УДК 621.395

АЛГОРИТМЫ ВЫЯВЛЕНИЯ СТЕГАНОГРАФИЧЕСКОГО СКРЫТИЯ ИНФОРМАЦИИ В JPEG-ФАЙЛАХ

М. А. Дрюченко

Воронежский государственный университет

В работе представлены алгоритмы выявления стеганографического скрытия информации в JPEG изображениях. Рассмотрены методы анализа JPEG-файлов, основанные как на использовании статистических характеристик, так и на использовании особенностей формата JPEG, возникающих при декомпрессии.

1. ВВЕДЕНИЕ

К числу эффективных современных методов защиты информации относятся методы компьютерной стеганографии, использование которых позволяет обеспечить скрытое хранение и передачу информации.

В настоящее время стеганографические системы активно используются для решения следующих основных задач:

— защита конфиденциальной информации от несанкционированного доступа;

— преодоление систем мониторинга и управления сетевыми ресурсами;

— камуфлирование программного обеспечения;

— защита авторских прав на некоторые виды интеллектуальной собственности.

Применение компьютерной стеганографии является одним из возможных путей защиты авторских прав и противодействия пиратству.

Список стеганографических методов ежегодно пополняется, изобретают более надежные и оригинальные способы сокрытия информации, для нейтрализации которых потребуются новые, эффективные методы анализа.

Для скрытия информации в стеганографии используется т.н. *контейнер* — это любой файл или поток данных, структура и размер которого позволяют спрятать необходимые данные. Объем скрываемой с помощью стеганографии информации напрямую зависит от объема контейнера. Чем больше объем контейнера, тем больше информации можно спрятать. К числу наиболее используемых в качестве контейнеров файлов относятся изображения, текстовые, аудио- и видео-файлы.

Среди графических форматов, формат JPEG на сегодняшний день является самым распространенным. Его поддерживают почти все приложения, работа которых так или иначе связана с цифровой графикой. Файлы формата JPEG предназначены для хранения полноцветных многоградационных изображений с глубиной от 6 до 24 бит/пиксел. JPEG — это схема сжатия изображений, основанная на дискретных косинусных преобразованиях, позволяющая достичь очень высоких коэффициентов сжатия. Принципы компрессии подробно описаны в [1]. С позиции стеганографии файлы формата JPEG являются контейнерами, позволяющими надежно скрывать сравнительно большие объемы информации.

Обнаружением скрытой информации занимается *стеганоанализ*, а защита от обнаружения считается основной задачей стеганографии. Совместное использование стеганографии и криптографии на сегодняшний день обеспечивает практически непреодолимый барьер для раскрытия информации.

Стеганоанализ — новое и пока не достаточно развитое направление, хотя потребность в нем, крайне велика. Как ни странно, стеганоанализ входит в поле деятельности спецслужб ряда государств. Юридические запреты традиционно накладывают некоторые ограничения на использование криптографии, в то время как стеганография не подпадает под действие этих ограничений и является эффективным инструментом тайной передачи информации. В США ведутся исследования стегопрограмм, с целью определения их слабых мест. Компания WetStone Technologies, работающая для ВВС США, разрабатывает специальную программу «blind steganography detection prototype», которая статистическими методами вычисляет вероятность наличия спрятанного в изображении сообщения и возможную программу, которая использовалась для кодирования. С помощью

современных методов стеганоанализа успешно можно анализировать работу стеганографических программ первого поколения, новые же программы реализуют весьма совершенные алгоритмы, хорошо защищающие информацию, так что выявление факта ее скрытия крайне сложно.

Данная работа посвящена разработке программного обеспечения, способного обнаруживать скрытую в JPEG-файлах информацию или ее присутствие в файле-контейнере. В статье рассмотрены некоторые распространенные статистические методы анализа, а также предложено несколько способов анализа файлов формата JPEG, основанных как на использовании статистических характеристик, так и на использовании особенностей формата JPEG, возникающих при декомпрессии.

2. МЕТОДЫ И ПРОГРАММНЫЕ ПРОДУКТЫ СТЕГАНОГРАФИЧЕСКОГО СКРЫТИЯ ИНФОРМАЦИИ В JPEG- ФАЙЛАХ

Чтобы наметить пути разработки методов обнаружения стеганографического скрытия информации рассмотрим основные методы стего-скрытия в JPEG.

Пожалуй, самым распространенным на сегодня, но наименее стойким к обнаружению методом скрытия, является метод замены младших или менее значимых бит (LSB — Least Significant Bit метод) [2]. Цифровые изображения представляют из себя матрицу пикселов. Младший значащий бит (LSB) изображения несет в себе меньше всего информации и человек обычно не способен заметить изменение в этом бите. Поэтому его можно использовать для встраивания информации. Если модифицировать два младших бита (что также почти незаметно), то можно скрытно передать вдвое больший объем данных.

Еще один распространенный метод стего-скрытия информации основан на использовании особенностей файлов, сжатых с потерей данных (JPEG). При скрытии в файлы JPEG информация как правило прячется не в значения цветовых составляющих отдельных пикселей, а в дискретные косинусные коэффициенты. В случае с JPEG, визуально определить присутствие скрытой информации невозможно. Даже если в результате скрытия возникают незначительные искажения изображения, их можно объяснить изменением степени сжатия изображения. Метод скрытия в JPEG-файлах также более стоек к геометрическим преобразованиям.

