Федеральное Агентство Связи

Ордена Трудового Красного Знамени

Федеральное государственное бюджетное

Учреждение высшего образования

Московский технический университет связи и информатики

Кафедра “Информатика”

Курсовая работа по дисциплине «Основы программирования»

На тему: «Стеганоанализ файлов»

Выполнил студент группы БСТ2002

Палибза С.А.

Проверил: доцент кафедры “Информатика”

Гуриков С.Р.

Москва 2020

Содержание

[Введение 3](#_Toc54852831)

[Техническое задание 5](#_Toc54852832)

[Глава 1. Теоретическая часть 8](#_Toc54852833)

[1 Классификация методов стеганоанализа 8](#_Toc54852834)

[1.2 Стеганоанализ наименьших значащих бит 9](#_Toc54852835)

[1.2.1. Визуальные атаки 9](#_Toc54852836)

[1.2.2. Использование для стеганоанализа статистики хи-квадрат 14](#_Toc54852837)

[1.2.3. Использование для стеганоанализа функции искажение и скорость 17](#_Toc54852838)

[1.2.4. Использование для стеганоанализа НЗБ статистики пар значений 22](#_Toc54852839)

[1.2.5. Использование для стеганоанализа НЗБ вейвлет декомпозиции 24](#_Toc54852840)

[1.2.6. Стеганоанализ НЗБ на основе статистики числа переходов 28](#_Toc54852841)

[1.2.7. Стеганоанализ НЗБ на основе статистики длин битовых серий 30](#_Toc54852842)

[1.2.8. Критерии информативности системы признаков 31](#_Toc54852843)

# Введение

**Актуальность темы.** В современном мире проблема безопасной и гарантирующей целостность передачи данных **имеет особую актуальность**. Задача стегоанализа состоит в обнаружении факта передачи секретного сообщения. Стеганоанализ получил **широкое применение** в сфере обеспечения информационной безопасности и, в частности, для борьбы с незаконной передачей информации. **Например,** в некоторых отечественных и иностранных компаниях служба безопасности проверяет исходящую электронную почту сотрудников для пресечения утечки закрытой коммерческой информации. **Принимая во внимание** широкую доступность и разнообразие программных продуктов, позволяющих встраивать скрытую информацию в обычные «невинные» письма, становится очевидной актуальность совершенствования методов стеганоанализа. **Учитывая** большой объем передаваемых данных, перспективными следует считать методы компьютерного анализа, работающие без участия человека. Стегоанализ также может быть применен злоумышленником. **Например,** для случаев с цифровыми отпечатками пальцев в программе, атакующий может выявить факт существования специальных меток в программе и попытаться их исказить или удалить. В таком случае развитие методов стеганоанализа необходимо для установления потенциальных возможностей злоумышленника и, соответственно, для корректировки схем внедрения скрытой информации. В настоящее время проводится множество конференций, по проблемам информационной безопасности. С каждым годом растет число публикаций, посвященных методам стеганографии и стеганоанализа. В этом направлении науки работают многие российские и зарубежные ученые. **Все это еще раз доказывает актуальность данной проблемы**. АБЗАЦЫ БУДУТ??? ИНТЕРВАЛ КАКОЙ В ГОСТ НАПИСАН??

**Объектом исследования**, проведенного в рамках данного курсового проекта, является стеганоанализ файлов.

**Предметом исследования** является классификация методов стеганоанализа и стеганоанализ наименьших значащих бит.

**Цели работы и задачи исследования.** Цель работы – исследование методов стеганоанализа, а также стеганоанализ наименьших значащих бит и разработка программы для выведения найденной информации на экран. Поставленная цель определила следующие основные задачи исследования:

1. Проанализировать найденный материал по теме исследования.
2. Разработать программу для вывода найденной информации на экран. ЭТО КАК??

**Методы исследования.** Для решения поставленных задач были использованы теоретические методы исследования. Теоретическую основу исследования составили труды в области стеганографических систем Грибунин В.Г., Костюков В.Е., Мартынов А.П., Николаев Д.Б., Фомченко В.Н.

# Техническое задание ЗДЕСЬ И ДАЛЕЕ НАДО ЧИТАТЬ ГОСТ В ЧАСТИ ОФОРМЛЕНИЯ ЗАГОЛОВКОВ

* 1. **Основание для разработки**

Основанием для разработки является работа, выполненная в соответствии с заданием, полученным от кафедры “Информатика” Московского Технического Университета Связи и Информатики и утвержденным научным руководителем доцентом кафедры “Информатика” к.т.н. Гуриковым С.Р.

