

СТЕГАНОАНАЛИТИЧЕСКИЙ АЛГОРИТМ ДЕТЕКТИРОВАНИЯ МЕТОДА МОДИФИКАЦИИ НАИМЕНЬШЕГО ЗНАЧАЩЕГО БИТА ДЛЯ КОНТЕЙНЕРОВ, ХРАНИМЫХ В ФОРМАТАХ БЕЗ ПОТЕРЬ

В.Н. Рудницкий¹, И.А. Узун²

¹ Черкасский государственный технологический университет,
бул. Шевченко, 460, Черкассы, 18006, Украина; e-mail: rvn_2008@ukr.net

² Одесский национальный политехнический университет,
просп. Шевченко, 1, Одесса, 65044, Украина

Работа посвящена разработке стеганоаналитического алгоритма определения наличия секретного сообщения, погруженного в цифровое изображение, хранимое в формате без потерь. Алгоритм основан на анализе пар цветов с использованием метода модификации наименьшего значащего бита.

Ключевые слова: стеганоанализ, цифровое изображение, метод модификации наименьшего значащего бита

Введение

Повышенный интерес к стеганографии [1, 2] - науке о скрытой передаче информации путём сохранения в тайне самого факта передачи, обусловленный запретом или ограничением на законодательном уровне во многих странах мира использования шифрования после трагических событий 11 сентября 2011 года, 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 стеганообразованием (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.

Каждый день во всемирной паутине на сайтах и 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.

Одним из широко используемых методов стеганографии для изображений является метод модификации наименьших значащих битов – LSB (*least significant bit*) [1], «долгожевучесть» и популярность которого обусловлена простотой его реализации,

а также возможностью обеспечения при его применении значительной скрытой пропускной способности с одновременным соблюдением надежности восприятия соответствующего СС [1], 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] (рис. 1). 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 эти методы, являясь достаточно эффективными, обладают несогласованностью 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. Большинство проанализированных методов применяется лишь в изображениях с количеством уникальных [5, 6] пикселей, не превышающим половину от общего количества пикселей ЦИ.



а

б

Рис. 1. Результат работы метода модификации наименьшего значащего бита: а – изображение-контейнер; б – стеганосообщение

Цель статьи и постановка исследований

Целью работы является повышение эффективности процесса стеганоанализа путем разработки стеганоаналитического алгоритма (САА) детектирования метода LSB, 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 о размере внедренного сообщения. Кроме этого предлагаемое решение не требует предварительной классификации изображений по категориям и последующего определения пороговых значений для каждой категории,

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 пикселей, превышающим половину общего числа пикселей ЦИ (в качестве таких ЦИ были взяты изображения в формате BMP, не подвергавшиеся операции сжатия), что также отличает его от существующих аналогов. В результате работы алгоритм должен дать ответ, является ли анализируемое ЦИ стеганосообщением, или не содержит никакой внедренной секретной информации.

Для достижения поставленной цели необходимо решить следующие задачи:

1) Определить понятие уникального цвета в силу существующих противоречий по поводу данного определения в различных вариациях метода [6, 7];

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;

4) Исходя из результатов решения предыдущих задач, разработать САА для выявления СС, полученных при использовании метода LSB к изображениям в формате BMP.

Основная часть

Введем необходимые обозначения и определения. Под цветом будем понимать в дальнейшем 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].

Буквой P обозначим число близких пар цветов в изображении. При этом два цвета (R_1, G_1, B_1) и (R_2, G_2, B_2) будем называть близкими [5], если для них справедливо следующее соотношение:

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

Второй статистической величиной, используемой при анализе, является количество уникальных цветов. В отношении уникальных цветов изображения существует несколько подходов. В одном случае [5, 6, 9] под уникальными цветами понимаются просто все уникальные пиксели (т.е. различные триплеты значений (R, G, B)) ЦИ. Данная величина обозначается буквой U . Данный подход применяют, когда анализируется ЦИ, в котором количество уникальных пикселей не превышает 50% всех пикселей [5]. Количество всех пар уникальных цветов, представляется как количество сочетаний из U по 2:

$$C_2^U \geq P$$

Будем обозначать R – отношение между числом близких пар цветов и количеством уникальных пар цветов:

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

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.

