УДК 621.397.3

КОМБИНИРОВАННЫЙ СТЕГАНОГРАФИЧЕСКИЙ АЛГОРИТМ ВСТРАИВАНИЯ КОНФИДЕНЦИАЛЬНОЙ ИНФОРМАЦИИ В ЦИФРОВЫЕ ИЗОБРАЖЕНИЯ ФОРМАТА JPEG

STEGANOGRAPHIC COMBINATION ALGORITHM OF EMBEDDING THE CONFIDENTIAL INFORMATION INTO THE JPEG DIGITAL IMAGES

С.В. Радаев, Д.В. Орлов, О.О. Басов
S.V. Radaev, D.V. Orlov, O.O. Basov

Федеральное государственное казённое военное образовательное учреждение высшего образования «Академия Федеральной службы охраны Российской Федерации», Россия, 302034, г. Орёл, ул. Приборостроительная, д. 35

Federal state military educational institution of higher professional education "Academy of the Federal security service of the Russian Federation", 35 Priborostroitelnaya St, Orel, 302034, Russia

E-mail: radik0782@mail.ru, dimaorlov56@gmail.com, oobasov@mail.ru

Аннотация
Представлен анализ организации файловой структуры формата JPEG с точки зрения стеганографического контейнера. Особенностью организации файловой структуры формата является то, что JPEG может быть рассмотрен с точки зрения встраивания информации форматным и неформатным методами на основе стеганографического программного продукта JPHide и, непосредственно, добавлением неосновных маркеров цифрового изображения JPEG, в том числе и искусственным путём при их отсутствии. В форматную часть целесообразно встраивать хэш-код, в свою очередь, неформатная часть является более подходящей для встраивания смысловой нагрузки сообщения. Предложен подход, идея которого заключается в комбинации форматных и неформатных методов стеганографии с целью осуществления процедуры аутентификации передаваемой информации. Таким образом, в результате информационного взаимодействия будет обеспечиваться не только скрытая передача конфиденциальной информации, но и существует возможность проверки целостности стеганографического сообщения путём сравнения извлечённого хэш-кода и вычисленного хэш-кода от полученного стегособщения.

Abstract
The analysis of organization of file structure of JPEG format from the point of view of steganographic container is presented. The peculiarity of the organization of the file structure is that JPEG can be considered from the point of view of embedding information in format and non-format methods based on the steganographic software product JPHide and, directly, adding non-main markers of the digital JPEG image, including artificially in the absence thereof. In the format part, it is advisable to embed a hash code, in turn, the non-format part is more suitable for embedding information. The proposed approach, which consists in the combination of formatted and unformatted methods of steganography to implement the authentication procedure of information to be transmitted. Thus, as a result of information interaction will not only hidden transfer of confidential information, but it is possible to check the integrity of the steganographic message by comparing the extracted hash and computed hash from the received hidden message.

Ключевые слова: стеганография, встраивание, защита информации, конфиденциальная информация.
Keywords: steganography, embedding, information protection, confidential information.

В настоящее время всё чаще используются различные технологии скрытия процесса информационного взаимодействия с целью маскирования информации при её последующей передаче по открытым каналам связи. Маскирование в этом случае служит для обеспечения скрытности сообщения, содержащегося в носителе (контейнере), и выполняется стеганографическими методами [Pfitzmann, 1996]. При этом всё большую актуальность наряду с конфиденциальностью приобретает обеспечение целостности передаваемых по незащищённым каналам данных. При этом в качестве основного метода защиты информации от нелегитимных пользователей при организации информационного взаимодействия применяется криптографическая защита, основанная на гарантированной стойкости современных систем шифрования.

Одним из ее перспективных направлений является разработка стеганографических методов и средств передачи закрытой информации. Это обусловлено, во-первых, стремительным развитием вычислительной техники, во-вторых, тем, что ограничения, накладываемые в большинстве стран на криптографические системы (передача ключей государству, регистрация, лицензирование и др.), не распространяются на стеганографические средства [Постановление Правительства РФ, 2012].

К настоящему времени нормативно-правовая база в области использования стеганографических технологий не разработана, что позволяет бесконтрольно использовать методы и средства стеганографии для организации скрытых каналов передачи данных, что и обуславливает перспективность направления разработки методов средств и для скрытой передачи информации с ограниченным доступом [Грибунин, 2002; Аграновский, 2003].

