

УДК 004.056.5

И.А. Узун,
аспирант Одесского национального
политехнического университета

СТЕГАНОАНАЛИТИЧЕСКИЙ АЛГОРИТМ, ОСНОВАННЫЙ НА АНАЛИЗЕ ПАР ЦВЕТОВ

Статья посвящена стеганоаналитическому алгоритму определения наличия секретного сообщения, погруженного в цифровое изображение, хранимое в несжатом виде. Алгоритм основан на анализе пар цветов с использованием метода модификации наименьшего значащего бита.

Ключевые слова: сообщение, алгоритм, изображение, анализ пар цветов.

Статтю присвячено стеганоаналітичному алгоритму визначення наявності секретного повідомлення, зануреного в цифрове зображення, яке зберігається в незжатому вигляді. Алгоритм заснований на аналізі пар кольорів з використанням методу модифікації найменшого значущого біту.

Ключові слова: повідомлення, алгоритм, зображення, аналіз пар кольорів.

Paper is devoted to the steganalysis algorithm of the determination of a secret report presence, submerged into the digital image, storable in the uncompressed kind. An algorithm is based on the analysis of the pairs of colors with the use of a method of modification of the least meaning bit.

Keywords: report, algorithm, image, analysis of the pairs of colors.

Повышенный интерес к стеганографии [1; 2] – науке о скрытой передаче информации путем сохранения в тайне самого факта передачи, привел к появлению огромного количества программных продуктов (Steganos, StegHide, S-tools и др.), многие из которых свободно распространяются в интернете. Используя данные стеганографические средства, можно в любой мультимедиа-объект (цифровое изображение (ЦИ), видео или аудио) внедрить дополнительную или секретную информацию (ДИ) таким образом, что это не приведет к каким-либо заметным изменениям этих объектов. Процесс погружения ДИ в контейнер или основное сообщение (ОС) будем называть стеганопреобразованием (СП), а результат погружения – стеганосообщением (СС). ДИ, вносимая в контейнер, может быть предварительно зашифрована, чтобы усложнить задачу стеганоаналитика [1]. Основная задача стеганоанализа (СА) [1; 2] – установление факта присутствия в контейнере скрытой информации.

Одним из широко используемых методов стеганографии для изображений является метод LSB (least significant bit), смысл которого заключается в перезаписи младших бит ОС [1]. Часто в качестве контейнера используются разного рода графические файлы. Визуально изображение при СП не изменяется, особенно если в качестве ОС выбрано многоцветное изображение с большим количеством деталей, то есть информационно нагруженное. Если, например, взять ЦИ в формате BMP, 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 статистические методы стегоанализа. К таким методам можно отнести алгоритмы, основанные на анализе пар цветов [5–9]. Проводя анализ в данном направлении, было замечено, что в существующих методах есть разногласия в определении понятий уникальных цветов ЦИ [6–7]. Помимо этого, некоторые методы [7; 8] для проведения детектирования наличия ДИ в анализируемом ЦИ требуют выполнение предварительной классификации ЦИ по категориям с целью определения пороговых значений для каждой категории, причем целесообразность предлагаемой авторами классификации является мало убедительной и не выдерживает критики при желании повторить описываемые в работах вычислительные эксперименты (рис. 1). Большинство проанализированных методов применяется лишь в изображениях с количеством уникальных [5; 6] пикселей, не превышающим половину от общего количества пикселей ЦИ.

Каждый день во всемирной паутине на сайтах и в социальных сетях миллионами публикуются фотографии, которые могут служить каналом передачи скрытой информации. Подобный способ передачи секретной информации не привлекает к себе внимания и может быть легко использован криминальными структурами и террористами [4]. Поэтому развитие методов стеганоанализа на сегодняшний день чрезвычайно актуальная задача.