На сегодняшний день разработан ряд программ, реализующих скрытие в файлах JPEG. Анализ таких программ показал, что некоторые из них для скрытия используют коэффициенты дискретного косинусного разложения, другие — специфичные для JPEG области (таблицы квантования, маркеры комментариев). Следует отметить, что большинство стеганографических продуктов “скрывают” информацию, просто дописывая ее в конец файла-контейнера. Очевидно, подобные программы в действительности не являются стеганографическими, так как нарушают основные принципы стеганографии — местоположение, а зачастую и сама скрытая информация может быть легко получена. Мало того, создатели некоторых программных продуктов просто обманывают пользователей, заявляя на своих сайтах о стопроцентной надежности и невозможности для третьих лиц обнаружить факт скрытия данных, сильных криптографических алгоритмах и поддержке всех известных форматов файлов. Как правило, данные программы написаны для коммерческого использования или являются условно бесплатными. Но для обычного пользователя, решившего защитить конкретную информацию, громкая реклама как правило, играет решающую роль в выборе инструмента стего-скрытия. Действительно качественными и надежными являются программы F5, Outguess, JPHide и пр., которые удовлетворяют принципам стеганографии: о недопустимости обнаружения информации или невозможности ее получения.

Псевдо-стеганографические программы такие как Camoufladge, Safe and Quick File Hide, Steganography(v1.6, v4.0), Data Stashv1.5, Fort-Knox 3.55 как правило, используют следующие алгоритмы:

– дописывание данных в конец файла. Это наиболее простой метод. Недостатки его в том, что не все форматы файлов будут корректно обрабатываться. Данный метод не удовлетворяет требованию внешней неизменности файла-контейнера и потенциальный стегоаналитик легко определит не только наличие, но и местоположение, а также длину скрытого сообщения.

– скрытие данных с использованием областей форматов файлов предназначенных для

комментариев. Зная формат файла всегда можно вычленить области комментариев и проанализировать их на наличие скрытой информации.

Обилие подобных некачественных программных продуктов свидетельствует об откровенной недобросовестности разработчиков и ограниченности пользователей, приобретающих такие программы.

3. СТЕГАНОАНАЛИЗ JPEG-КОНТЕЙНЕРОВ

Для выявления скрытой информации применяются *стеганографические атаки*. Классификация и описание основных видов атак приводятся в [2]. Возможных методов выявления стего существует достаточно много, начиная от визуального анализа, заканчивая различными статистическими методами. Большая часть статистических атак основана на известной математической модели контейнера или его части. Задача стегоаналитика состоит в выявлении различий между моделями пустого и заполненного контейнеров. Ниже рассмотрены несколько вариантов статистических атак на файлы формата JPEG.

4. СТАТИСТИЧЕСКИЕ АТАКИ НА ОСНОВЕ КРИТЕРИЯ ХИ-КВАДРАТ

Для анализа эффективности различных методов стеганоанализа в работе была реализована классическая атака хи-квадрат. Ее идея и обоснование приводятся в [5]. Младшие биты изображений не являются чисто случайными, что в равной мере справедливо и для младших бит DCT коэффициентов JPEG. Westfeld и Pfitzmann утверждают, что частоты двух соседних дискретных коэффициентов должны находиться достаточно далеко от значения частоты среднего арифметического этих коэффициентов. В "чистом" изображении ситуация когда частоты коэффициентов со значениями 2N и 2N+1 близки по значению встречается достаточно редко. При встраивании информации данные частоты сближаются или становятся равными. Идея атаки хи-квадрат и заключается в поиске этих близких значений и высчитывании вероятности встраивания на основе того, как близко располагаются значения частот четных и нечетных коэффициентов DCT. Особенностью алгоритма является последовательный анализ всего изображения и, соответственно, накапливание частот DCT. Метод хи-квадрат, в принципе, является универсальным — подходит для анализа изображений, созданных различными программами скрытия. Однако результаты атаки хи-квадрат в значительной мере зависят от способа скрытия данных. При последовательной записи в НЗБ дискретных коэффициентов метод обеспечивает хорошие результаты, а при псевдослучайном выборе младших бит и рассеивании сообщения по всей длине контейнера атака не срабатывает.

В работе с целью увеличения надежности обнаружения был предложен вариант этой атаки, когда изображение анализируются по частям. Принципиальное отличие от классического хи-квадрат заключается в том, что для каждого блока изображения рассчитываются свои наборы частот дискретных косинусных коэффициентов. В результате как бы анализируется несколько изображений. Это позволяет выявлять присутствие информации, скрытой в дискретных коэффициентах, выбранных псевдослучайным образом. В обычном хи-квадрат при анализе изображения происходит последовательное накопление частот DCT для всей картинки — это не дает возможности сформировать правильные статистики для изображений, сформированных утилитами, реализующими псевдослучайное скрытие.

Существует несколько параметров, при изменении которых могут быть получены различные варианты работы данного алгоритма. Один из таких параметров — это размер анализируемого блока изображения. Если выбирается блок сравнительно небольшого размера, то вероятность ложного обнаружения возрастает, и наоборот, при увеличении длин блоков вероятность ложных обнаружений падает, а вместе с ней падает и вероятность правильного обнаружения.

На сегодняшний день пока не разработан критерий, позволяющего оптимальным образом выбирать размеры блоков, хотя в перспективе математическая оценка длины блоков представляется необходимой.