В качестве носителей для скрытой передачи информации чаще всего используются данные мультимедийного характера. При этом большинство известных стеганографических методов скрывают данные в графических изображениях по причине наибольшей распространённости последних в сети Интернет. В свою очередь, среди различных графических форматов большую популярность у пользователей получил формат JPEG (Joint Photographic Experts Group – объединённая группа экспертов по фотографическим изображениям) в связи с небольшим размером файлов, получаемым за счёт использования сжатия с потерями. Вместе с тем использование сжатия с потерями в контейнерах формата JPEG [Wallace, 1991] является причиной возникновения ряда трудностей, ограничивающих возможности сокрытия данных. Поскольку из контейнеров удаляется практически вся избыточная для восприятия информация, данные, скрываемые в контейнере после такого удаления, могут вызывать визуальные искажения [Жиляков, 2011]. Если же скрывать данные до этапа удаления избыточности, появляется опасность нарушения целостности встроенных данных и, как следствие, некорректного их извлечения.

Достаточно большое число существующих стеганографических программных продуктов вносит данные в служебные поля контейнеров (или дописывает в конец файла). Исходя из чего, в зависимости от используемой структуры формата файла все методы стеганографии подразделяются на форматные и неформатные [Alturki, 2001; Farid, 2001].

Разработка неформатных методов стеганографии базируется на модификации параметров, кодирующих непосредственно данные самого изображения [Жиляков, 2014]. При этом различают методы сокрытия в файлах, использующих сжатие без потерь, и в файлах, предусматривающих сжатие с потерями.

Использование форматных методов стеганографии для встраивания дополнительной информации в мультимедийные данные представляет большой интерес вследствие простоты их реализации, идея которой основывается на использовании служебных полей (заголовки, флаги, идентификаторы, маркеры, комментарии, неиспользуемые биты в палитре цветов, косвенные данные) в структуре файла, изменение которых не сказывается на визуальном качестве изображений.

В настоящее время в свободном доступе имеется достаточно большое количество стеганографических программных продуктов. По данным Интернет-сайта

http://www.jjtc.com_количество таких программных продуктов составляет 146 единиц, из них четвёртая часть работает с изображениями формата JPEG (Invisible Secret 4.0., JpegX, Puff v.101, Image Hide, Hide and Seek, Image Hide, SecurEngine 4.0 и др.).

В статье проведён стеганографический анализ модифицированного цифрового изображения (ЦИ) JPEG на примере портативного стеганографического программного приложения JPHide, которое можно скачать с Интернет-сайта http://linux01.gwdg.de/~alatham/stego.htm.

В результате проведённого эксперимента выявлено, что встраивание по алгоритму JPHide осуществляется неформатным методом непосредственно в данные, кодирующие само ЦИ. Суть эксперимента заключалась в следующем.

1. С помощью программного продукта XnView для Windows (версия 1.98.1) было смоделировано 24-битное ЦИ *.jpg размером 50×50 пикселей.

2. Полученное ЦИ *.jpg было загружено в приложение JPHide. После чего смоделировано два ЦИ: со встроенной информацией (модифицированное, размер встроенного стегосообщения, представленного файлом в формате *.txt, равен 6 байтам) и пустое (исходное).

3. В программной среде Matlab (R2012a) исходное и модифицированное ЦИ были декомпозированы на цветовые составляющие (R, G и B), представленные в виде массивов целочисленных значений интенсивностей пикселей в диапазоне от 0 до 255;

4. Была определена разница между соответствующими значениями интенсивностей пикселей цветовых компонент R (красная), G (зелёная) и B (синяя) исходного и модифицированного ЦИ. Результат представлен на рисунке 1.

Как видно из рисунка, модификации подверглись значения интенсивностей трёх цветовых составляющих, причём синяя компонента вместила больше всего информации (1803 пикселя). При этом общее число модифицированных значений интенсивностей цветовых компонент оказалось равным 2832.

Встраивание по алгоритму программы JPHide осуществляется методом LSB (Least Significant Bits – наименее значимые биты) [Hursev, 2004; Chandramouli, 2001; Fridrich, 2004], т. е. по аналогии с большинством стегопрограмм, поддерживающих формат BMP. Следует заметить, что модификации подверглись также младшие пиксели 6 разряда, что, в свою очередь, является демаскирующим признаком, так как происходит нарушение статистики распределения младших бит изображения [Fridrich, 2007].