Дата утверждения: 2 октября 2020г.

* 1. **Назначение разработки**

Разработанный программный продукт предназначен для ознакомления с классификацией методов стеганоанализа и стеганоанализа наименьших значащих бит, а также проверки знаний пользователя.

* 1. **Требования к программе**
     1. **Требования к функциональным характеристикам**

Разработанный программный продукт должен обеспечить выполнение следующих функций:

1. Возможность ознакомления пользователя с классификацией методов стеганоанализа и стеганоанализа наименьших значащих бит.
2. Возможность пройти тестирование на основе прочитанного материала, проверяющее знания пользователя по теоретической части
3. Возможность авторизации пользователя
4. Возможность вывода результатов тестирования в файл .txt.
5. Запрашивание пароля при запуске приложения в целях зашиты от несанкционированного доступа.
   * 1. **Требования к надежности**

Разрабатываемое программное обеспечение должно обеспечивать:

1. Устойчивую работу при соответствующим алгоритме функционирования
2. Выдачу сообщений об ошибках
3. Поддержание диалогового режима в рамках предоставляемых пользователю возможностей
4. Защиту приложения при помощи пароля, который будет запрашиваться при запуске программы
   * 1. **Требования к составу и параметрам технических средств**
     2. **Минимальные системные требования**: ФОРМАТИРОВАНИЕ??

* Процессор с тактовой частотой не ниже 1,8 ГГц. Рекомендуется использовать как минимум двухъядерный процессор.
* 2 ГБ ОЗУ; рекомендуется 8 ГБ ОЗУ (минимум 2,5 ГБ при выполнении на виртуальной машине)
* Место на жестком диске: до 210 ГБ (минимум 800 МБ) свободного места в зависимости от установленных компонентов; обычно для установки требуется от 20 до 50 ГБ свободного места.
* Скорость жесткого диска: для повышения производительности установите Windows и Visual Studio на твердотельный накопитель (SSD)
* Видеоадаптер с минимальным разрешением 720p (1280 на 720 пикселей); для оптимальной работы Visual Studio рекомендуется разрешение WXGA (1366 на 768 пикселей) или более высокое.
  + 1. **Требования к информационно-программной совместимости**

Программа должна легко инсталлироваться и корректно работать при наличии следующего программного обеспечения:

1. Операционной системы Windows 7, Windows 8, Windows 8.1 Windows 10
2. Приложения “Блокнот”
   * 1. **Требования к транспортированию и хранению**

Программа должна поставляться на flash-карте. Документация к программе предоставляется в электронном и печатном виде.

* + 1. **Требования к программной документации**

В ходе разработки программы были разработаны: программный код и его описание, методика испытаний, а также руководство пользователя.

* + 1. **Стадии и этапы разработки**

**ЗДЕСЬ И ДАЛЕЕ НАДО ЧИТАТЬ ГОСТ В ЧАСТИ ОФОРМЛЕНИЯ ТАБЛИЦ**

Таблица 1 – Стадии разработки

|  |  |  |  |
| --- | --- | --- | --- |
| № этапа | Название этапа | Срок выполнения | Отчетность |
| 1 | Утверждение темы | До 02.10.2020 |  |
| 2 | Написание введения | До 12.10.2020 |  |
| 3 | Составление ТЗ и написание теоретической части | До 29.10.2020 |  |

# Глава 1. Теоретическая часть

# 1 ПУНКТА 1 НЕ МОЖЕТ БЫТЬКлассификация методов стеганоанализа

Обычно под стеганоанализом понимается совокупность способов, методов и приемов определения факта наличия стегановнедрения в контейнере. Также, в известных работах можно найти следующие понимания целей стеганоанализа:

1) обнаружение факта наличия в контейнере внедренной информации (и, возможно, определение ее объема);

2) восстановление смыслового содержания внедренной информации (на самом деле, это не является целью стеганоанализа и далее рассматриваться не будет)

3) оценка местоположения цифрового водяного знака (ЦВЗ) с целью его дальнейшего удаления;

4) искажение контейнера без оценки ЦВЗ так, чтобы ЦВЗ не обнаруживался детектором они также применимы для разрушения скрытого сообщения

5) нарушение работы системы ЦВЗ так, чтобы она не детектировала имеющийся ЦВЗ (компрометация системы – не является целью стеганоанализа).