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.

На основании данных наблюдений Д. Фридрих [6] был предложен САА, согласно которому анализируемое изображение подвергается внедрению дополнительной информации. Чтобы ответить на вопрос, являлось ли исходное изображение СС или нет, анализируются показатели R и R' . Если $R \cong R'$ – то делается вывод, что анализируемое изображение содержало дополнительную информацию, в противном

случае если $R' > R$, то считается, что анализируемое изображение было чистым. Подобный вид имеют и алгоритмы, основанные на (2), которые помимо всего еще используют пороговые значения категорий классификации ЦИ.

В настоящей работе рассматриваются изображения в формате BMP, которые не подвергались операции сжатия. Было отобрано 100 изображений размером 300 на 200 пикселей, которые были конвертированы в формат BMP из TIF. Посредством LSB метода было получено еще 100 СС. В качестве дополнительной информации внедрялись Ц. При анализе количества уникальных цветов, было установлено, что у 76% анализируемых чистых изображений уникальными являются более 50% пикселей. А данный факт свидетельствует о том, что приведенные выше алгоритмы не применим к анализируемым изображениям, поскольку количество уникальных пикселей слишком велико.

Согласно вычислениям, полученным при анализе чистых ЦИ, было замечено, что если ЦИ не подвергалось сжатию и количество уникальных цветов в изображении превышает 50% (или близко к 50%) от общего количества пикселей, то это ЦИ уже содержит значительное число близких пар цветов. И при внедрении дополнительной информации LSB методом, данный показатель будет стремительно уменьшаться (от P к P_1). При этом уменьшается и количество уникальных цветов, однако с гораздо меньшей скоростью (от U к U_1).

Если рассмотреть аналогичные значения для изображений, в которые была предварительно внедрена дополнительная информация, то здесь нет такой однозначной тенденции в изменении количества близких пар цветов или уникальных цветов. Изменения коэффициентов P и U происходят как в сторону увеличения, так и уменьшения.

Как уже было сказано ранее, в предлагаемом методе было решено для получения новых статистических данных проводить над анализируемым ЦИ стеганопреобразования, т.е. сначала необходимо получить исходные данные количеств уникальных цветов и близких пар цветов (P и U) в исходном изображении, а потом эти же показатели (P_1 и U_1) должны считываться в СС, получаемых посредством внедрения в НЗБ исходного изображения бит случайного ЦИ. Над каждым изображением из базы (100 чистых и 100 стеганосообщений) производилось по 15 внедрений.

Данные особенности будет удобно рассматривать на графиках (рис. 2).

Анализ подобных графиков для всех тестируемых изображений обратил внимание на различия в разбросе коэффициентов R_1 в чистых изображениях и стеганосообщениях (рис. 2): 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 контейнеров

$$\max(R_1) - R < 0,$$

а для стеганосообщений –

$$\max(R_1) - R \geq 0.$$

На основании проделанных исследований предлагается САА, 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