Таким образом, можно сделать вывод, что стеганографическая стойкость встроенной по алгоритму JPHide информации является недостаточной. Кроме того, гарантия целостности встроенной информации отсутствует. Следовательно, целесообразно воспользоваться особенностями гибкой архитектуры организации файлового формата JPEG [Recommendation T.81, 1993] и разработать комбинированный алгоритм встраивания дополнительной информации, позволяющий на приёмной стороне при извлечении стегосообщения сделать вывод о состоянии целостности переданной информации. При этом предлагается содержательную часть стегосообщения встраивать в частотную область ЦИ [Жиляков, 2016], а значение хэш-кода, вычисленного от встроенного сообщения, внедрять в форматную часть структуры файла.

Согласно спецификации T.81 формат JPEG состоит из упорядоченного набора параметров и маркеров, описывающих сжатые данные. Параметры и маркеры в свою очередь образуют сегменты.

Файл JPEG содержит последовательность маркеров, каждый из которых начинается с байта 0xFF, свидетельствующего о начале маркера, и байта-идентификатора. Некоторые маркеры состоят только из этой пары байтов, другие же содержат дополнительный данные, состоящие из двухбайтового поля с длиной информационной части маркера (включая длину этого поля, но за вычетом двух байтов начала маркера т.е. 0xFF и идентификатора) и собственно данных.

Рис. 1. Результат стегоанализа модифицированного ЦИ формата JPEG
Fig. 1. The result of stegoanalysis of the modified DI JPEG format

Маркеры служат для идентификации различных структурных частей формата JPEG (рис.2).

Идентификатор	Длина	Данные маркера

Рис. 2. Структура маркера
Fig. 2. The structure of the marker

Идентификатором являются два байта, обязательно в формате 0xFFC0, по которым можно идентифицировать тип маркера.

Длина так же, как и идентификатор, состоит из двух байт, значение которых складывается из длины данной секции и длины данных маркера в байтах (в обратном порядке). Нужно отметить, что не все маркеры имеют длину (например, маркеры TEM, RST0…RST7, SOI, EOI не содержат значение длины).

Данные маркера – набор байт, которые требуют обработки в соответствии с типом маркера.

При этом должны соблюдаться следующие правила размещения маркеров в jpg-файле:

– файл всегда начинается с маркера SOI и заканчивается маркером EOI;

– если данные из одного маркера нужны для обработки второго маркера, первый маркер должен располагаться до второго;

– сжатые данные компонентов не встречаются внутри маркера. Они всегда следуют сразу после маркера SOS. Т.к. в нем нет информации о длине, необходимо просканировать данные до следующего маркера (отличного от RSTN), чтобы найти конец сжатых данных без их восстановления;

– маркеры RSTN встречаются внутри сжатых данных, но не встречаются в маркерах.

Наипростейшим с точки зрения реализации вариантом встраивания является метод дописывания стегосообщения в конец файла. Модификация файла формата JPEG производилась в редакторе файлов Hex Workshop v*6.8*. Результат встраивания показан на рисунке 3 (текст стегосообщения отмечен черным цветом). Следует отметить, что, несмотря на простоту реализации, этот метод встраивания имеет крайне низкую стеганографическую стойкость.

