УДК 004.415.24/004.56.5

Н.В. Кошкина

Стеганоанализ изображений в формате *jpeg* на базе атаки контрольным внедрением

Описаны особенности использования метода наименьшего значащего бита для внедрения информации в файлы формата *jpeg*. Стеганоаналитический метод на базе атаки контрольным внедрением адаптирован для обнаружения *jpeg*-стеганоконтейнеров. Исследована зависимость точности выявления стеганоконтейнеров от их наполненности. Получены численные оценки точности при выявлении стеганоконтейнеров, созданных программами *jsteg* и *jphide*.

The singularity of the least significant bits method for embedding the information to the JPEG files is described. The method of steganalysis which is based on control embedding attack has been adapted to detect the JPEG stego images. The dependence of accuracy of the steganocontainers detection on their steganographic capacity is examined. The numerical estimation of the detection the stego images accuracy, created with jsteg and jphide applications, is obtained.

Описано особливості використання методу найменшого значущого біту для вкраплення інформації в файли формату *jpeg*. Стеганоаналітичний метод на базі атаки контрольним вкрапленням адаптовано для виявлення *jpeg*-стеганоконтейнерів. Досліджено залежність точності виявлення стеганоконтейнерів від їх заповненості. Отримано чисельні оцінки точності при виявленні стеганоконтейнерів, створених за допомогою програм *jsteg* та *jphide*.

Введение. Доступность и простота применения современных программных продуктов для стеганографического сокрытия информации в файлах, впоследствии передаваемых на сайты мультимедиа-контента и файлообменные серверы, позволяет скрытно организовывать и координировать проведение различного рода противоправных действий. Поэтому особенно актуальной становится задача обнаружения скрытой информации в изображениях, аудиосигналах и других объектах, типичных для цифровых сред.

В работах [1–4] исследован метод стеганоанализа на основе атаки контрольным внедрением. Присутствие межэлементной и внутриэлементной корреляции аудиосигналов и изображений позволяет выявлять наличие стегановложений в этих типах контейнеров статистическими методами. В частности, если для сокрытия данных используется метод *наименьшего значащего бита* (НЗБ) и сообщение перед внедрением было зашифровано, то младшие биты элементов контейнера, обладающие своей, «естественной», статистикой, будут заменены равномерно распределенной битовой последовательностью. Хотя такая модификация мультимедийных файлов и не влияет на их функциональность, однако она приводит к ощутимому изменению статистических характеристик создаваемых стеганоконтейнеров относительно их прототипов (для тех их частей, которые служат непосредственными носителями битов сообщения). Повторное и все дальнейшие стеганопреобразования этих контейнеров с помощью того же программного продукта незначительно изменят статистику в сравнении с первым [1]. С учетом этой закономерности был построен метод стеганоанализа на основе различий в изменении статистических характеристик пустых и заполненных контейнеров после применения к ним атаки контрольным внедрением. Метод предполагает формирование восьмимерного характеристического вектора каждого подлежащего проверке контейнера и дальнейшую *Support Vector Mashine* (*SVM*)-классификацию [5–6] полученных векторов. Элементами характеристического вектора выступают разницы значений математического ожидания, дисперсии, асимметрии и эксцесса составляющих самого контейнера и ошибки его линейного предсказания [7].

Для некоторых элементов характеристического вектора их значения монотонно увеличиваются (уменьшаются) с увеличением длины сообщения, используемого во время атаки контрольным внедрением (наличие таких элементов можно проверить эмпирическим путем). Аналитически доказуемо, что если использовать во

время контрольного внедрения сообщения максимально возможной длины, то при любых первично скрытых данных значения таких элементов для пустых контейнеров будут не меньше их значений для соответствующих заполненных. Если стеганографический программный продукт всегда использует один и тот же стеганопуть, т.е. сохраняет порядок выбора местоположений скрываемых битов, это делает целесообразным использование при контрольном внедрении сообщений относительно малой длины и облегчает стеганоанализ. Также в этом случае имеет смысл осуществлять не одну, а несколько атак с разной длиной внедряемых сообщений [3].