Программная реализация предложенного алгоритма предполагала возможность пользователю самостоятельно указывать размер блока для анализа — это позволяло увеличить или уменьшить "чувствительность" алгоритма. Вместе с тем, экспериментально были получены хорошие результаты обнаружения факта скры-

тия информации при анализе блоков размером порядка 1 % от объема контейнера (размер изображения-контейнера предполагается более 200 Кб). При анализе меньших по размеру JPEG-файлов приходится увеличивать размер блока, чтобы уменьшить вероятность ложных обнаружений. Алгоритм продемонстрировал корректную работу со стегоизображениями, сформированными утилитами Jsteg, JPHide (v 3.0, v 5.0), CriptoBola JPEG.

Кратко рассмотрим алгоритмы работы протестированных утилит и результаты работы предложенных статистических атак.

Программа Jsteg использует последовательное встраивание данных в квантованные дискретные косинусные коэффициенты. С точки зрения стеганографической стойкости ее возможности довольно средние, поэтому статистические атаки оказываются эффективными. Блочный критерий хи-квадрат, равно как и обычный достаточно надежно обнаруживают присутствие спрятанной информации. На рис. 2 показаны результаты работы алгоритма при анализе изображения, размером 750 Кб, сформированного Jsteg и содержащего 70 Кб скрытой информации.

Алгоритм работы JPHide несколько отличается от Jsteg. Утилита для скрытия данных выбирает дискретные косинусные коэффициенты псевдослучайным образом. В результате при незначительном заполнении контейнера классический критерий хи-квадрат не срабатывает, в то время как применением блочного хи-квадрат были получены неплохие результаты. При значительном объеме встраиваемой информации порядка 75 % и более от возможного максимального объема скрываемых данных обе статистические атаки показывают присутствие стего. Разница в том, что обычный хи-квадрат не дает представления о реальной длине скрытых данных, в то время как блочный хи-квадрат позволяет стегоаналитику приблизительно оценить длину стего. На рис. 1 показаны результаты работы алгоритма при анализе изображения, размером 750 Кб, сформированного JPHide и содержащего 28 Кб скрытой информации.

Алгоритм работы CriptoBola JPEG похож на Jsteg. Принципиальное различие в том, что первый для скрытия данных использует не только самый младший 8 бит дискретных коэффициентов, но и более старшие 7, 6. В результате программа способна скрывать больше информации в файлы JPEG по сравнению с прочими протестированными утилитами. Поскольку CriptoBola JPEG не использует каких-то оригинальных алгоритмов скрытия, то увели-

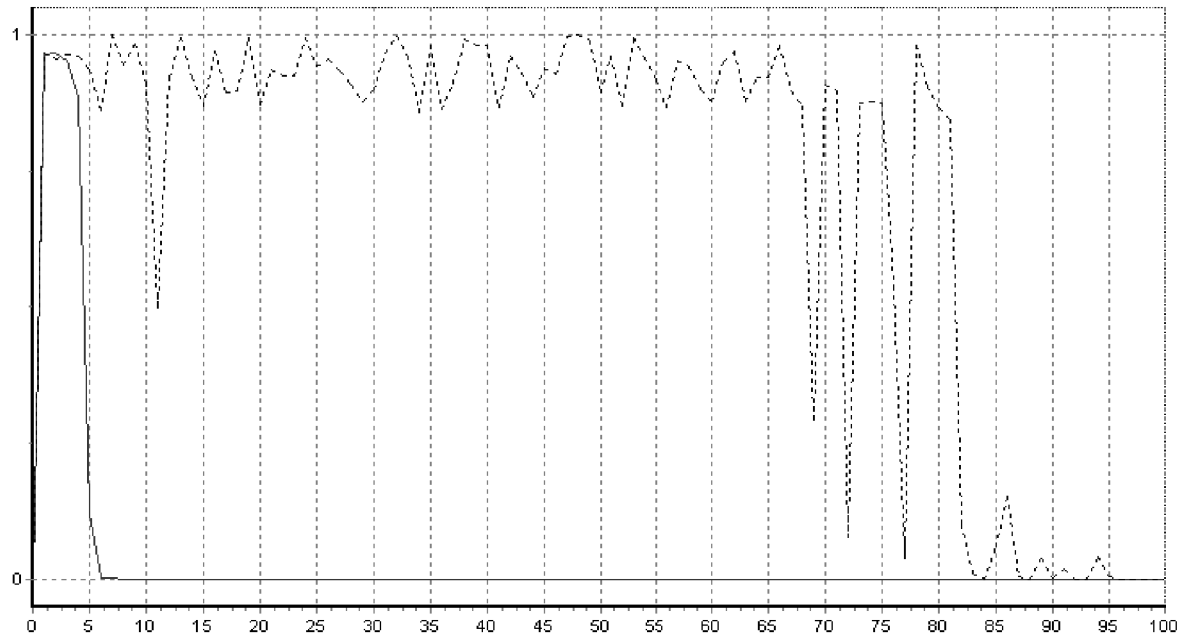

Рис. 1. Графики, показывающие вероятность скрытия информации в изображении: по оси X — процент проверенных DCT, по оси Y — вероятность встраивания. Сплошная линия — обычный хи-квадрат, пунктир — блочный хи-квадрат. Анализировалось изображение размером 750 Кб, содержащее 28 Кб скрытой информации

