.Н., Узун И.А.// Інформатика та математичні методи в моделюванні. – 2012. – Т.2, №3. – С. 210-220.

[7]. Швидченко И.В. Анализ криптостеганографических алгоритмов // Проблемы управления и информатики, 2007. – № 4. – С. 149-155.

[8]. Fridrich J., Du R., Long M. Steganalysis of LSB encoding in Color images, IEEE International Conference on Multimedia and Expo, vol.3, 2000, pp. 1279-1282.

[9]. Geetha S., Silva S.Sivatha Sindhu, Kamaraj N. Close color pair signature ensemble adaptive threshold based steganalysis for LSB embedding in digital images. Transactions on Data Privacy, 1(2009) 140-161.

[10]. Gul G., Kurugollu F. SVD-Based Universal Spatial Domain Image Steganalysis / IEEE Transactions on Information Forensics and Security. – 2010. – Vol. 5, No. 2. – P. 349-353.

[11]. Johnson Neil F., Jajodia Sushil. Steganography: Seeing the Unseen. IEEE Computer, February 1998, pp. 26-34.

[12]. Kelley J. Terrorist instructions hidden online, USA Today – 2001.
[13]. Mitra S., T.Roy, D.Mazumadar, A.Saha. Steganalysis of LSB encoding in uncompressed images by close color pair analysys. CDAC, Kolkata
[14]. Seymer P., Dimitoglou G., Performance Optimization of Close-Color Pair Steganalysis, Proceedings of the 2007 International Conference on Security & Management, pp. 123-127, Las Vegas, USA. 2007.
[15]. Stuart Fox. How Russian spies hid secret codes in online photos. Tech News, the Christian Science Monitor. 2010.

REFERENCES

[1]. Bobok I.I., Kobozeva A.A. Steganalysis as a special case of the analysis of the information system, Modern special-purpose machinery, 2011, No 1, pp. 25-36.
[2]. Bobok I.I. Steganalysis algorithm for the basic message stored in the lossy formats, Visnyk of national technical university «HPI», 2012, No 29, pp. 41-49.
[3]. Gonsales R., Vuds R.. Digital image processing, M.: Techno sphere, 2005, 1072 p.
[4]. Gribunin V.G., Okov I.N., Turintsev I.V. Digital steganography, M.: Solon-Press, 2002, 272 p.
[5]. Konahovich G.F., Puzirenko A.U. Computer steganography. Theory and Practice, MK-Press, 2006.
[6]. Rudnitskiy V.N., Uzun I.A.Steganalysis algorithm based on the analysis of color pairs, Computer and mathematical methods in modeling, 2012, T.2, No 3, pp. 210-220.
[7]. Shvidchenko I.V. Analysis of the crypto-steganography algorithms, Problems of management and informatics, 2007, No 4, pp. 149-155.
[8]. Fridrich J., Du R., Long M. Steganalysis of LSB encoding in Color images, IEEE International Conference on Multimedia and Expo, vol.3, 2000, pp. 1279-1282.
[9]. Geetha S., Silva S.Sivatha Sindhu, Kamaraj N. Close color pair signature ensemble adaptive threshold based steganalysis for LSB embedding in digital images. Transactions on Data Privacy, 1(2009) 140-161.
[10]. Gul G., Kurugollu F. SVD-Based Universal Spatial Domain Image Steganalysis / IEEE Transactions on Information Forensics and Security, 2010, Vol. 5, No. 2, pp. 349-353.
[11]. Johnson Neil F., Jajodia Sushil. Steganography: Seeing the Unseen. IEEE Computer, February 1998, pp.26-34.
[12]. Kelley J. Terrorist instructions hidden online, USA Today, 2001.
[13]. Mitra S., T.Roy, D.Mazumadar, A.Saha. Steganalysis of LSB encoding in uncompressed images by close color pair analysis. CDAC, Kolkata
[14]. Seymer P., Dimitoglou G., Performance Optimization of Close-Color Pair Steganalysis, Proceedings of the 2007 International Conference on Security & Management, pp. 123-127, Las Vegas, USA. 2007.
[15]. Stuart Fox. How Russian spies hid secret codes in online photos. Tech News, the Christian Science Monitor. 2010.

СТЕГАНОАНАЛІТИЧНИЙ АЛГОРИТМ ДЛЯ ЗОБРАЖЕНЬ, ЯКІ ПІДДАВАЛИСЯ ОПЕРАЦІЇ СТИСНЕННЯ З ВТРАТАМИ

Легкість у застосуванні, а також маса програмних засобів як платних, так і безкоштовних, вільно розповсюджуваних по мережі, зробили стеганографію дуже популярним інструментом, за рахунок якого можна забезпечити простий спосіб організації витоку цінної інформації і неконтрольований обмін інформацією в протизаконних цілях. Дані обставини змушують активізувати зусилля з розробки алгоритмів стеганоаналізу. Одним із видів таких алгоритмів є алгоритми, що базуються на аналізі пар кольорів цифрового зображення. Більшість існуючих подібних засобів орієнтована на роботу із зображеннями, що зберігаються у форматах без втрат, що значно звужує сферу їх застосування. З урахуванням цього, в роботі був запропонований стеганоаналітичний алгоритм визначення наявності секретного повідомлення, зануреного в цифрове зображення, збережене в форматі з втратами (JPEG), за допомогою методу модифікації найменшого значущого біта. Отримані результати показали, що принцип аналізу пар кольорів може бути успішно застосований і для стеганоаналізу цифрових зображень, які піддавалися операції стиснення з втратами.

Ключові слова: стеганографія, стеганоаналіз, близькі пари кольорів, унікальні пари кольорів, приховування інформації.

STEGANALYSIS ALGORITHM FOR IMAGES THAT HAVE BEEN LOSSY COMPRESSED

Easy to use, and also a lot of software tools both paid and free, freely distributed on network, made steganography very popular tool by which you can provide a simple way of diversion of valuable information and uncontrolled exchange of information for illegal purposes. These circumstances make greater efforts to develop algorithms for steganalysis. One type of such algorithms are algorithms that are based on the analysis of color pairs of digital images. Most of the existing similar means focused on working with images that are stored in a lossless format, which significantly narrows the scope of their application. Taking this into account in the steganalysis algorithm has been proposed for determining presence a secret message embedded into digital image which has been lossy compressed, using method of least significant bit. The results showed that the principle of the analysis color pairs can be successfully applied for steganalysis digital image is subjected to a lossy compression.

Index Terms: Steganography, steganalysis, close-color pairs, unique colors, information hiding.