Постановка задачи

Большинство существующего стеганографического программного обеспечения (ПО) реализует ту или иную модификацию метода НЗБ. Как правило, данное ПО хорошо задокументировано и не нуждается в наличии специальных знаний для его использования, что делает его доступным широкой общественности и, как следствие, обуславливает первоочередность решения задачи выявления фактов его противоправного использования.

Один из возможных путей решения данной задачи состоит в применении описанного стеганоаналитического метода на базе атаки контрольным внедрением. Выполненные ранее численные эксперименты подтвердили, что он эффективен при выявлении скрытых сообщений в аудиосигналах и изображениях, хранимых в форматах без потерь [1–4]. Вопрос о применении метода для выявления стеганоконтейнеров в форматах с потерями ранее не рассматривался. Вместе с тем в настоящее время в связи со значительным увеличением объемов цифровой информации ее хранение и передача в подавляющем большинстве случаев осуществляется в сжатом состоянии. Учитывая наличие стеганографического ПО, позволяющего скрывать информацию в сжатых с потерями звуковых или графических файлах, задача их стеганоанализа актуальна.

Анализ существующего стеганографического программного обеспечения [8] показал, что среди форматов с потерями как контейнеры для сокрытия информации чаще всего используются файлы формата *jpeg*. Поэтому было решено остановиться на более детальном изучении именно этого формата.

Цель данной статьи – адаптация метода на базе атаки контрольным внедрением для решения задачи выявления НЗБ-стегановложений в файлы формата *jpeg*, реализация адаптированной версии и получение оценок точности стеганоанализа сжатых изображений.

Сокрытие информации в файлах формата *jpeg*

Процесс сжатия данных при создании *jpeg*-файлов состоит из таких этапов:

1. *Преобразование изображения в оптимальное цветовое* пространство. Лучшая степень сжатия достигается в случае применения цветового пространства яркость/цветность, поэтому обычно изображение из *RGB* преобразуют в пространство *YCbCr*.

2. *Субдискретизация компонентов цветности усреднением групп пикселей*. Поскольку известно, что система человеческого зрения более чувствительна к компонентам яркости (*Y*), на этом этапе происходит уменьшение количества блоков данных каналов цветности (*Cb* и *Cr*) при неизменном количестве блоков *Y*. Наиболее часто используют субдискретизацию 4 : 1 : 1. Уменьшение объема данных на 50 процентов, которого при этом добиваются, практически незаметно отражается на качестве большинства изображений.

3. *Применение дискретного косинусного преобразования* (ДКП) к блокам 8×8 данных каждого компонента цветовой модели:

$$S(u,v) = \frac{2}{N} C(u)C(v) \times$$

$$\times \left[\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} s(x,y) \cos\left(\frac{\pi u(2x+1)}{2N} \right) \cos\left(\frac{\pi v(2y+1)}{2N} \right) \right],$$

где $N = 8$, $C(z) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{при } z = 0 \\ 1, & \text{при } z \neq 0 \end{cases}$.

Переход от пространственного представления к частотному позволяет перераспределить

энергию изображения так, чтобы основная ее часть содержалась в сравнительно малом количестве коэффициентов. Последние, содержащие информацию о несущественных деталях, т.е. высокочастотные, могут быть отброшены без ущерба для визуального восприятия.

4. *Квантование каждого блока коэффициентов* ДКП с применением весовых функций, оптимизированных с учетом визуального восприятия человеком. Прежде чем отбросить определенный объем информации, компрессор делит каждое выходное значение ДКП на соответствующий ему коэффициент квантования, округляя результат до целого. Именно на этом шаге происходят необратимые потери данных. Каждая из 64 позиций выходного блока имеет собственный коэффициент квантования. Чем больше коэффициент квантования, тем больше данных теряется. Высокочастотные элементы блоков ДКП квантуются с большим коэффициентом, чем низкочастотные. Кроме того, для данных яркости и цветности применяются отдельные таблицы квантования, позволяющие квантовать данные цветности с большими коэффициентами, чем данные яркости.

Пример одного из тестовых изображений, сжатого с коэффициентом качества 75, и визуализация его квантованных коэффициентов ДКП для яркостной составляющей представлены на рис. 1,*a*,*b* соответственно.