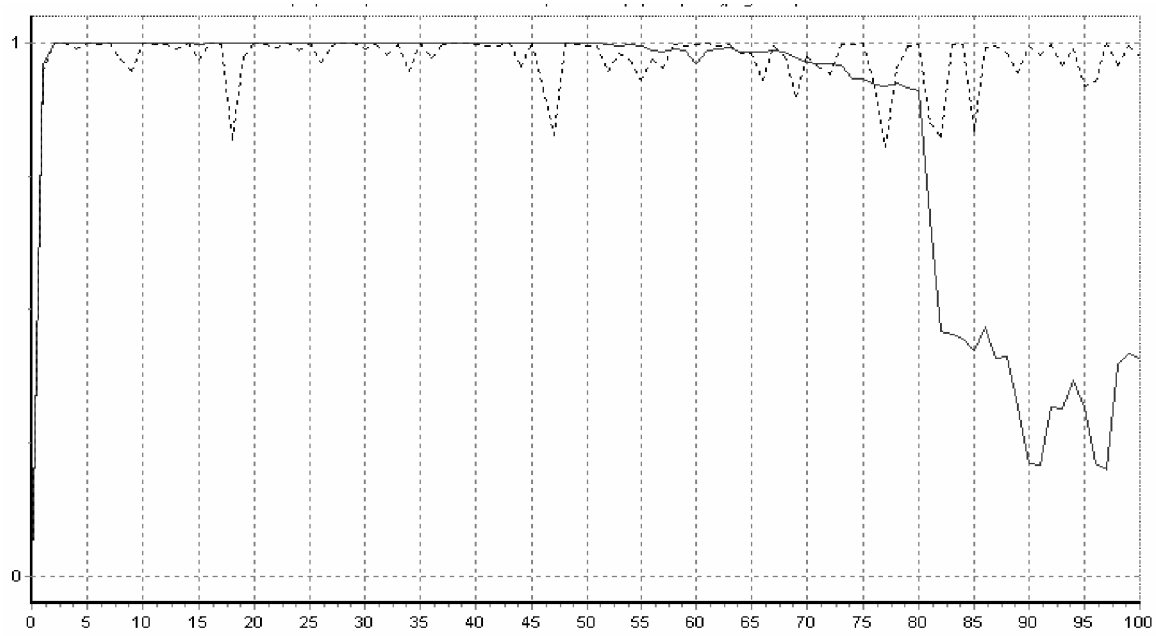

Рис. 2. Графики, показывающие вероятность скрытия информации в изображении: по оси X — процент проверенных DCT, по Y — вероятность встраивания. Сплошная линия — обычный хи-квадрат, пунктир — блочный хи-квадрат. Анализировалось изображение размером 750 Кб, содержащее 70 Кб скрытой информации

чение объема скрытой информации влечет ухудшение качества картинки. Если изменение восьмого бита дискретных коэффициентов практически не вносит искажений в изображение, то замена более старших бит вносит заметные изменения, которые проявляются в появлении блочности. Такой же эффект наблюдается при сильном сжатии JPEG. Разработанные статистические атаки показали предельно корректную работу при анализе изображений, сформированных CriptoBola JPEG.

При тестировании утилит Outguess и F5 разработанная программа не позволяет с достаточной уверенностью судить о наличии или отсутствии скрытой информации. Это объясняется тем, что данные программы используют несколько более совершенные методы скрытия, стойкие к статистическому анализу. Так в большинстве проведенных тестов обычный хи-квадрат не выявляет присутствия скрытой информации. Гистограммы блочного хи-квадрат как правило имеют скачкообразный характер. Не смотря на то, что вероятность скрытия нигде строго не равна 1, подобное поведение графика должно, по меньшей мере вызвать определенные подозрения у стегоаналитика. На рис. 3 показаны результаты работы алгоритма при анализе изображения, сформированного Outguess.

5. АНАЛИЗ ЧИСЛА ОКРУГЛЕНИЙ ЗНАЧЕНИЙ ЦВЕТОВЫХ СОСТАВЛЯЮЩИХ ПРИ ДЕКОМПРЕССИИ JPEG ИЗОБРАЖЕНИЯ

Один из статистических критериев анализа JPEG-файлов, который разрабатывался в работе, непосредственно связан с особенностями JPEG формата. В процессе JPEG декомпрессии осуществляется переход от цветового пространства YCbCr (интенсивности, цветности) к модели RGB. Значения каждого цветового канала RGB лежат в допустимых пределах от 0 до 255. Обычным явлением при декомпрессии являются величины, не попадающие в этот интервал, и декодеры JPEG в таких случаях предусматривают их округление. На рис. 4 приведены блоки значений цветовых составляющих оригинального изображения, а на рис. 5 - значения, полученные в процессе декомпрессии JPEG. В незакрашенных клетках стоят величины, не попадающие в интервал [0..255], именно они будут округляться.

Число значений, подлежащих округлению, напрямую зависит от коэффициента сжатия JPEG — чем сильнее сжатие, тем больше ошибок, и наоборот при малом сжатии будет получено меньше ошибок. Также при сокрытии информации в JPEG увеличивается число по-