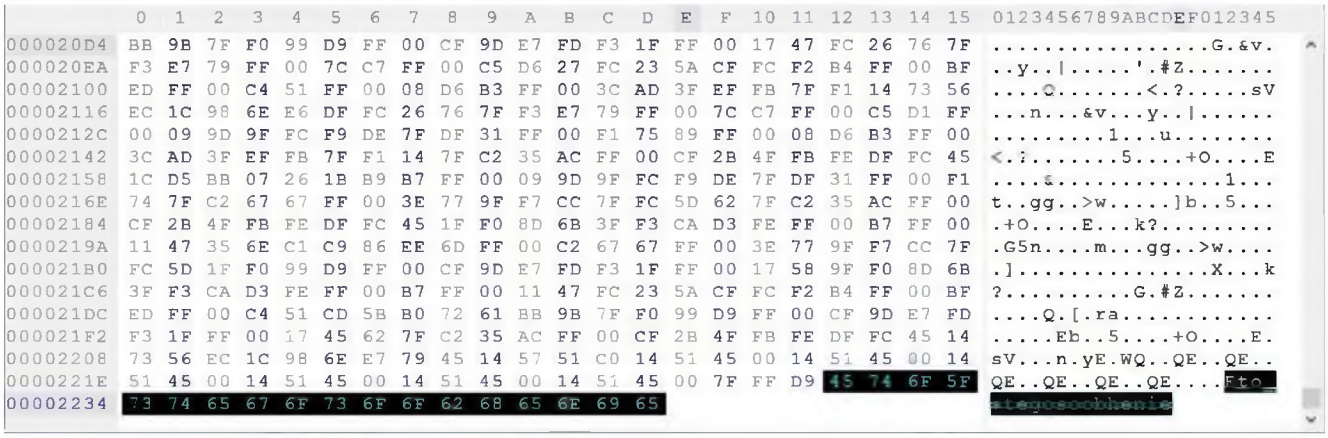

Рис. 3. Встраивание методом дописывания в конец файла

Fig. 3. Embedding by appending to the end of the file

Кроме того встроить стегосообщение можно непосредственно в необязательный маркер изображения, предварительно его добавив. Для этого необходимо выполнить следующее:

1. Создать новый маркер в изображении, которого ранее в нем не находилось, не являющийся обязательным и не оказывающий влияние на процесс декодирования данных об интенсивностях пикселей ЦИ. Например: маркер со значением FFC8 является таковым.

2. В новом маркере после его значения следует записать само стегосообщение. На рисунке 4 представлено встроенное в маркер FFC8 стегосообщение размером 32 байта (выделено чёрным цветом).

Следует отметить, что в случае некорректной модификации значений полей маркеров в большинстве случаев кодер будет идентифицировать файл изображения как неизвестный, что будет являться демаскирующим признаком.

Таким образом, в форматной части реализовано встраивание значение 256-битного хэш-кода, вычисленного по алгоритму криптографической хэш-функции [ГОСТ Р 34.11–2012].

В свою очередь, встраивание содержательной части стегосообщения предлагается осуществлять в DC-коэффициенты частотной области изображения.

Данный выбор обусловлен тем, что при модификации AC-коэффициентов величина ошибки линейного предсказания значений соседних пикселей изменяется сильнее, чем при модификации DC-коэффициентов. Выбор DC-коэффициентов в качестве скрывающих позволяет сохранять статистические соотношения между пикселями внутри

одного блока (блок коэффициентов размером 8x8). Поскольку DC-коэффициенты дискретного косинусного преобразования (ДКП) могут рассматриваться в качестве средних значений соответствующих блоков пикселей, матрица DC-коэффициентов обладает всеми свойствами изображения и может быть представлена как исходное изображение меньшего разрешения. DC-коэффициент представляет собой значение точечной статистики от значений пикселей блока. Из этого следует, что сохранение статистических соотношений между DC-коэффициентами позволит сохранить статистические соотношения между блоками в целом, а значит, и во всём изображении. Так как DC-коэффициенты ДКП могут восприниматься как уменьшенная копия изображения, является целесообразным применение к ним пространственных методов сокрытия, в частности, основанных на вейвлет-преобразовании (ВП) [Ching-Yu Yang, 2012.]. Использование базовых пространственных методов сокрытия приводит к заметным визуальным искажениям. С другой стороны, сокрытие информации в вейвлет-коэффициентах вызывает гораздо меньшее искажение контейнера [Sakkara, 2012].

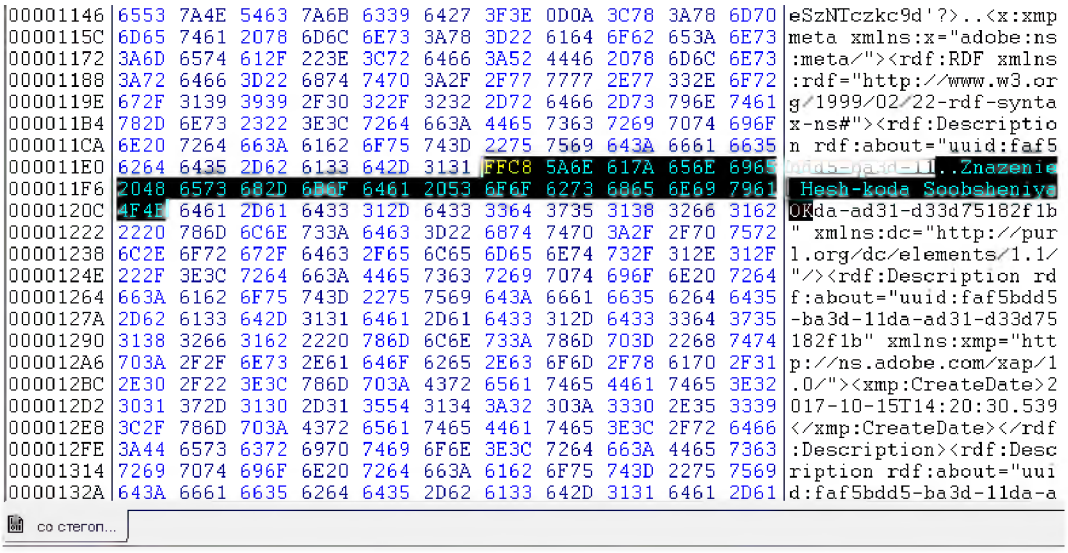

Рис. 4. Встраивание методом добавления необязательного маркера в изображение
Fig. 4. Embedding by adding an optional marker to the image