5. *Кодирование результирующих коэффициентов* с применением алгоритма Хаффмана для удаления статистической избыточности информации.

Стеганографические программы, работающие с *jpeg*-контейнерами, такие как *jsteg*, *jphide*, *Steganos Privacy Suite* 2012 (модуль *crypt&hide*) и другие, как правило, повторяют этапы 1–4, после чего они заменяют младшие биты блоков квантованных ДКП-коэффициентов битами внедряемого сообщения и для полученного результата выполняют этап 5. При этом для улучшения стойкости скрытых данных к некоторым преобразованиям стеганоконтейнера, в частности его возможному зашумлению или повторному сжатию с потерями, могут пропускаться нулевые и единичные коэффициенты. Кроме того, после квантования большинство высоко- и среднечастотных коэффициентов будут иметь нулевые значения и их изменение демаскировало бы стегановмешательство. Для усиления визуальной незаметности стеганографического вмешательства в отдельных реализациях могут пропускаться *DC*-коэффициенты, т.е. статическая компонента (рис. 1,*c*).

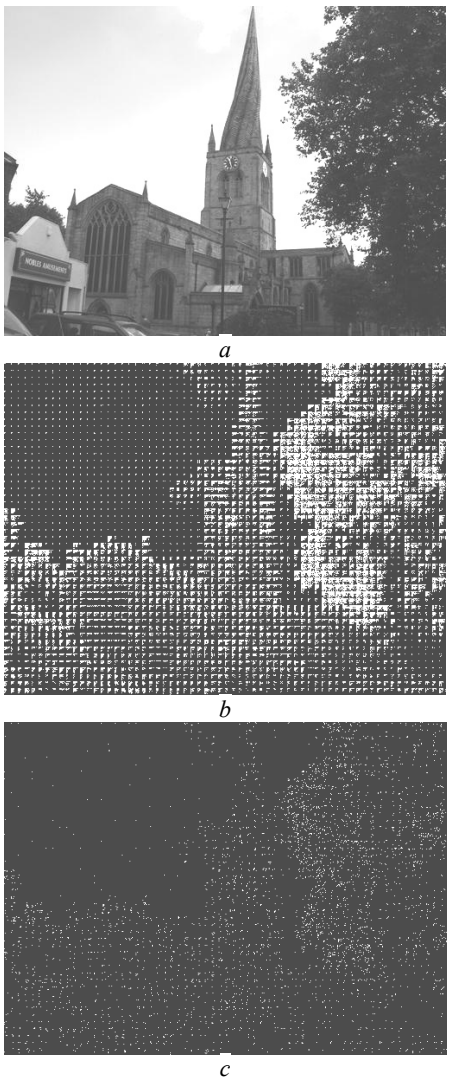

Рис. 1. Примеры: *a* – тестового изображения; *b* – графического представления квантованных ДКП-коэффициентов яркости; *c* – местоположений битов сообщения, скрытого в яркостной составляющей с помощью *Steganos Privacy Suite* 2012

Отметим, что младшие биты могут быть модифицированы для компонентов и яркости и цветности, но, учитывая усреднение и грубое

квантование данных цветности, большая часть скрываемого сообщения будет содержаться в канале яркости.

Если оценивать пропускную способность, образуемую при применении описанных вариантов стеганопреобразования, то при использовании *DC*-коэффициентов она более чем в пять раз меньше той, которую можно получить, применяя метод наименьшего значащего бита в пространственной области неупакованных изображений. Так, в цветном изображении *bmp*-формата размером 512×384 пикселя можно скрыть 72 Кб дополнительной информации. А, например, программа *jphide* после преобразования этого изображения в формат *jpeg* с коэффициентом качества 75 позволяет скрыть от двух до 13 Кб, данных (в зависимости от содержимого изображения и, как следствие, количества пригодных для внедрения коэффициентов ДКП). При отбрасывании *DC*-коэффициентов теряется возможность внедрить дополнительно количество бит, равное количеству блоков ДКП-изображения.