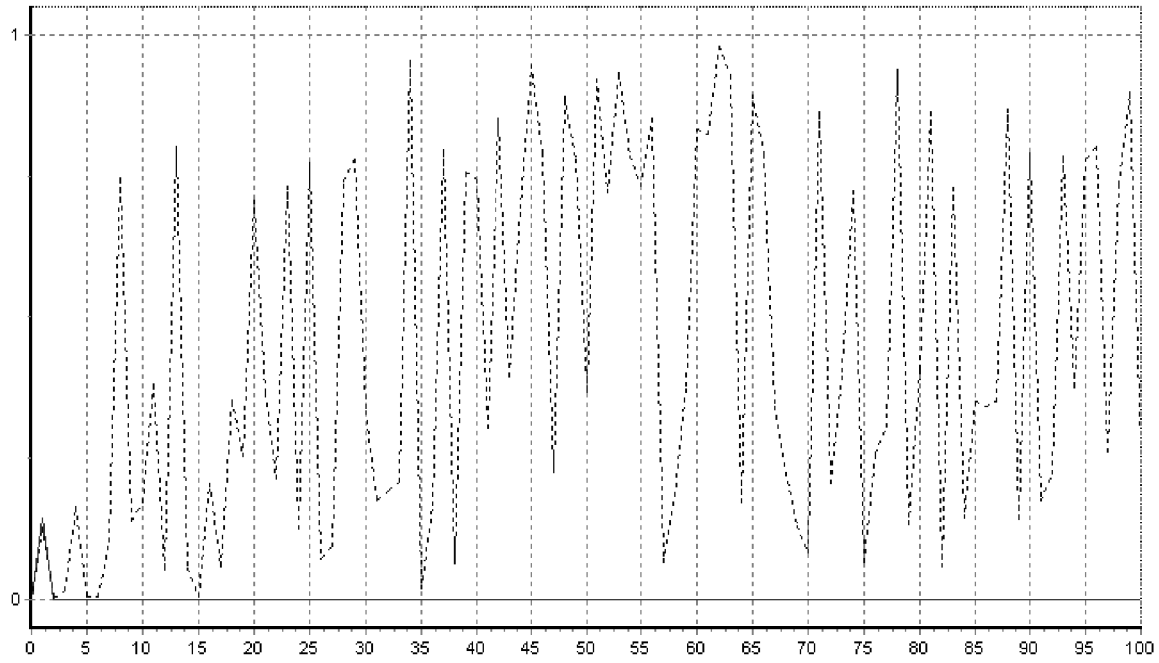

Рис. 3. Графики, показывающие вероятность скрытия информации в изображении: по оси X — процент проверенных DCT, по Y — вероятность встраивания. Сплошная линия — обычный хи-квадрат, пунктир — блочный хи-квадрат. Анализировалось изображение размером 1170 Кб, сформированное Outguess, и содержащее 114 Кб скрытой информации

лучаемых при декодировании ошибок. В работа [6] разработан статистический критерий, по которому можно судить о наличии скрытой информации. Метод включает несколько шагов. Сначала в анализируемом JPEG изображении подсчитывается число округлений значений цветовых составляющих получаемых при его декомпрессии — C0. Далее в НЗБ квантованных дискретных косинусных коэффициентов изображения (отличных от 0 и 1) встраивается случайно сгенерированная битовая последовательность и подсчитывается число округлений C1. Этот шаг повторяется N раз с различными битовыми последовательностями. Таким образом получают набор {C1, C2, ... CN}. После этого находится минимальный элемент Cmin из {C1, ... CN}. Оценка строится следующим образом: если $C0 \geq Cmin$, то принимается гипотеза о том, что изображение содержит скрытые данные, в противном случае гипотеза отвергается.

Несмотря на очевидность и простоту метод имеет ряд недостатков. Во-первых, невозможно заранее определить величину N — минимальное необходимое число случайно сгенерированных последовательностей, по которым накапливаются статистики {C1, ... CN}. Во-вторых, при достаточно больших N процедура анализа даже одного изображения занимает сравнительно много времени. В-третьих, и это пожалуй главное, метод корректно работает с контейнерами имеющими 100 % заполнение, причем данные должны встраиваться по алгоритму подобному используемому в программе Jsteg, то есть путем последовательного изменения НЗБ дискретных косинусных коэффициентов. Для Jsteg-подобных алгоритмов проще использовать другие статистические критерии, например хи-квадрат.

Как видно из вышеизложенного, предложенный в [6] вариант статистической атаки на JPEG изображения пригоден для анализа стегоконтейнеров заполнением, близким к 100 %. Рассмотрим метод анализа JPEG, применимый при частичном заполнении контейнера. Идея метода заключается в блочном анализе числа округлений значений цветовых составляющих, получаемых при декомпрессии. Изображение разбивается на N блоков, для каждого из них вычисляется количество округлений. Таким образом получаем набор {C1, ... CN} и по нему строится гистограмма рис. 6. Если используется метод последовательного встраивания информации, то даже при частичном заполнении контейнера, в блоках изображения, содержа-

R							
14	10	6	3	8	9	7	4
14	14	10	4	5	13	14	7
14	14	13	12	7	9	10	9
34	22	13	14	11	6	7	11
89	47	24	8	12	14	14	9
173	72	42	12	9	12	12	11
223	123	59	17	11	14	16	15
235	181	92	40	16	18	21	23

G							
11	11	11	12	11	11	12	13
12	10	10	12	14	11	11	14
11	14	12	11	13	12	12	14
3	11	12	10	12	16	17	13
4	3	9	15	12	12	13	13
42	0	7	15	16	11	10	12
102	20	1	8	14	15	15	13
123	60	0	3	12	15	18	18

B							
2	4	6	6	2	1	1	2
2	4	3	4	3	3	2	2
8	5	2	1	3	5	5	2
16	12	3	0	2	6	6	2
39	19	10	1	0	3	3	1
106	24	13	2	0	1	3	7
170	66	15	4	3	6	10	9
193	120	37	17	8	10	13	10