Здесь будут рассмотрены методы выявления факта наличия в контейнере внедренной информации. Наиболее простой стеганоанализ заключается просто в осмотре контейнера. Если в контейнере содержится достаточно большой объем внедренной информации и внедрение выполнялось без учета локальных свойств контейнера (яркости изображения и т.п.), то в результате визуального осмотра может быть выявлен факт внедрения (подозрительный шум в гладких областях). Также внедренная информация может быть легко обнаружена, если внедрение осуществлялось форматными методами.

Второй подход состоит в том, чтобы искать аномалии в предполагаемой области внедрения. Например, внедрение в наиболее младший бит (НЗБ) в палитровое изображение может явиться причиной большого количества «двойных» цветов, где идентичные (или почти идентичные) цвета дважды появляются в палитре и имеют отличия только в самом младшем двоичном разряде. Анализ аномалий структур может позволить выявить сигнатуру конкретного стеганографического алгоритма. Третьим подходом является так называемый целевой стеганоанализ, заключающийся в поиске известных артефактов внедрения различных стеганоалгоритмов. Четвертым, и наиболее распространенным, подходом является статистический стеганоанализ, называемый еще иногда слепым стеганоанализом. В этом случае от контейнера вычисляются некоторые характеристики и анализируются статистики первого и высших порядков. ПАРАГРАФ НЕ МОЖЕТ БЫТЬ ТАКИМ КОРОТКИМ

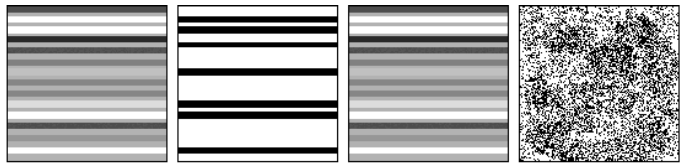
## 1.2 Стеганоанализ наименьших значащих бит

### 1.2.1. Визуальные атаки **НЕ НАДО ДЕЛАТЬ 3-Х УРОВНЕВОЕ ВЛОЖЕНИЕ**

Метод внедрения НЗБ (LSB – Least Significant Bits), или метод замены младших бит внедряемым сообщением, был предложен Е. Адельсоном в 1990 г. На сегодняшний день он является одним из наиболее широко используемых методов сокрытия информации. Основная идея метода заключается в замене от одного до четырех младших битов в байтах цветового представления точек исходного изображения битами скрываемого сообщения. Сама возможность такой замены обусловлена некоторой избыточностью представления цвета и зачастую случайным поведением младших битов изображения.

Полноцветные изображения в режиме True Color хранят информацию о цвете каждой точки в трех байтах. Каждый из байтов содержит информацию об интенсивности одной из трех составляющих цвета (палитра RGB): красного, синего, зеленого. Так как под хранение каждой из составляющих отводится по одному байту, то соответствующие им интенсивности цветов изменяются в пределах от 0 до 255. Таким образом, общее число возможны х цветов равно 2563 ≈ 1,6·107. Считается, что человеческий глаз способен различить лишь порядка четырех тысяч цветов, значит для кодирования достаточно всего четырех бит на цвет. При использовании метода НЗБ предполагается, что цвет точки практически в полной мере определяется старшими четырьмя битами каждого из трех байтов представления RGB, а оставшиеся четыре младших бита можно использовать для внедрения скрываемой информации.

Рассмотрим метод визуального анализа битовых срезов, приведенный в. Он заключается в том, что с помощью несложной программы изображение просматривают по слоям, т.е. берутся битовые срезы изображения. Учитывая то, что интенсивность каждого цвета определяется ровно одним байтом, всего необходимо будет просмотреть 8 таких срезов. Для каждого из трех цветов первый срез – это изображение, построенное самыми младшими битами, второй срез – изображение, построенное вторыми битами и т.д. Для большей четкости битового среза в те байты, в срезе которых стоит единица, заносят максимальное значение (255). Далее полученное изображение битового среза просматривают и визуально сравнивают с исходным изображением. На рис. 1.1 приведен пример подобного анализа.