Насколько незаметно стегановмешательство в частотную область изображений? Если модификации затрагивали *DC*-коэффициенты, то, занявшись поиском визуальных отличий между пустыми и заполненными контейнерами при значительном их увеличении, в первую очередь можно обнаружить усиление шумовых ореолов с артефактами сжатия вокруг резких границ изображений (рис. 2,*b*). Но это усиление практически не идентифицируется без наличия оригинального контейнера. Если были модифицированы не только младшие, но и вторые биты коэффициентов ДКП, как например, это реализовано в программе *jphide*, то стегановмешательство заметно проявит себя на гладких участках контейнера (рис. 2,*c*).

Особенности метода в применении к *jpeg*-контейнерам

Даже если стегановмешательство не заметно визуально, его можно выявить статистическими методами, в частности методом на основе атаки контрольным внедрением. В отличие от варианта, направленного на выявление стегановложений в контейнерах, хранимых в форматах без потерь, т.е. *bmp*, *ppm*, *wav* и других, вариант для *jpeg*-контейнеров сравнивает разности статистик до и после контрольного стеганопреобразования не в пространственной области изображений, а в частотной – именно там, где происходит замена младших битов элементов контейнера битами секретного сообщения. В целом процесс стеганоанализа будет состоять из трех шагов:

a

b

c

Рис. 2. Увеличенные фрагменты: *a* – исходного изображения; *b* – стеганоконтейнера, созданного программой *jsteg*; *c* – стеганоконтейнера, созданного программой *jphide*

1. Формирование характеристических векторов тестируемых контейнеров.

2. Обучение *SVM*-классификатора на выборке контейнеров того же типа, что и подле-

жащие проверке, но для которых известна принадлежность к одному из классов – *пустой* или *заполненный*.

3. *SVM*-классификация контейнеров с неизвестной меткой класса.

Шаги 2 и 3 не зависят от формата контейнеров и при переходе от *bmp*-контейнеров к *jpeg* остаются без изменений [3, 9].

На первом шаге из *jpeg*-контейнера извлекаются квантованные ДКП-коэффициенты, для которых вычисляется ошибка линейного предсказания. Рассчитывается математическое ожидание, дисперсия, асимметрия и эксцесс для самих ДКП-коэффициентов и для полученной ошибки. Далее тестовый контейнер подвергается *контрольному стеганопреобразованию* (КСП) посредством определенного стеганографического программного продукта, использование которого предполагается при создании искомых аналитиком стеганоконтейнеров. Для образованного после КСП стеганоконтейнера рассчитывается математическое ожидание, дисперсия, асимметрия, эксцесс его ДКП-коэффициентов и ошибки их предсказания. Вычисляются разности соответствующих статистик до и после КСП-контейнера. Эти разности и образуют характеристический вектор. На этом шаге целесообразно учитывать также и особенности стеганопреобразования, реализованного в том или ином программном продукте. Например, программы *jsteg*, *jphide* не изменяют нулевые и единичные коэффициенты, следовательно, их можно не учитывать и при вычислении статистических характеристик.

Исследование зависимости точности стеганоанализа от наполненности стеганоконтейнеров

Для такого исследования проверяемых контейнеров с помощью пакета *Matlab* было реализовано последовательное НЗБ внедрение случайных битовых сообщений с равномерным распределением в значения квантованных ДКП-коэффициентов яркостной составляющей этих изображений. При этом из области внедрения исключались коэффициенты, равные минус единице, нулю и единице. В ходе численных экспериментов использовался набор из 1330 цветных изображений 384×512 пикселей в *jpeg*-формате, сжатых с коэффициентом качества 75. Размеры файлов с изображениями варьировались от восьми до 82 Кб. На рис. 3 отображено количество ДКП-коэффициентов, не равных минус единице, нулю или единице в яркостной составляющей *Y* каждого из тестовых контейнеров.