Рис. 4. Блоки значений цветовых составляющих исходного (оригинального) изображения

R							
6	7	4	0	2	11	17	18
9	10	8	4	4	8	12	14
17	15	13	10	6	4	6	9
37	27	17	12	7	2	3	9
79	51	24	14	9	4	6	14
145	92	39	18	13	9	12	21
218	140	59	25	20	17	19	26
266	173	75	31	25	22	23	29

G							
9	12	12	9	9	12	11	8
5	10	14	15	14	13	11	8
-1	6	14	21	20	15	10	8
-3	-1	8	20	23	16	11	11
10	-1	-1	13	22	17	12	13
44	13	-6	4	18	17	12	13
89	37	-5	-1	15	16	10	10
122	55	-1	-3	14	16	9	7

B							
1	-1	-7	-13	-11	0	9	13
2	2	-2	-2	-7	0	7	10
6	5	3	1	-1	0	4	8
17	10	5	5	3	0	2	8
47	24	6	4	5	1	3	10
97	52	12	3	6	4	4	11
156	86	23	4	7	6	5	10
197	114	33	6	9	7	5	8

Рис. 5. Блоки значений цветовых составляющих изображения, полученные при декомпрессии

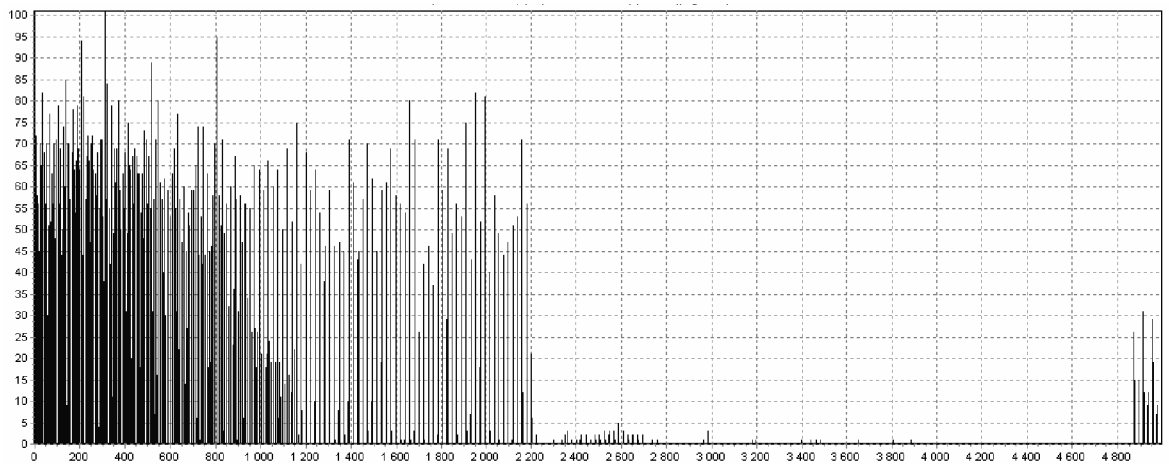

Рис. 6. Гистограмма частот округлений при декомпрессии JPEG изображения: контейнер содержит порядка 70 % скрытой информации от максимально возможного объема. По оси X — число анализируемых блоков, по Y — частота округлений для каждого блока

щих стего, значения из набора {C1,...CN} , будут больше, чем в блоках, не содержащих стего. Между тем, результаты работы данного метода могут в значительной степени зависеть от качества, характера, а также от размера изображения. Картинка сравнительно малого размера не позволит сформировать правильный дискретный набор {C1, ... CN}. Это справедливо и для большинства статистических атак — чем меньше набор отсчетов, тем хуже окончательная статистика. Качество изображения — также немаловажный фактор. Для цифровых изображений формата JPEG, имеющих большой коэффициент сжатия, возможно получение набора {C1,...CN} с очень большими дискретными значениями. Таким образом, сделать вывод о наличии скрытой информации будет невозможно.

Такая статистическая атака не рассматривается как основная, ее использование целесообразно в дополнение к другим стегоатакам при анализе достаточно больших изображений, содержащих много деталей (как следствие — много не нулевых дискретных косинусных коэффициентов). Как не странно изображения такого типа обычно и рекомендуется использовать в качестве стего-контейнеров. Полученные в результате гистограммы, позволяют стегоаналитику, если не сделать вывод о присутствии стего, то получить представление о структуре дискретных коэффициентов и распределении по длине файла потенциально округляемых значений цветов при декомпрессии. При принятии решения о наличии скрытой информации данный метод является дополнительным, но не окончательным.