а б в г

Рисунок 1.1 – Сравнение битовых срезов: а – исходное изображение; б – битовый срез по последнему биту; в – изображение с внедренным в последние биты сообщением; г – битовый срез полученного изображения

Как видно из примера, здесь нарушается предположение о том, что младшие биты всегда случайны. Более того, между младшими битами существуют некоторые закономерности, и их поведение совсем не похоже на случайное, а изображение, построенное на их основе, на шум. Например, в изображениях довольно часто встречаются длинные серии из одинаковых бит и практически любое изображение содержит серию минимум из 14 одинаковых бит. При ситуации, когда в младшие биты изображения происходит внедрение информации, эти закономерности нарушаются.

Кроме того, внедряемые сообщения могут не содержать столь длинных серий из одинаковых бит. Довольно часто внедрение сообщений в изображения осуществляется побайтно. При таком способе внедрения длинные серии могут получиться только в случае, если сообщение содержит рядом расположенные байты, равные 0 или 255. Однако практически всегда до внедрения сообщения в контейнер проводится его предварительное шифрование и (или) сжатие. В результате сообщение представляет собой нечто похожее на случайную последовательность бит. Как следствие, в срезе, в который внедряется полученное в результате шифрования и (или) сжатия сообщение, число длинных серий, равно как и их длина, сильно сокращается. На данном факте основывается также целый ряд методов статистического стеганоанализа.

Различные исследования в области особенностей человеческого зрения показывают, что порог чувствительности глаза к изменению освещенности при средних ее значениях составляет *ΔI =* 0,01 *÷* 0,03*I* или 1~3 % (рисунок 1.2). Заметим, что использование для внедрения информации четырех младших разрядов в байтах исходного изображения может привести к изменению интенсивности порядка 6 %, что в два раза превышает порог чувствительности человеческого глаза.

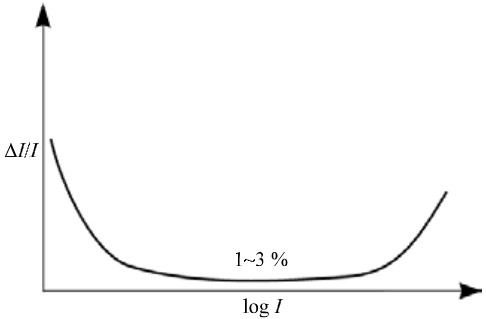


Рисунок 1.2 - Порог чувствительности человеческого глаза

Проведенные эксперименты подтвердили, что замена не четырех, а даже трех младших битов (~3 %) вносит заметные для человеческого глаза искажения. Изменение же яркости в переделах 1–1,5 % в действительности оказалось незаметным. Следовательно, для того чтобы внедрение в изображение дополнительной информации оказалось незаметным для человеческого глаза, наиболее оптимальным будет подвергнуть модификации либо первый и второй разряды (максимальное искажение 1,17 %), либо только третий разряд (искажение 1,5 6 %) каждого из трех байтов, отвечающих за цвет точки.

Надо отметить, что изображения, полученные со сканера или фотоаппарата и созданные на компьютере, отличаются друг от друга.

Например, сканированное изображение, в которое ничего не внедрялось, выглядит так (рисунок 1.3):



Рисунок 1.3 - Сканированное изображение без внедренной информации

А вот – изначально компьютерное (рисунок 1.4):

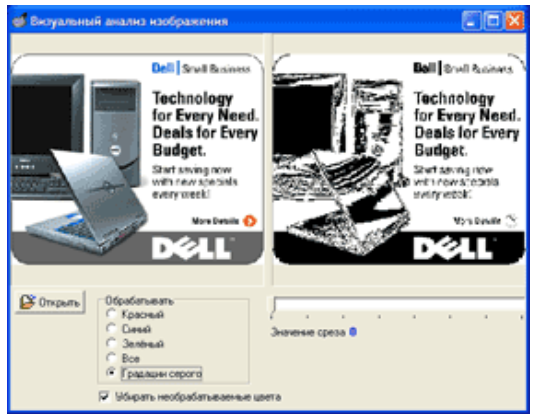


Рисунок 1.4 - Изначально компьютерное изображение без внедренной информации

Это можно объяснить шумом матрицы сканера (или цифровой камеры). Необходимо упомянуть, что просмотром и анализом распределения НЗБ иногда очень просто определить факт компьютерной обработки (подчистки) картинки. Например, на рисунке 1.5 представлен образец изображения.