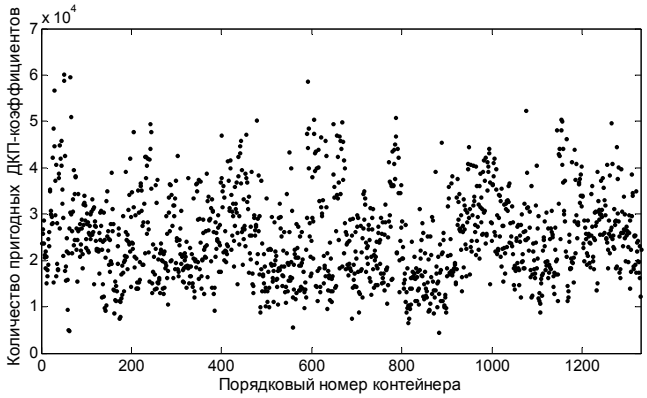

Рис. 3. Количество пригодных для внедрения ДКП-коэффициентов в яркостной составляющей каждого тестового изображения

Отметим, что количество пригодных ДКП-коэффициентов в двух каналах цветности существенно меньше. Так, для используемого набора контейнеров в среднем в компоненте *Y* можно скрыть в 10 раз больше данных, чем в *Cb* и *Cr* вместе взятых.

В ходе исследований установлено, что как и для случаев использования контейнеров в форматах без потерь, в данном случае также имеет место закономерность: чем выше наполненность стеганоконтейнера, подлежащего проверке, тем лучше точность его выявления. Так, на рис. 4 показана зависимость точности стеганоанализа от наполненности стеганоконтейнеров, полученная при использовании во время атаки контрольным внедрением случайного сообщения максимальной длины. Для создания стеганоконтейнеров использован описанный Матлаб модуль, *SVM*-классификация также осуществлялась посредством пакета Матлаб.

Значения точности усреднялись по 100 экспериментам, в каждом из которых обучающая выборка формировалась из 250 пустых и 250 стеганоконтейнеров той же наполненности, что и стеганоконтейнеры контрольной выборки. А

она в свою очередь состояла из 1080 пустых и 1080 заполненных контейнеров, не вошедших в обучающую выборку. Ядро классификатора Гаусово, а его параметры подбирались по дискретной сетке значений. Отметим, что эти условия стеганоанализа сохранялись и в дальнейших экспериментах.

Первый график на рис. 4 соответствует расчету статистики по всем абсолютным значениям квантованных ДКП-коэффициентов яркостной составляющей, второй – с исключением нулевых и единичных элементов. Как видим, второй вариант подсчета статистики дает лучшую точность.

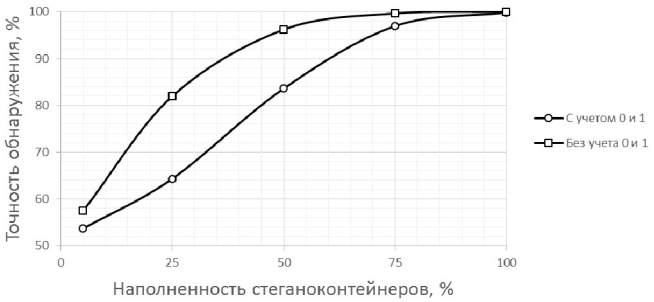

Рис. 4. Зависимость точности стеганоанализа от наполненности стеганоконтейнеров

Стеганоаналитик, как правило, не владеет информацией о длине скрытых сообщений. Предположим, что решается задача выявления стеганоконтейнеров, наполненность которых равномерно распределена в диапазоне от пяти до 100 процентов. Контрольные контейнеры можно последовательно проверить на наборе бинарных классификаторов, наученных на пустых и стеганоконтейнерах, одинаковой или близкой наполненности. В этом случае итоговое решение будет принято путем взвешенного голосования всех классификаторов.

Возможно также ускорить процесс стеганоанализа, допустив при этом некоторое ухудшение точности классификации. При выборе данного варианта используется один классификатор, обученный на пустых и стеганоконтейнерах, наполненность которых, как и в контрольной выборке, распределена в диапазоне от пяти до 100 процентов.

Второй вариант, реализованный при распознавании набора контейнеров разной наполненности, показал итоговую точность 88,12 процента при исключении нулевых и единичных коэффициентов из области сбора статистики. На рис. 5 сравниваются графики зависимости точности классификации от наполненности стеганоконтейнеров контрольной выборки для этого случая и для случая, когда наполненность стеганоконтейнеров обучающей и контрольной выборок совпадает. Как видим, присутствие в обучающей выборке контейнеров с наполненностью, отличающейся от наполненности контрольных, ухудшило точность классификации на 1–5 процентов.