6. АНАЛИЗ ЧИСЛА ПЕРЕХОДОВ ЗНАЧЕНИЙ МЛАДШИХ БИТ В СОСЕДНИХ ДИСКРЕТНЫХ КОЭФФИЦИЕНТАХ

Еще один вариант статистической атаки основан на анализе переходов НЗБ в соседних байтах квантованных дискретных косинусных коэффициентов потенциально использующихся для скрытия в них информации. Еще в работе [4] обосновывается возможность использования данной атаки для выявления факта скрытия в графических файлах формата BMP. Младшие биты в таких файлах не являются чисто случайными. Между младшими битами соседних элементов естественных контейнеров имеются существенные корреляционные связи. Также выявлены зависимости между НЗБ и остальными битами элементов естественных контейнеров. Все что касается файлов формата BMP оказывается применимо и к JPEG. Если в BMP анализируются младшие биты цветовых составляющих рядом стоящих пикселей, то в JPEG изображениях младшие биты соседних дискретных косинусных коэффициентов (соответственно отличных от 0 и 1). Зависимость между битами в соответствующих разрядах дискретных коэффициентов имеет марковский характер. При этом параметры зависимости определяются номером разряда. Число переходов в потоке НЗБ из 0 в 0, из 0 в 1, из 1 в 0, из 1 в 1 различно для файла-контейнера и файла, содержащего встроенную информацию. Распределение НЗБ файла со стего-вложением зачастую имеет случайный характер. Соответственно число переходов в потоке НЗБ для всех состояний будут примерно одинаковы, чего нельзя сказать о пустом файле-контейнере (конечно, если распределение НЗБ его дискретных коэффициентов не случайно). Рассмотренный в работе метод не дает абсолютной гарантии обнаружения скрытой информации, а позволяет с некоторой вероятностью подтверждать или опровергать присутствие скрытых вложений. Этот метод может быть полезен при принятии решения как дополнение к другим статистическим стегоатакам.

7. АНАЛИЗ ГИСТОГРАММ, ПОСТРОЕННЫХ ПО ЧАСТОТАМ КВАНТОВАННЫХ ДИСКРЕТНЫХ КОСИНУСОВЫХ КОЭФФИЦИЕНТОВ JPEG ИЗОБРАЖЕНИЯ

Еще один вариант анализа предусматривает построение гистограммы частот квантованных дискретных косинусных коэффициентов. Экспериментально обнаружено, что огибающая гистограммы ‘чистого’ изображения имеет более гладкий характер, по сравнению с гистограммами изображений, содержащими стего. Конечно, в зависимости от характера и степени сжатия изображения, гистограммы могут изменяться— в них могут появляться скачки и провалы, но важно то, что сокрытие информации меняет общий вид гистограмм. Большинство стеганографических утилит, работающих с JPEG, скрывают данные в младшие биты дискретных коэффициентов отличных от 0 и 1. Как следствие частоты 0-х и 1-х DCT не изменяются, в то время как все остальные частоты либо уменьшают-

ся, либо увеличиваются в зависимости от алгоритма встраивания. При значительных объемах скрываемой информации гистограммы часто приобретают ступенчатый характер, что нетипично для обычных JPEG изображений. Это хорошо видно на рис. 7, 8.

Разработка математической модели, позволяющей по виду гистограмм судить о наличии скрытой в JPEG-файле информации, представляется мало перспективной ввиду случайности получаемых результатов для различных изображений, и различных программ скрытия информации в JPEG. Однако данный метод имеет право на существование, так как продемонстрировал неплохие результаты при работе с Jsteg и ей подобными утилитами.