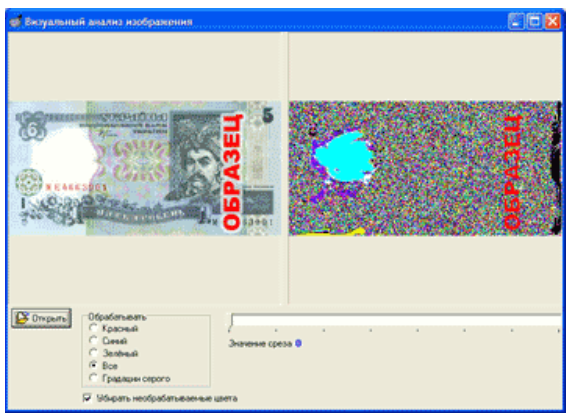


Рисунок 1.5 - Образец изображения

А вот это же изображение, но с внедренной информацией (рисунок 1.6):

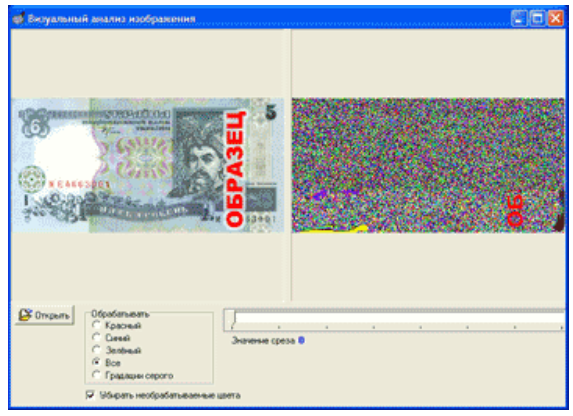


Рисунок 1.6 - Образец предыдущего изображения, но в внедрением

Здесь четко видна граница между «своим» шумом и шумом, возникшим в результате внедрения сообщения. Так появляется еще одно необходимое правило: всегда нужно распределять биты сообщения по всем младшим битам контейнера (например, внедрять не последовательно, а в каждый 2-й, 3-й и т.д. НЗБ) или же дополнять чем-то (шумом с тем же законом распределения) длину сообщения так, чтобы она стала равна объему НЗБ контейнера.

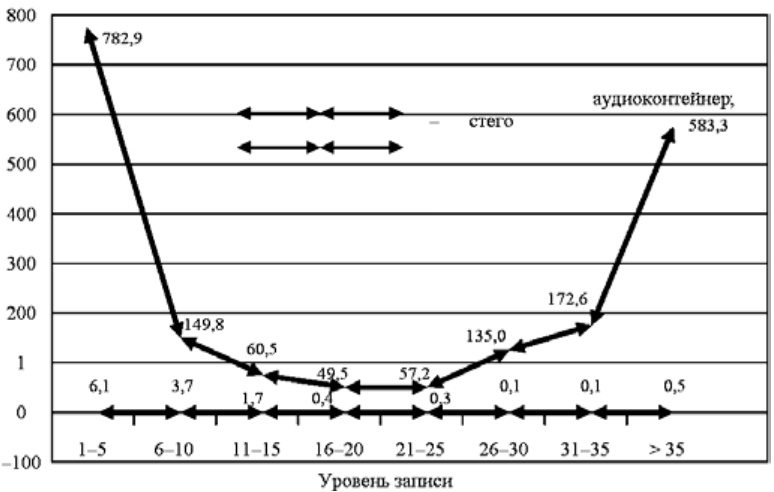
### 1.2.2. Использование для стеганоанализа статистики хи-квадрат

Другим наиболее простым известным способом стеганографического анализа методов внедрения в НЗБ является оценка статистики хи-квадрат распределения этих бит. ГДЕ ИСПОЛЬЗОВАНИЕ НАУЧНОГО СТИЛЯ??

В работе показано, что следы скрытия проявляются при анализе таких статистических характеристик речи и музыки, как, например, распределение НЗБ отсчетов стеганограммы, условные распределения младших и остальных разрядов отсчетов, величины коэффициента корреляции между соседними отсчетами и т.п.

При исследовании метода НЗБ было обработано свыше 1200 аудиофайлов стеганограмм и пустых контейнеров, записанных на CD дисках и представляющих собой различные музыкальные и вокальные произведения разных авторов. Показано, что для пустых аудио-контейнеров НЗБ и остальные биты статистически взаимно зависимы, причем на характер этой зависимости влияет уровень записи (усредненная амплитуда отсчетов аудиосигнала).

На рисунке 1.7 показана полученная для аудиофайлов зависимость статистики хи-квадрат. По критерию хи квадрат вычислялась степень различия между распределением пустых и заполненных контейнеров от характерного для стеганограмм бернуллиевского распределения.