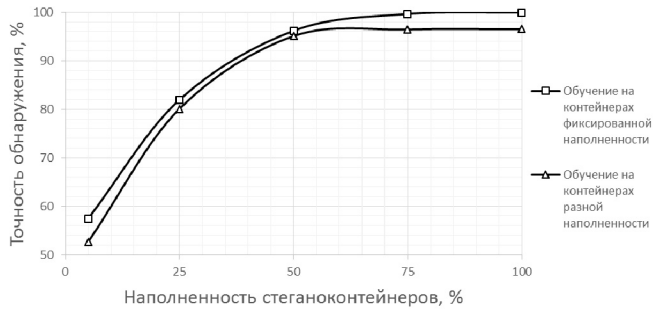

Рис. 5. Зависимость точности стеганоанализа от наполненности стеганоконтейнеров при разных вариантах обучения *SVM*

При использованных для сокрытия каналах цветности и все том же распределении наполненности стеганоконтейнеров от пяти до 100 процентов, точность стеганоанализа для этих контейнеров составила 90,66 процента.

В работе [3] показано, что при последовательном НЗБ-внедрении сообщений и во всех других случаях фиксированного стеганопути, для повышения точности стеганоанализа контейнеров следует заменить 100 процентов наполняемости контейнеров во время КСП относительно малой наполняемостью. Это минимизирует изменение статистических характеристик заполненных контейнеров после КСП и улучшает точность классификации стеганоконтейнеров относительно малой наполненности.

Описанное справедливо для НЗБ-сокрытия информации как в пространственной, так и в частотной области. Так, точность, полученная после дополнения характеристического вектора элементами, отражающими изменение ста-

тистик после КСП с 10 и пятью процентами наполняемости, при использованной только яркостной компоненте возросла с 88,12 до 92,06 процента, а с учетом всех трех (*Y*, *Cb* и *Cr*) – с 90,66 до 95,2 процента.

Стеганоанализ контейнеров, создаваемых программами *jsteg* и *jphide*

Эксперименты, подобные представленным, были реализованы для выявления стеганоконтейнеров, создаваемых программами *jsteg* и *jphide*. Анализ изменений статистики выполнялся на множестве не равных минус единице, нулю и единице квантованных ДКП-коэффициентов *Y*, *Cb* и *Cr* составляющих.

Программа *jsteg* скрывает дополнительную информацию во время перекодирования неупакованных файлов с изображениями в формат *jpeg*. Метод сокрытия – последовательное НЗБ квантованных ДКП-коэффициентов за исключением нулевых и единичных значений. Если сообщение не помещается в избранный для него контейнер, в нем скрывается только помещающаяся часть. Особенность стеганоанализа этой программы та, что перед КСП все тестируемые контейнеры должны быть перекодированы в исходный формат без потерь.

В ходе экспериментов из того же набора 1330 цветных изображений, который использовался в предыдущих тестах, был получен набор стеганоконтейнеров, где каждое изображение содержало 1 Кб скрытой информации. В зависимости от содержимого изображений и, как следствие, пригодных для внедрения ДКП-коэффициентов, наполненность стеганоконтейнеров при этом варьировалась от 10 до 100 процентов. Точность стеганоанализа, полученная при выявлении этих стеганоконтейнеров в варианте с КСП со 100 процентами наполняемости, равна 85,87 процента. В варианте, когда при КСП внедрялось сообщение меньшей длины, чем длина искомого, а именно 0,5 Кб, была получена точность 96,61 процента. Объединение значений изменений статистики, полученных после обеих атак, в один характеристический вектор привело к улучшению точности стеганоанализа до 97,41 процента.

В отличие от *jsteg*, программа *jphide* есть ключевая система. Значения ключевых элементов зависят от пароля пользователя и длины скрываемого сообщения. Для определения порядка изменения ДКП-коэффициентов *jphide* с помощью заданной таблицы делит все подходящие коэффициенты на классы для определения порядка их изменения. Первый класс в таблице составляют *DC*-коэффициенты, далее идут классы *AC*-коэффициентов, модуль которых больше задаваемого таблицей значения. НЗБ-сокрытие данных продолжается в поточном классе даже после внедрения всего сообщения. Кроме того, *jphide* может изменять не только первые, а при необходимости и вторые значащие биты коэффициентов–носителей. Если сообщение не помещается в выбранном контейнере, программа выдает ошибку и не создает стеганоконтейнер.