Один из классов ВП, допускающих минимум ошибок декодирования, являются целочисленные ВП. Для безошибочного извлечения сообщения после сокрытия в области ВП над контейнером (носителем) не должно производиться никаких других преобразований, допускающих удаление избыточности (в частности, цветоразностное преобразование, квантование и ДКП).

Предложен следующий алгоритм встраивания стегосообщения:

– осуществляется декодирование битового потока jpg-файла до этапа получения матрицы коэффициентов ДКП после процедуры квантования;

– DC-коэффициенты ДКП представляются в виде уменьшенной копии изображения, т. е. в виде матрицы;

– к матрице DC-коэффициентов применяется целочисленное дискретное вейвлет-преобразование (ДВП) [Taubman, 2002];

– осуществляется встраивание бит стегосообщения путём модификации наименее значимых бит коэффициентов ДВП;

– над модифицированными коэффициентами ДВП осуществляется обратное целочисленное ДВП;

– над полученной матрицей коэффициентов ДКП осуществляют дальнейшие преобразования согласно алгоритму JPEG до получения структуры битового потока jpg-файла.

При использовании лифтинг-схемы реализация целочисленного ДВП осуществляется с использованием следующих выражений [Adams, 2002]:

$$Y(2n+1) = X_{ext}(2n+1) - \left\lfloor \frac{X_{ext}(2n) + X_{ext}(2n+2)}{2} \right\rfloor, \quad (1)$$

$$Y(2n) = X_{ext}(2n) + \left\lfloor \frac{Y(2n-1) + Y(2n+1)}{4} \right\rfloor, \quad (2)$$

где $Y(2n+1)$ и $Y(2n)$ - соответственно нечётные и чётные выходящие значения яркости пикселей (значения вейвлет-коэффициентов); $X_{ext}(2n+1)$ и $X_{ext}(2n)$ – соответственно нечётные и чётные входящие значения яркости пикселей (индекс *ext* означает симметричное расширение значений яркости пикселей на границах изображения);

Таким образом, в результате информационного взаимодействия будет обеспечиваться не только скрытая передача конфиденциальной информации, но и существует возможность проверки целостности стеганографического сообщения и контейнера путём сравнения извлечённого хэш-кода и вычисленного хэш-кода от полученного стегосообщения.

Однако следует иметь в виду, что гарантированную аутентификацию передаваемых данных могут обеспечить только сертифицированные методы и средства [ГОСТ 28147–89; ГОСТ Р 34.10–2012], поэтому выбор тех или иных способов и алгоритмов защиты информации зависит от степени важности передаваемой информации и должен учитывать возможные риски в случае её несанкционированной модификации или умышленного уничтожения.

Заключение

В большинстве развитых стран в связи с возрастающим количеством информационных преступлений на криптографические системы накладываются ограничения. В свою очередь, на современные системы маскирования и стеганографии таких ограничений нет, что говорит о перспективности развития данной тематики. Тем не менее, стоит иметь в виду, что выбор способов и алгоритмов защиты информации должен учитывать всевозможные риски при передаче конфиденциальной информации, вплоть до её умышленного уничтожения.

Список литературы
References

1. Аграновский А.В., Девянин П.Н., Хади Р.А., Черемушкин А.В. 2003. Основы компьютерной стеганографии. М., Радио и связь: 90–131.

Agranovskij A.V., Devjanin P.N., Hadi R.A., Cheremushkin A.V. 2003. The basics of digital steganography. M., Radio i svjaz: 90–131.

2. ГОСТ Р 34.11–2012. Информационная технология. Криптографическая защита информации. Функция хэширования. Дата введения 01.07.2012.

GOST R 34.11–2012. Information technology. Cryptographic protection of information. Hash function. Date of introduction 01.07.2012.

3. ГОСТ 28147–89. Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования. Дата введения 01.12.1989.

GOST 28147–89. Information processing systems. Cryptographic protection. Cryptographic transformation algorithm. Date of introduction 01.12.1989.