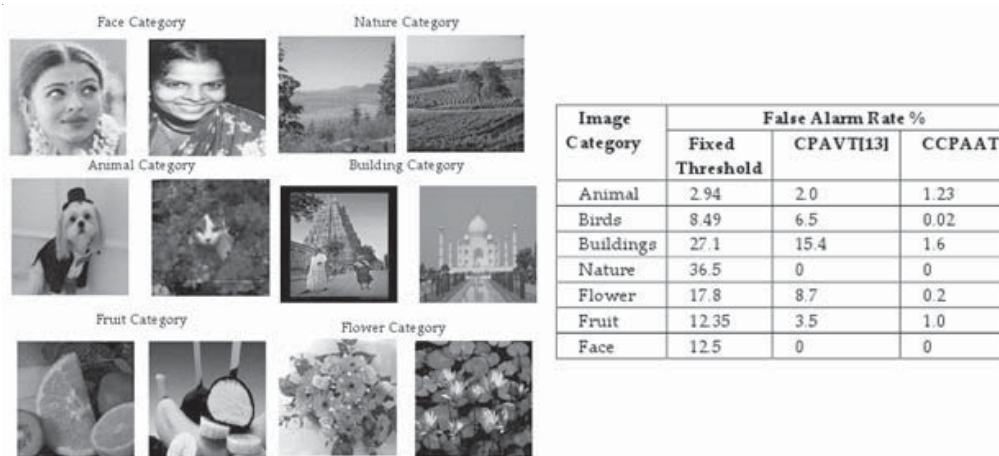


Рис. 1. Классификация изображений и определение пороговых значений для категорий классификации

Целью исследования является разработка стеганоаналитического алгоритма (САА), основанного на анализе количества близких пар цветов и уникальных цветов. Особенностью и ценностью предлагаемого метода является то, что для анализа требуется лишь изображение, которое нужно проверить на факт содержания ДИ. При этом нет необходимости в информации ни о примененном методе внедрения ДИ, ни о размере внедренного сообщения. Кроме этого предлагаемое решение не требует предварительной классификации изображений по категориям и последующего определения пороговых значений для каждой категории (рис.1), как это требуется в подобных алгоритмах [7; 8]. Данная ситуация является наиболее часто возникающей на практике. Областью применимости предлагаемого метода являются изображения с количеством уникальных пиксе-

лей, перевищуючим половину загального числа пікселів ЦІ (в якості таких ЦІ були взяті зображення в форматі ВМР, не підвергавшись операції стиснення), що також відділяє його від існуючих аналогів. В результаті роботи алгоритм повинен дати відповідь, чи є аналізоване ЦІ стеганоповідомленням чи не містить жодної вкраденої секретної інформації.

Для досягнення поставленої мети необхідно вирішити наступні *задачі*:

1. Визначити статистичні характеристики, що використовуються для перевірки ОС і СС, аналіз яких дозволить відокремити ЦІ, підвергавшись СП, від ЦІ, що не містять ДІ;

2. Визначити поняття унікального кольору в силу існуючих суперечностей по відношенню до даного визначення в різних методах [6; 7];

3. Виявити характерні особливості і відмінності вихідних ЦІ, що не підвергавшись СП, від СС – отриманих після впровадження ДІ в ході СП за допомогою модифікації НЗБ;

4. Виявити характерні особливості і відмінності зображень, що вже підвергавшись СП за допомогою модифікації НЗБ, від СС, отриманих після повторного СП, т.е. вже змінені бітами випадкового ЦІ, НЗБ аналізованого зображення, повторно модифікуються в ході СП бітами іншого випадкового ЦІ;

5. Виходячи з результатів рішення попередніх завдань, розробити САА для виявлення СС, отриманих при використанні методу LSB над зображеннями в форматі ВМР.

Введемо необхідні позначення і визначення. Під кольором будемо розуміти в подальшому трійку компонентів (R, G, B) або піксель, який також розуміється як триплет значень (R, G, B) , де R – червоний, G – зелений і B – синій компоненти в кольоровій моделі RGB [3].

Буквою P позначимо кількість близьких пар кольорів в зображенні. При цьому дві кольори (R_1, G_1, B_1) і (R_2, G_2, B_2) будемо називати близькими [5], якщо для них справедливо наступне співвідношення:

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

Другою статистичною величиною, що використовується при аналізі, є кількість унікальних кольорів. В відношенню до унікальних кольорів зображення існують декілька підходів. В одному випадку [5; 6; 9] під унікальними кольорами розуміють просто всі унікальні пікселі (т.е. різні триплети значень (R, G, B) ЦІ). Данна величина позначається буквою U . Цей підхід застосовують, коли аналізується ЦІ, в якій кількість унікальних пікселів не перевищує 50 % всіх пікселів [5]. Кількість всіх пар унікальних кольорів, представляється як кількість комбінацій з U по 2:

$$C_2^U \geq P$$

Будемо позначати R – відношення між кількістю близьких пар кольорів і кількістю унікальних пар кольорів:

$$R = \frac{P}{C_2^U} \quad (1)$$

Согласно второму подходу [7, 8] в определении понятия уникального цвета, два цвета (R_1, G_1, B_1) и (R_2, G_2, B_2) называются уникальными, если выполняется, хотя бы одно из условий:

$$|R_1 - R_2| \leq 1,$$

$$|G_1 - G_2| \leq 1,$$

$$|B_1 - B_2| \leq 1.$$

При этом R определяется как отношение количества близких пар цветов к количеству уникальных цветов U :

$$R = \frac{P}{U} \quad (2)$$

Как уже говорилось ранее, в ЦИ в формате BMP каждый цвет состоит из трех компонент, на каждую из которых отведено по 8 бит. Таким образом, в каждом цвете НЗБ содержатся в каждой компоненте триплета значений (R, G, B) . Тогда, можно сказать, что два цвета составляют близкую пару цветов, если они отличаются между собой лишь наименьшими значащими битами. Данное определение было введено исследователями Джонсоном и Яджоди [1], которые указали, что после внедрения ДИ в ЦИ методом LSB полученная цветовая палитра будет иметь отличительную особенность – большое количество пар близких цветов. При этом ими предполагалось, что данный тип артефакта можно наблюдать только в изображениях в градациях серого.

Позже данными особенностями заинтересовались и другие исследователи [6], которыми было замечено, что число уникальных цветов в ЦИ обычно значительно меньше общего числа пикселей ЦИ. Отношение числа уникальных цветов к количеству пикселей составляет примерно от 1:2 для изображений, хранящихся в форматах без потерь (TIF, BMP), до 1:6, или даже ниже, для JPEG изображений. Количество уникальных цветов в JPEG изображениях тем меньше, чем выше параметр сжатия. Данное наблюдение очень важно, поскольку означает, что большинство ЦИ имеет небольшое количество уникальных цветов, относительно общего количества пикселей ЦИ. Таким образом, было показано, что появление большого количества близких пар цветов закономерно и для полноценных цветных изображений.

Рассмотрим данную особенность на примере (рис. 2). В исходное ЦИ, не содержащее внедрений, встраивается секретное сообщение методом модификации НЗБ. Как видно, после внедрения ДИ в “чистое” изображение количество близких пар цветов увеличилось значительно.



Рис. 2. Внедрение сообщения в изображение, не содержащее ДИ

Если же внедрение ДИ происходит в СС, то при этом не происходит значительных изменений в количестве близких пар цветов (рис. 3).

Отношение R (как для первого (1), так и для второго подхода (2)) дает идею относительно близких пар цветов в ЦИ. Так, после внедрения число уникальных цветов в ЦИ возрастет до U' , и мы можем посчитать число близких пар цветов P' и отношение R' для полученного СС. Таким образом, идеей является то, что для изображения, которое не содержит секретного сообщения, число близких пар цветов относительно показателя всех возможных пар цветов будет значительно меньше, чем для изображения, которое уже содержит в себе внедренную ДИ. Однако оказывается, что практически невозможно найти порог для отношения R для всех изображений в силу большой вариации уникальных цветов U . Данную проблему помогло решить наблюдение, позволяющее различать изображения, содержащие и не содержащие ДИ. Так, можно заметить, что если изображение уже содержит в себе внедренную информацию, то внедрение другого сообщения в него не изменит отношение R значительно. С другой стороны, если изображение является чистым, то показатель R значительно возрастет после внедрения.

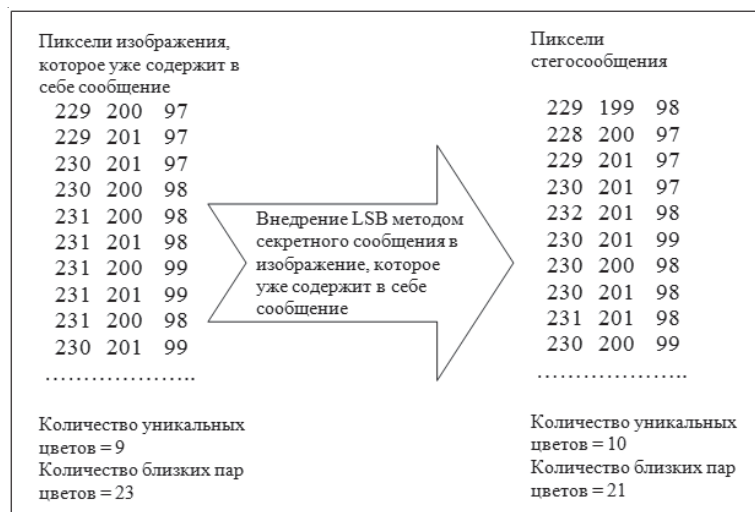


Рис. 3. Внедрение сообщения в изображение, которое уже содержит ДИ

На основаниі даних спостережень Д. Фридрих [6] запропонував САА, згідно з яким аналізуване зображення піддається введенню додаткової інформації. Щоб відповісти на питання, чи є початкове зображення СС чи ні, аналізуються показники R і R' . Якщо $R \cong R'$, то робиться висновок, що аналізуване зображення містило додаткову інформацію, в протилежному випадку якщо $R' > R$, то вважається, що аналізуване зображення було чистим. Подібний вид мають і алгоритми, засновані на (2), які крім того ще використовують порогові значення категорій класифікації ЦІ.