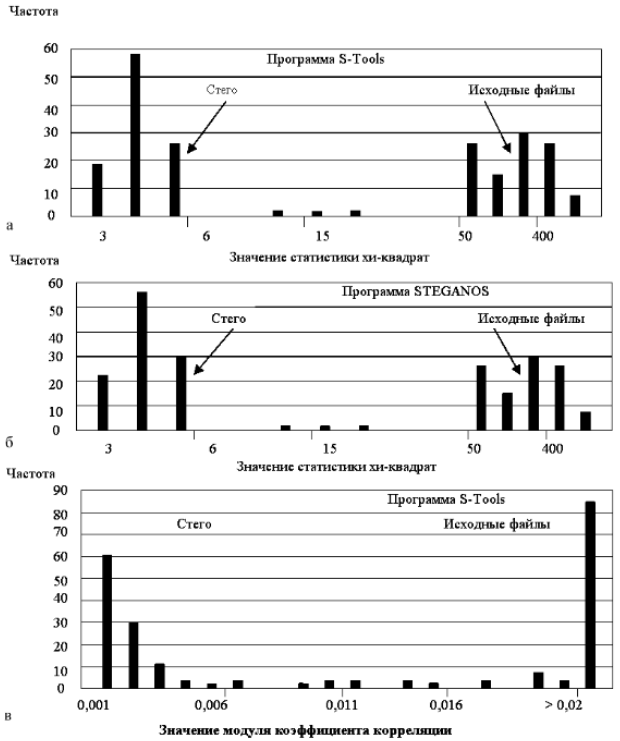


*Рисунок 1.7 - Зависимость величины статистики хи-квадрат от амплитуды отсчетов аудиосигнала*

На данный момент известны результаты анализа стегостойкости некоторых программных средств скрытия информации в аудиофайлах. Используя статистику хи квадрат и коэффициент корреляции, проведен стеганоанализ программ Steganos (version 1.0) и S-Tools (Steganography Tools for Windows, version 2.0), которые скрывают информацию в наименее значимых битах звуковых отсчетов. В качестве исходных контейнеров исследовались считанные с CD дисков 100 музыкальных фрагментов различных исполнителей длительностью звучания 15 с каждый (как со стандартных музыкальных компакт-дисков, так и с дисков в формате МР3). В качестве скрываемого сообщения использовалась псевдослучайная последовательность объемом 83 кбайт, которая побитно внедрялась в каждый НЗБ отсчетов контейнера. По критерию хи-квадрат определялась степень отличия распределения НЗБ отсчетов исследуемой последовательности от бернуллиевского распределения.

Результаты статистических вычислений для музыкальных контейнеров и сформированных из них полностью заполненных стегопредставлены в виде гистограмм на рисунке 1.8. При этом область значений статистики (ось абсцисс) разбита на непересекающиеся и различные по размерам интервалы. Высота столбца (ось ординат) показывает число значений статистики, попавших в заданный интервал. На рисунке приведена частота встречаемости значений статистики хи-квадрат (а – для STools, б– для Steganos) и коэффициента корреляции (в – для STools). Правые столбцы соответствуют пустым контейнерам, а левые – стегограммам. Для стегограмм величина хи-квадрат была равна единицам, а для пустых контейнеров – десяткам и сотням. Также было выявлено, что после внедрения среднее значение коэффициента корреляции соседних отсчетов уменьшилось в десятки раз.

Необходимо обратить внимание на то, что диапазоны значений статистики хи-квадрат, вычисленные до и после внедрения дополнительной информации, и диапазоны значений коэффициентов корреляции не пересекаются. Данные признаки позволяют при использовании статистических атак с большой вероятностью отделить пустые аудио-контейнеры от стегограмм.



*Рисунок 1.8 - Статистические различия стего и пустых аудио-контейнеров: а, б – по критерию хи-квадрат, в – по модулю коэффициента корреляции*

### 1.2.3. Использование для стеганоанализа функции искажение и скорость

Другим методом определения наличия внедренного в НЗБ сообщения является метод оценки функции искажение – скорость. Указанный метод был предложен М. Селиком и др. в 2004 г. для изображений. В основе метода лежат два постулата:

1. Внедрение данных обычно увеличивает энтропию контейнера. Для того чтобы скрыть способы стеганографического внедрения, изменяют элементы контейнера. Эти изменения обычно не связаны со статистикой контейнера, и п о тому увеличивают его энтропию.
2. Методы внедрения данных ограничены малыми изменениями. Стеганографические методы внедрения информации, как правило, должны обеспечивать незаметность их применения. Значимые для восприятия человека элементы сообщения должны остаться неизменными.