4. ГОСТ Р 34.10–2012. Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи. Дата введения 01.07.2012.

GOST R 34.10–2012. Information technology. Cryptographic protection of information. Processes of formation and verification of electronic digital signature. Date of introduction 01.07.2012.

5. Грибунин В.Г., Оков И.Н., Туринцев И.В. 2002. Цифровая стеганография. М., Солон-Пресс, 272.

Gribunin V.G., Okov I.N., Turincev I.V. 2002. Digital steganography. M., Solon-Press, 272.

6. Жиляков Е.Г., Черноморец А.А., Голощапова В.А.. 2011. Реализация алгоритма внедрения изображений на основе использования неинформационных частотных интервалов изображения-контейнера. Вопросы радиоэлектроники. 4(1): 96–104.

Zhilyakov E.G., Chernomorets A.A., Goloshhapova V.A. 2011. Implementation of the algorithm for introducing images based on the use of non-information frequency intervals of the image-container. Radio electronic. 4(1): 96–104.

7. Жиляков Е.Г., Черноморец А.А., Болгова Е.В., Гахова Н.Н. 2014. Исследование устойчивости стеганографии в изображениях. Научные ведомости БелгГУ. Экономика. Информатика. 29(172): 168–174.

Zhilyakov E.G., Chernomorets A.A., Bolgova E.V., Gakhova N.N. 2014. Study of steganography stability in images. Belgorod State University Scientific Bulletin. Economics Information technologies. 29(172): 168–174.

8. Жиляков Е.Г., Черноморец А.А., Болгова Е.В. 2016. Об информационных подобластях пространственных частот изображений. Научные ведомости Белгородского государственного университета. Серия «Экономика. Информатика». Белгород, Белгородский государственный национальный исследовательский университет. 23(244): 87–92.

Zhiljakov E.G., Chernomorec A.A., Bolgova E.V. 2016. On information subregions of spatial frequencies of images. Belgorod State University Scientific Bulletin. Economics Information technologies. 23(244): 87–92.

9. Постановление Правительства РФ от 16.04.2012 № 313.

Decree of government of the Russia Federation No 313, 16.04.2012

10. Adams M.D. 2002. The JPEG-2000 still image compression standard. Avaible at: www.ece.uvic.ca/mdadams.

11. Alturki F., Mersereau R. 2001. A Novel Approach for Increasing Security and Data Embedding Capacityin Images for Data Hiding Applications. Proc. of ITCC. LasVegas, Nevada.

12. Chandramouli R., Memon N. 2001. Analysis of LSB Based Image Steganography Techniques. Proceedings of ICIP. Thessaloniki, Greece.

13. Ching-Yu Yang, Yu Chih-Hung Lin, Wu-Chih Hu. 2012. Reversible Data Hiding for High-Quality Images Based on Integer Wavelet Transform. Journal of Information Hiding and Multimedia Signal Proceccing. Vol 3, Number 2: 142.

14. Farid H. 2001.Detecting Steganographic Message in Digital Images. Technical Report, TR2001-412. Dartmouth College, New Hampshire.

15. Fridrich J. Goljan M. 2004. On Estimation of Secret Message Length in LSB Steganography in Spatial Domain. EI SPIE Electronic Imaging. San Jose.

16. Fridrich J. Pevný T., Kodovský J. 2007. Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities. Proceedings of the 9-th ACM Multimedia & Security Workshop. Dallas, TX: 3–14.

17. Hursev T.S. Remkumar M., Akansu A.N. 2004. Data Hiding Fundamentals and Applications. Content Security in Digital Multimedia. ELSEVIER science and technology books, 254.

18. Pfitzmann B. 1996. Information hiding terminology, information hiding. First international workshop of lecture notes in computer science. Berlin, Springer-Verlag. Vol. 1174: 347–350.

19. Recommendation T.81. 1993. Information technology – digital compression and coding of continuous-tone still images requirements and guidelines. The International telegraph and telephone consultative committee: 186.

20. Sakkara S., Akkamahadevi D., Somashekar K. 2012. Integer Wavelet based Secret Data Hiding By Selecting Variable Bit Length. International Journal of Computer Applications. Vol. 48, Number 19: 7–11.

21. Taubman D.S., Marcellin M.W. 2002. JPEG 2000 Image compression fundamentals, standards and practice. Norwell, Massachusetts. Kluwer AcademicPublishers, 773.

22. Wallace G.K. 1991. The JPEG still picture compression standard. Commun. ACM.