В даній роботі розглядаються зображення в форматі BMP, які не піддавалися операції стиснення. Було обрано 100 зображень розміром 300 на 200 пікселів, які були конвертовані в формат BMP з TIF. За допомогою LSB методу було отримано ще 100 СС. Як додаткову інформацію вбудували ЦІ. При аналізі кількості унікальних кольорів, було встановлено, що у 76 % аналізованих чистих зображень унікальними є більше 50 % пікселів. А даний факт свідчить про те, що запропоновані вище алгоритми не застосовні до аналізованих зображень, оскільки кількість унікальних пікселів занадто велика.

Згідно з розрахунками, отриманими при аналізі чистих ЦІ, було помічено (рис. 4), що якщо ЦІ не піддавалося стисненню і кількість унікальних кольорів в зображенні перевищує 50 % (або близько до 50 %) від загальної кількості пікселів, то це ЦІ вже містить значуще число близьких пар кольорів. І при введенні додаткової інформації методом LSB даний показник буде наполегливо зменшуватися (від P до P_1 на рис. 4). При цьому зменшується і кількість унікальних кольорів, однак з набагато меншою швидкістю (від U до U_1).

Якщо розглянути аналогічні значення (рис. 5) для зображень, в які було попередньо вбудовано додаткову інформацію, то тут немає такої однозначної тенденції в зміні кількості близьких пар кольорів або унікальних кольорів. Зміни коефіцієнтів P і U відбуваються як в бік збільшення, так і зменшення.

Як було сказано раніше, в запропонованому методі було вирішено для отримання нових статистичних даних проводити над аналізованим ЦІ стеганообробку, тобто спочатку необхідно отримати початкові дані кількості унікальних кольорів і близьких пар кольорів (P і U) в початковому зображенні, а потім ці показники (P_1 і U_1) повинні враховуватися в СС, отриманих за допомогою введення в НЗБ початкового зображення біт випадкового ЦІ. Над кожним зображенням з бази (100 чистих і 100 стегоповідомлень) було виконано по 15 вбудовувань.

Контейнер	%U	Сообщение	%U	P	P1	U	U1
1602_P5120003.tif.bmp'	0.66'	4450_P5120029.tif.bmp'	0.82'	195709	93630	39365	32072
1777_P5120005.tif.bmp'	0.91'	2528_P5120011.tif.bmp'	0.69'	112253	48134	54475	44287
1049_P4044437.tif.bmp'	0.44'	1294_P5120001.tif.bmp'	0.65'	84624	21392	26446	18793
1711_P5120004.tif.bmp'	0.57'	3557_P5120021.tif.bmp'	0.56'	230628	22151	34376	14091
1128_P4184478.tif.bmp'	0.54'	2558_P5120012.tif.bmp'	0.6'	172122	24267	32384	20635
1131_P4184478.tif.bmp'	0.43'	1039_P4044437.tif.bmp'	0.37'	116064	10925	25594	14698
1216_P5120001.tif.bmp'	0.51'	3559_P5120021.tif.bmp'	0.65'	184969	31121	30879	16554
1232_P5120001.tif.bmp'	0.59'	937_P4044417.tif.bmp'	0.19'	236299	17802	35530	16308
1239_P5120001.tif.bmp'	0.53'	2493_P5120011.tif.bmp'	0.64'	231519	24688	31906	14339
1249_P5120001.tif.bmp'	0.49'	1783_P5120005.tif.bmp'	0.86'	164392	30166	29632	18140
1458_P5120002.tif.bmp'	0.59'	1341_P5120002.tif.bmp'	0.8'	139570	23563	35316	25979
1341_P5120002.tif.bmp'	0.8'	2495_P5120011.tif.bmp'	0.73'	169857	68248	47938	34915
1571_P5120003.tif.bmp'	0.74'	2376_P5120010.tif.bmp'	0.73'	192956	90038	44371	34219
1502_P5120002.tif.bmp'	0.65'	2592_P5120012.tif.bmp'	0.67'	148926	28057	39111	26674