Стохастическое внедрение, как правило, не приводит к образованию в контейнере изменений, которые бы однозначно свидетельствовали о наличии встроенного сообщения. Исходя из этого для выявления тоже должны быть использованы некоторые статистические характеристики. Принцип выявления факта стеганографического внедрения может быть описан с помощью схемы, показанной на рисунке 1.9.



*Рисунок 1.9 - Процесс выявления факта внедрения*

В качестве оцениваемых характеристик авторами были использованы значения искажений при заданных скоростях (объемах). В качестве метрик искажений использовались среднеквадратическая ошибка, средняя абсолютная ошибка и взвешенная среднеквадратическая ошибка. В целях уменьшения разнообразия классов контейнеров и стеганобразов значения скоростей определялись для сжатия без потерь. Авторами был использован байесов классификатор, перед которым над вектором характеристик было осуществлено *KL*-преобразование.

Классификатор и *KL*-преобразование могут быть описаны следующим образом. **ЗДЕСЬ НАДО ЧИТАТЬ ГОСТ В ЧАСТИ ОФОРМЛЕНИЯ ФОРМУЛ**

Пусть *wi– i*-й стеганографический метод, где 1 ≤ *i* ≤ *M*, *M* – количество существующих методов, а *x* – *L*-мерный вектор характеристик:

где *bfµi* = *E* [*x*] – среднее значение класса *wi*; *∑i* – ковариационная матрица; *|∑i|* – детерминант *∑i; E*[ ] – математическое ожидание случайной величины.

В качестве примера авторами было рассмотрено обнаружение факта внедрения в НЗБ.

Пусть для передачи 7 старших битов изображения *I* требуется скорость *R7MSB*(*I*), аналогично для передачи в каждом пикселе информации на 1 бит больше потребуется *R7MSB*(*I*+1), Принимая во внимание статистику натурального изображения, можно предположить, что *R7MSB*(*I*) ≈ *R7MSB*(*I*+1).

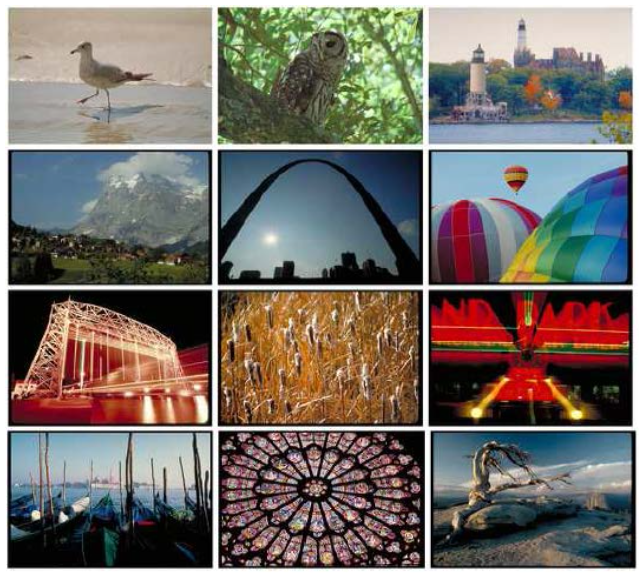
Но процесс внедрения информации в НЗБ вносит некоторые изменения в отношения статистики изображения и его двоичного представления. В общем случае *R7MSB*(*IS*) ≤ *R7MSB*(*IS*+1).

Разность между скоростями может быть определена как *∆I =*

*= |R7MSB*(*IS*+1) - *R7MSB*(*IS*)*|*. *∆I* будет хорошим средством для стеганографического анализа внедрения в НЗБ. Но опыты, проведенные авторами, показали, что в этом случае высока вероятность ложных тревог. Поэтому данная величина была нормализована. Для этой цели младшие биты изображения были заменены на случайные и получено изображение . В принципе, статистически соответствует стегограмме с полной загрузкой. В результате вычисляется характеристика .

Эта характеристика в ходе проведенных авторами экспериментов обеспечивала лучшую правильность определения наличия скрываемого сообщения.