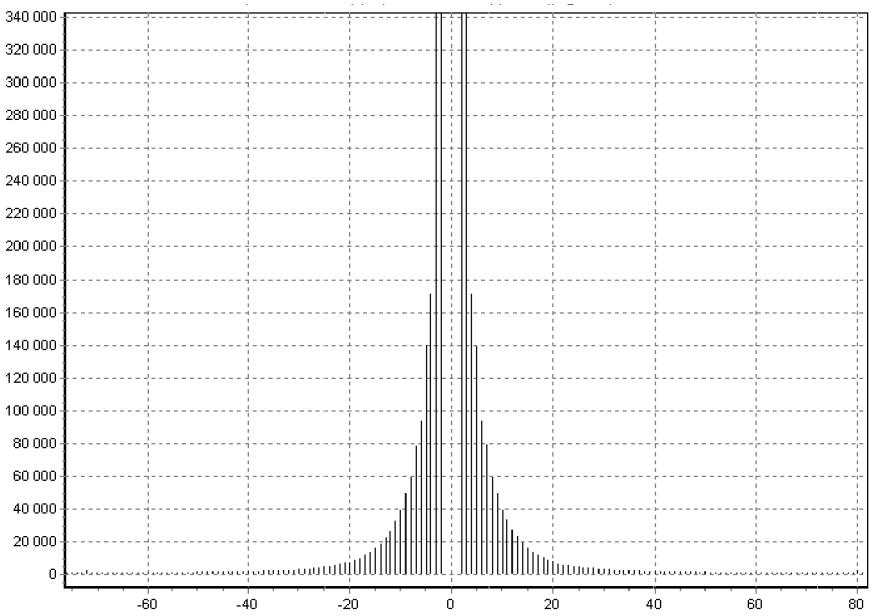

Рис. 7. Гистограмма частот DCT коэффициентов исходного изображения

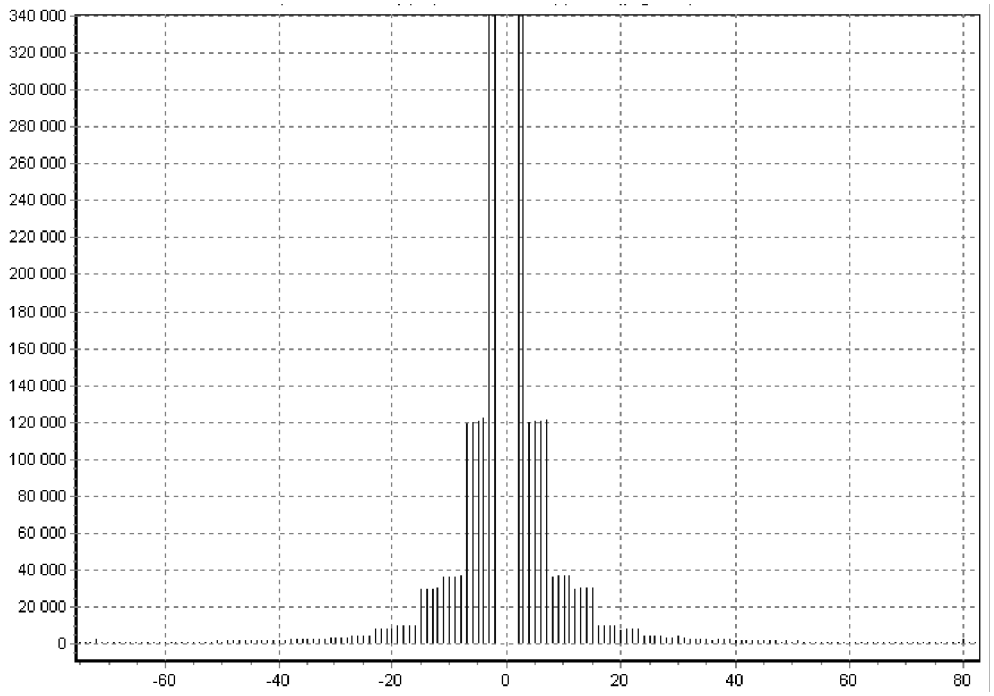

Рис. 8. Гистограмма частот DCT коэффициентов после встраивания данных

ЗАКЛЮЧЕНИЕ

Рассмотренные в статье алгоритмы, конечно, не являются средством, позволяющим со 100 % надежностью определять присутствие стеганографически скрытой информации. Как и все статистические методы, они дают возможность аналитику с определенной вероятностью судить о том, используется ли стеганография или нет.

Предложенные алгоритмы эффективны против большинства популярных стеганографических утилит, реализующих скрытие в JPEG, а разработанная программа представляет удобный инструмент для анализа и обнаружения факта скрытия информации. Также написана оригинальная программа, позволяющая скрывать информацию в JPEG файлах. Метод производит скрытие в один или два младших битов дискретных коэффициентов. Информация предварительно архивируется и шифруется.

При тестировании программы использовались утилиты, скрывающие информацию в JPEG-файлы Jsteg, CriptoBola JPEG, JPHide (v 3.0, v5.0), OutGuess, F5. Полученные результаты позволяют судить о корректности работы алгоритмов и высокой степени обнаружения (естественно при разумных соотношениях размеров скрываемого сообщения и файла-контейнера).

Предложенный вариант блочного хи-квадрат, анализирующего JPEG-файл по частям показал хорошие результаты при анализе изображений, сформированных утилитами использующими псевдослучайные алгоритмы выбора дискретных коэффициентов для скрытия в них информации. Притом результаты были существенно лучше, чем при обычной атаке хи-квадрат. Обычная атака хи-квадрат в таких случаях либо не выявляет факта скрытия, либо не позволяет оценить длину спрятанного сообщения.

Реализованная в программе статистическая атака на основе анализа числа округлений цветовых значений при декомпрессии JPEG-файла имеет вероятностный характер. Корректная работа алгоритма была доказана при анализе программ последовательно скрывающих информацию в дискретных коэффициентах JPEG (Jsteg, CriptoBola JPEG).

Атаки на основе анализа гистограмм частот дискретных косинусных коэффициентов в JPEG-файле, и гистограмм числа переходов значений НЗБ соседних дискретных коэффициентах позволяют делать выводы о наличии скрытой информации при значительном заполнении стегоконтейнера. При скрытии небольшого объема информации, эти атаки не эффективны.

На результат работы алгоритмов влияют размер, степень сжатия, характер изображения и, естественно, используемый метод скрытия информации. Как уже отмечалось, все статистические методы стеганоанализа имеют вероятностный характер. А значит, при анализе любого файла невозможно будет утверждать, что в нем не содержится скрытая информация.

ЛИТЕРАТУРА

1. *Миано Дж.* Форматы и алгоритмы сжатия изображений в действии / Дж. Миано —М. : ТРИУМФ, — 2005. — 330 с.

2. *Грибунин В.Г.* Цифровая стеганография / В. Г. Грибунин. — СПб. : СОЛОН-Пресс, — 2002. — 280 с.

3. *Генне О.В.* Защита информации / О. В. Генне // Конфидент. — 2000. — № 3. — С. 20—25.

4. *Барсуков В.С.* Оценка уровня скрытности мультимедийных стеганографических каналов хранения и передачи информации / В. С. Барсуков, А. П. Романцов // Специальная Техника. —2000. — № 1. (http://www.bnti.ru/dbtexts/analmat/2/barsukov.pdf)

5. *Westfeld A. Pfitzmann A.* Attacks on Steganographic Systems: Breaking the Steganographic Utilities EzStego, Jsteg, Steganos and S-Tools-and Some Lessons Learned // 3rd International Workshop on Information Hiding (2000). — (http://www.ece.cmu.edu/~adrian/487-s06/westfeld-pfitzmann-ihw99.pdf)

6. *Yeuan-Kuen Lee, Shih-Yu Huang* A novel quantity based on clipping statistics for Jsteg steganalysis // Submit to the 8th IASTED International Conference on SIGNAL & IMAGE PROCESSING. — (http://imedia.cce.mcu.edu.tw/interactive/resserver.php?blogId=17&resource=SIP06-JS-DRAFT.pdf)

Статья принята к опубликованию 25 декабря 2006 г.