Рис. 4. Аналіз показників близьких пар і унікальних кольорів для чистих зображень

Контейнер	%U	Сообщение	%U	P	P1	U	U1
1039_P4044437.tif_with_1bit_2163_P5120008.tif.bmp'	0.25'	4115_P5120026.tif.bmp'	0.69'	31749	29879	14972	14997
103_P3104346.tif_with_1bit_2163_P5120008.tif.bmp'	0.13'	1128_P4184478.tif.bmp'	0.54'	26765	9719	7881	5794
1049_P4044437.tif_with_1bit_2495_P5120011.tif.bmp'	0.33'	2528_P5120011.tif.bmp'	0.69'	27863	24055	19939	20127
1128_P4184478.tif_with_1bit_2817_P5120014.tif.bmp'	0.33'	2984_P5120016.tif.bmp'	0.5'	34737	47357	19982	21719
1131_P4184478.tif_with_1bit_4461_P5120029.tif.bmp'	0.33'	1458_P5120002.tif.bmp'	0.59'	38141	13810	19515	15055
1216_P5120001.tif_with_1bit_2302_P5120009.tif.bmp'	0.22'	1470_P5120002.tif.bmp'	0.47'	1460	18856	12965	15109
1232_P5120001.tif_with_1bit_1274_P5120001.tif.bmp'	0.21'	2302_P5120009.tif.bmp'	0.63'	1000	1286	12698	12822
1239_P5120001.tif_with_1bit_1216_P5120001.tif.bmp'	0.24'	252_P3114356.tif.bmp'	0.31'	36701	29819	14665	15269
1261_P5120001.tif_with_1bit_1335_P5120002.tif.bmp'	0.19'	2558_P5120012.tif.bmp'	0.6'	7444	15698	11686	14686
1272_P5120001.tif_with_1bit_192_P3114354.tif.bmp'	0.22'	1783_P5120005.tif.bmp'	0.86'	12808	39837	13336	18660
1274_P5120001.tif_with_1bit_1847_P5120005.tif.bmp'	0.32'	1701_P5120004.tif.bmp'	0.68'	63973	19397	19376	15463
1294_P5120001.tif_with_1bit_103_P3104346.tif.bmp'	0.32'	1701_P5120004.tif.bmp'	0.68'	3169	19887	18923	25023

Рис. 5. Анализ показателей близких пар и уникальных цветов для стеганосообщений

Данные особенности будет удобнее рассматривать на графиках (рис. 6), где по оси ординат – значения коэффициентов R (отношение числа близких пар цветов к количеству уникальных цветов), а по оси абсцисс – количество модифицируемых НЗБ.

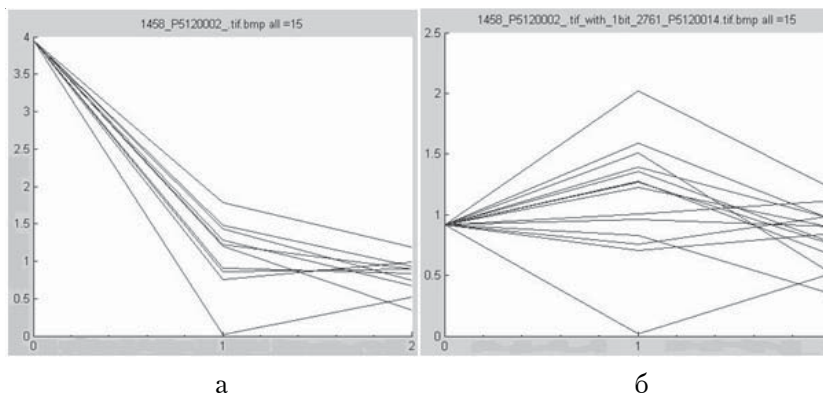
Анализ подобных графиков для всех тестируемых изображений обратил внимание на различия в разбросе коэффициентов R_1 в чистых изображениях и стеганосообщениях (рис. 6): практически во всех анализируемых изображениях разброс значений R_1 в СС превышает аналогичный показатель, полученный для ЦИ, не содержащих изначально внедренной ДИ. Поэтому в качестве порогового было взято значение разности между максимальным значением коэффициента R_1 и R :

$$\max(R_1) - R$$

Для чистых изображений $\max(R_1) - R < 0$, а для стеганосообщений – $\max(R_1) - R \geq 0$.