$$\max(R_1) - R < 0,$$

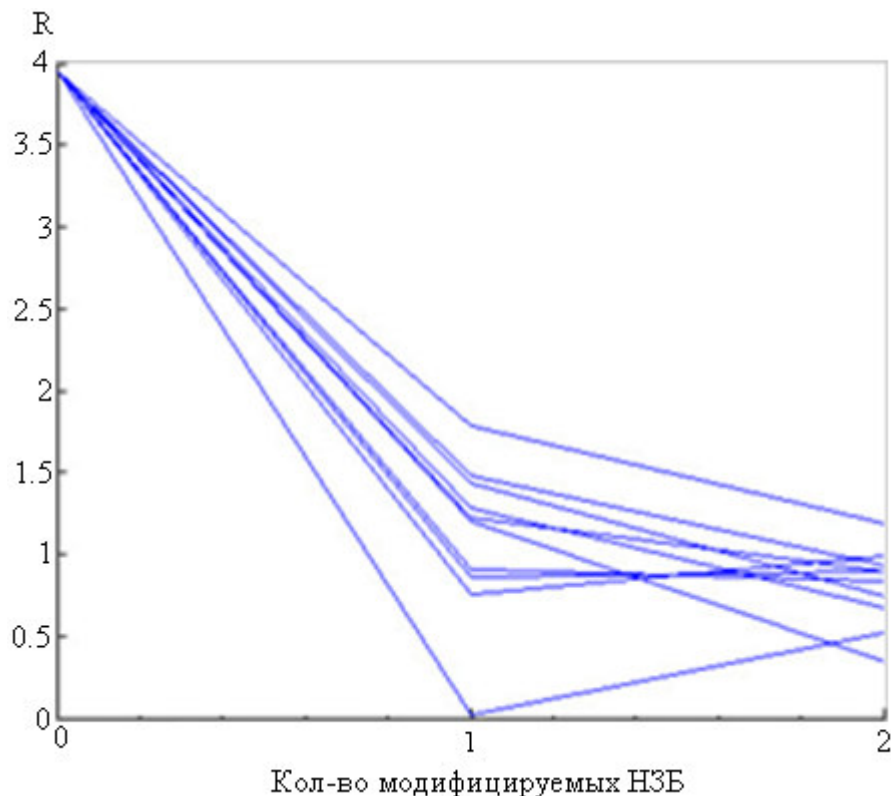
то

анализируемое ЦИ не подвергалось стегопреобразованию,

иначе

анализируемое изображение являлось стегосообщением.

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



а

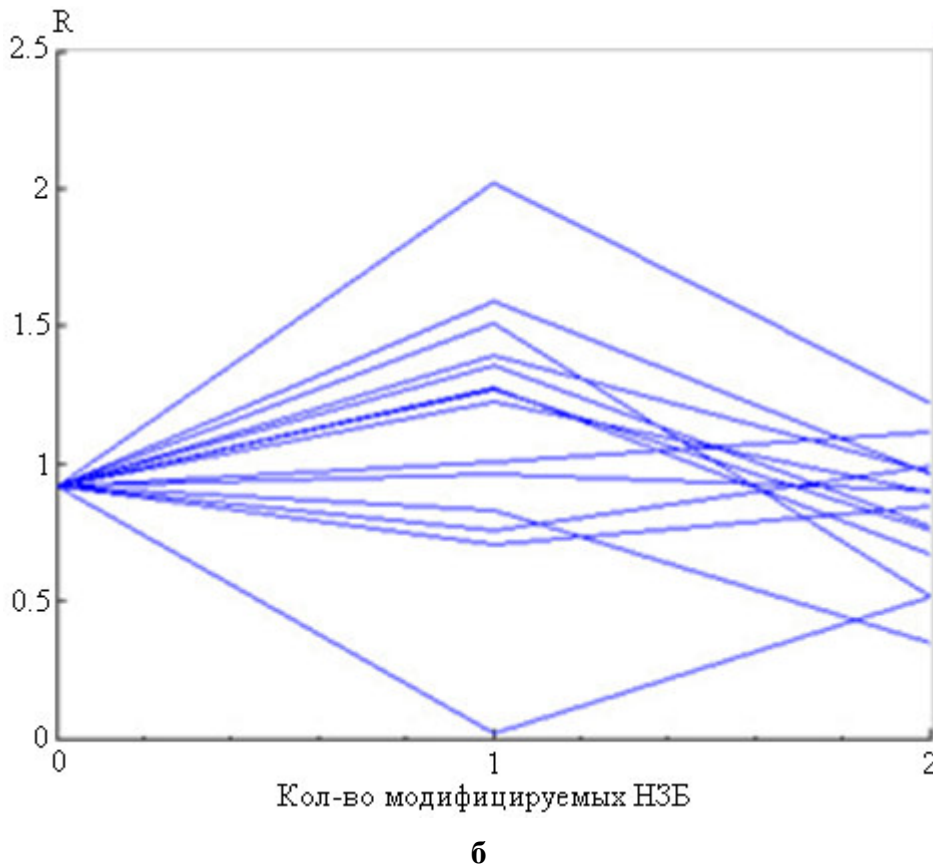


Рис. 2. Графики поведения коэффициентов R при стеганопреобразованиях: над ЦИ, не содержащими ДИ (а); над СС (б)

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 авторов направлены на изучение эффективности алгоритма при анализе СС с уменьшением количества погруженной в них информации, поскольку в настоящий момент рассматривалась скрытая пропускная способность в LSB 1 бит/пиксель, а также совершенствование порогового значения.