Тестовые *jphide*-стеганоконтейнеры были созданы из исходного набора 1330 цветных изображений путем внедрения в них 3 Кб дополнительной информации. Из-за сложностей получения большого количества 100 процентов заполненных с помощью *jphide*-стеганоконтейнеров, КСП в ходе экспериментов было организовано с помощью Матлаб-модуля, выполняющего НЗБ-внедрение во все не равные минус единице, нулю и единице ДКП-коэффициенты изображения. Итоговая точность выявления в этом случае составила 86,31 процента.

Путем использования КСП, осуществляемого программой *jphide* с паролем пользователя и длиной сообщения, совпадающими с их значениями при первичном стеганопреобразовании, был смоделирован вариант стеганоанализа в условиях известных местоположений внедренных битов. Точность стеганоанализа при этом составила 99,69 процента.

Таким образом, *jphide*-стеганоконтейнеры обладают лучшей стойкостью к выявлению данным методом в сравнении с контейнерами, создаваемыми *jsteg*. Эту стойкость в данном случае обеспечивает заложенный в *jphide* стеганоключ, определяющий местоположение модифицированных битов.

Окончание на стр. 17

Заключение. Несмотря на значительное уменьшение количества скрываемых данных при переходе от пространственной области сокрытия к частотной, стеганоаналитический метод на основе атаки контрольным внедрением остается эффективным средством выявления *jpeg*-стеганоконтейнеров при относительно большой их наполненности (больше 50 процентов). Учитывая, что использованные в данном исследовании изображения содержали чуть меньше 200 тыс. пикселей, в то время как современные цифровые фотографии состоят из нескольких миллионов, следует отметить, что для больших контейнеров метод целесообразно применять, предварительно разбив их на несколько десятков блоков. В таком случае все блоки, заполненные более чем на половину, с высокой степенью точности будут классифицированы правильно.

Как направление дальнейших исследований в первую очередь следует отметить исследование эффективности метода при выявлении *jpeg*-стеганоконтейнеров с отсутствующей модификацией *DC*-коэффициентов, а также сравнительный анализ точности данного метода с другими, направленными на выявление НЗБ-стеганографии в частотной области изображений.

1. *Ru X.*, *Zhuang Y.*, *Wu F.* Audio steganalysis based on “negative resonance phenomenon” caused by steganographic tools // J. of Zhejiang Univ. SCIENCE A. – 2006. – **7**(4). – P. 577–583.
2. *Кошкина Н.В.* Выявление в аудиосигналах скрытых сообщений, внедренных с помощью *Hide*4*PGP* // Проблемы управления и информатики. – 2013. – № 3. – С. 151–156.
3. *Кошкина Н.В.* Выявление в аудиосигналах скрытых сообщений, внедренных с помощью программы *S-Tools* // Захист інформації. – 2013. – № 4. – С. 316–326.
4. *Кошкина Н.В.* Стеганоаналіз цифрових зображень із застосуванням контрольного вкраплення // Матеріали 3 Міжнар. наук.-техн. конф. «Захист інформації і безпека інформаційних систем», 5–6 черв. 2014. – Львів: Львівська політехніка, 2014. – С. 98–100.
5. *Vapnik V.N.* Statistical Learning Theory. – New York: Wiley. – 1998. – 732 p.
6. *Vapnik V.N.* The nature of statistical learning theory. – New York: Springer-Verlag. – 2000. – 332 p.
7. *Прэтт У.* Цифровая обработка изображений. – М.: Мир, 1982. – Кн. 2 – 480 с.
8. *Steganography* software. – http://www.jjtc.com/Steganograp-hy/tools.html
9. *Кашкіна Н.В.* Стеганоаналіз МІК-стеганографії на базі матриці суміжності та методу опорних векторів // Искусственный интеллект. – 2012. – № 4.– С. 567–577.

Поступила 27.06.2014
+38 044 526-4569 (Киев)
E-mail: *nata.koshkina@gmail.com*