На основании проделанных исследований предлагается ССА, основными шагами которого являются:

1. Вычисление количеств близких пар цветов (P) и уникальных цветов (U) в анализируемом изображении, а также коэффициента R , равного их отношению;
2. Внедрение случайного изображения в анализируемое ЦИ с использованием LSB-преобразования с последующим подсчетом коэффициентов P_1 , U_1 и R_1 для полученного СС. Данная процедура повторяется 15 раз с различными случайными изображениями.
3. Определение максимального значения в массиве из 15 значений R_1 , полученных на 2 шаге.
4. *Результат.* Если величина $(\max(R_1) - R) < 0$, то анализируемое ЦИ не подвергалось стегопреобразованию, иначе – анализируемое изображение являлось стеганосообщением.



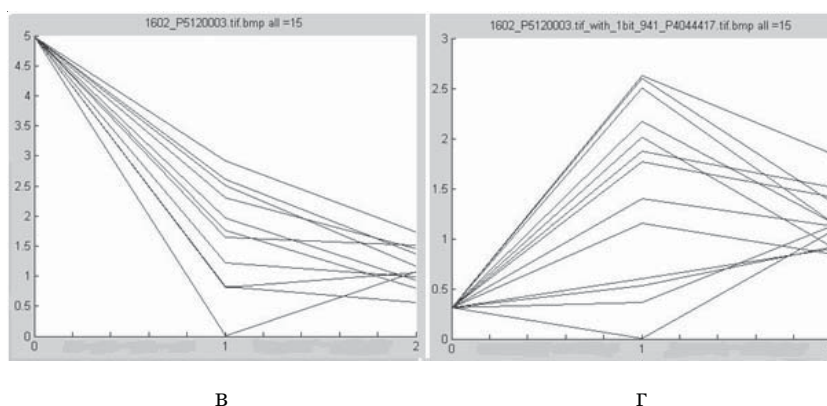


Рис. 6. Графики поведения коэффициентов R при стеганопреобразованиях:
а, в – над ЦИ, не содержащими ДИ; б, г – над СС.

Предлагаемый алгоритм был опробован на 100 чистых ЦИ и 100 стегосообщениях в формате BMP размером 300 на 200 пикселей. При этом ошибка первого рода, когда ЦИ изображение было ошибочно принято за стегосообщение, составила 1 %, а ошибка второго рода составила 5 %.

Выводы

Разработанный алгоритм дал возможность эффективного использования статистических значений близких пар цветов и уникальных цветов в ЦИ с количеством уникальных пикселей, превышающим 50% всех пикселей изображения, что никогда не делалось ранее.

В данный момент исследования автора направлены на изучение эффективности алгоритма при анализе СС с уменьшением количества погруженной в них информации, а также совершенствование порогового значения.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Грибунин В.Г. Цифровая стеганография / В.Г. Грибунин, И.Н. Оков, И.В. Туринцев. – М. : Солон-Пресс, 2002. – 272 с.
2. Конахович Г.Ф. Компьютерная стеганография. Теория и практика / Г.Ф. Конахович, А.Ю. Пузыренко. – М. : МК-Пресс, 2006 г.
3. Гонсалес Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1072 с.
4. Бобок И.И. Стеганоанализ как частный случай анализа информационной системы / И.И. Бобок, А.А. Кобозева // Сучасна спеціальна техніка. – 2011. – № 1. – С. 25–36.
5. Johnson, Neil F. and Jajodia, Sushil. Steganography : Seeing the Unseen. IEEE Computer, February, 1998. – pp. 26–34.
6. Fridrich J. Steganalysis of LSB encoding in Color images / J.Fridrich, R. Du, M. Long ; IEEE International Conference on Multimedia and Expo. – vol. 3. – 2000. – pp. 1279 – 1282.
7. Geetha S. Close color pair signature ensemble adaptive threshold based steganalysis for LSB embedding in digital images / S.Geetha, Silva S.Sivatha Sindhu, N. Kamaraj // Transactions on Data Privacy. – № 1(2009). – p. 140–161.
8. Mitra S. Steganalysis of LSB encoding in uncompressed images by close color pair analysis / S.Mitra, T.Roy, D.Mazumadar, A.Saha ; CDAC, Kolkata.
9. Seymer P. Performance Optimization of Close-Color Pair Steganalysis / Seymer P., Dimitoglou G. ; Proceedings of the 2007 International Conference on Security & Management, Las Vegas, USA, 2007. – pp. 123–127.

Отримано 07.03.2012