Список литературы

1. Грибунин, В.Г. Цифровая стеганография [Текст] : монография / В.Г. Грибунин, И.Н. Оков, И.В. Туринцев. — М. : СОЛОН-Пресс, 2002. — 272 с.
2. Конахович, Г.Ф. Компьютерная стеганография [Текст] : теория и практика / Г.Ф. Конахович, А.Ю. Пузыренко. — Киев : МК-Пресс, 2006. — 283 с.
3. Гонсалес, Р.С. Цифровая обработка изображений [Текст] : научное издание / Р.С. Гонсалес, Р. Вудс ; Пер. с англ. П.А. Чочиа. — М. : Техносфера, 2005. — 1070 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, N.F. Steganography: Seeing the Unseen / N.F. Johnson, S. Jajodia // IEEE Computer. — 1998. — Vol.31, Iss.2. — PP. 26–34.

6. Friedrich, J. Steganalysis of LSB Encoding in Color Image / J. Friedrich, R. Du, M. Long // IEEE International Conference on Multimedia and Expo. — 2000. — Vol.3. — PP. 1279–1282.
7. Geetha, S. Close color pair signature ensemble adaptive threshold based steganalysis for LSB embedding in digital images / S. Geetha, S. Sindhu, and N. Kamaraj // Transactions on Data Privacy. — 2008. — Vol.1, Iss.3. — PP. 140–161.
8. Mitra, S. Steganalysis of LSB Encoding in Uncompressed Images by Close Color Pair Analysis / S. Mitra, T. Roy, D. Mazumdar and A.B. Saha // IIT Kanpur Hackers' Workshop 2004 (ИТКНАСК04), 23–24 Feb 2004. — 2004. — PP. 23–24.
9. Seymer, P. Performance Optimization of Close-Color Pair Steganalysis / P. Seymer, G. Dimitoglou // Proceedings of the 2007 International Conference on Security & Management, Las Vegas, USA. — 2007. — PP. 123–127.

СТЕГАНОАНАЛІТИЧНИЙ АЛГОРИТМ ДЕТЕКТУВАННЯ МЕТОДУ МОДИФІКАЦІЇ НАЙМЕНШОГО ЗНАЧУЩОГО БІТА ДЛЯ КОНТЕЙНЕРІВ, ЗБЕРЕЖЕНИХ У ФОРМАТАХ БЕЗ ВТРАТ

В.М. Рудницький¹, І.О. Узун²

¹ Черкаський державний технологічний університет
бул. Шевченка, 460, Черкаси, 18006, Україна; e-mail: rvn_2008@ukr.net

² Одеський національний політехнічний університет,
просп. Шевченка, 1, Одеса, 65044, Україна

Робота присвячена розробці стеганоаналітичного алгоритму визначення наявності секретного повідомлення, вбудованого в цифрове зображення, яке зберігається в форматі без втрат. Алгоритм заснований на аналізі пар кольорів з використанням методу модифікації найменшого значущого біту.

Ключові слова: стеганоаналіз, цифрове зображення, метод модифікації найменшого значущого біта

METHOD OF LSB STEGANALYSIS FOR LOSSLESS COVER-IMAGES

Vladimir N. Rudnitsky¹, Ija A. Uzun²

¹ Cherkassy State Technological University
460 Shevchenko str., Cherkassy, 18006, Ukraine; e-mail: rvn_2008@ukr.net

² Odessa National Polytechnic University,
1 Shevchenko Ave., Odessa, 65044, Ukraine

This paper is devoted to steganalysis algorithm determining presence a secret message embedded into digital image which is stored in uncompressed form. The algorithm is based on the analysis of pairs of colors and uses the method of modifying the least significant bit.

Keywords: steganalysis, digital images, Least Significant Bit