Для получения репрезентативной выборки изображений авторами были использованы 108 фотографий с сайта компании «Кодак», на фотографиях изображены различные объекты, все изображения приведены в формат RGB TIFF с разрешением 768×512 пикселов. Приблизительная подборка изображений показана на рисунке 1.10.

**

*Рисунок 1.10 - Подборка изображений для анализа*

Авторами были проведены эксперименты по определению внедрения методом НЗБ. В ходе экспериментов использовались 6 значений скорости внедрения в НЗБ: 0,1; 0,2; 0,4; 0,6; 0,8 и 1,0 бит на пиксел. Для определения наличия скрываемого сообщения использовалась характеристика *Ψ ()* при пороговом значении 0,1. Таким образом, алгоритм стеганографического анализа рассматривался как способный обнаруживать внедрение в НЗБ при скорости внедрения от 0,1 бит на пиксел и выше. В таблице 1.1 показаны результаты попыток обнаружения внедрения в НЗБ.

Таблица 1.1 - Результаты проведения опытов

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Скорость внедрения | 0,1 | 0,2 | 0,4 | 0,6 | 0,8 | 1,0 |
| Процент ошибок | 35 | 15 | 3 | 0 | 0 | 0 |

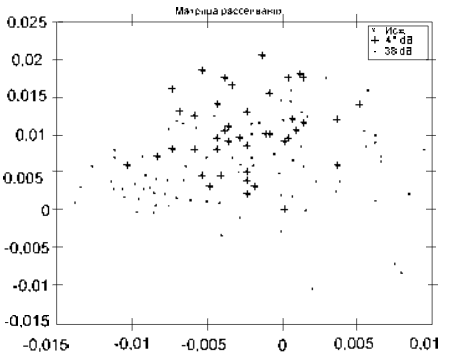
Кроме обнаружения сообщений, встроенных за счет использования НЗБ, М. Селик и др. предполагали, что метод может быть применен и для определения наличия сообщений, встроенных за счет ШПС.

По мнению авторов, в этом случае процесс внедрения без ограничения общности можно представить как процесс добавления к изображению шума. Хотя, в принципе, метод может предполагать шум произвольного распределения, проверка проводилась на гауссовом распределении. Скорость внедрения была 0,84 и 0,91 бит на пиксель.

Тренировочные и тестовые множества отбирались случайно и содержали равное количество изображений. Для каждого из изображений вычислялись среднеквадратическая ошибка, средняя абсолютная ошибка и взвешенная среднеквадратическая ошибка по отношению к сжатой версии. Сжатие проводилось по алгоритму JPEG2000 при коэффициенте сжатия 95, 90, 85, 80, 75, 70, 65, 60, 55, 50 %.

В ходе обучения векторы характеристик обрабатывались для получения оптимальной проекции на двумерное пространство характеристик. Затем обученный байесов классификатор разбивал тестовое множество на три класса (нет внедрения, встроено с низкой плотностью и встроено с высокой плотностью). Графики рассеяния показаны на рисунке 1.11.

В результате 16,7 % чистых изображений были ошибочно приняты за стегограммы и 12 % стегограмм были приняты за пустые контейнеры.

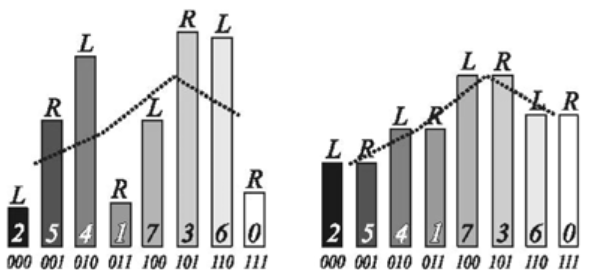


*Рисунок 1.11 - Результаты анализа методом М. Селика и др. факта внедрения за счет ШПС*

Из результатов проведенных авторами экспериментов видно, что разработанный ими способ стеганографического анализа уступает по показателям вероятности пропуска и вероятности ложной тревоги некоторым другим способам и, очевидно, нуждается в некоторой доработке.

### 1.2.4. Использование для стеганоанализа НЗБ статистики пар значений

Еще одним из известных методов стеганоанализа НЗБ является анализ статистики пар значений. Метод основан на поиске закономерности в вероятностях появления значений яркости в естественных изображениях и изображениях со внедренной информацией. При замене младшего бита цветовой компоненты очередного пиксела изображения на очередной бит предварительно зашифрованной или сжатой информацией значение яркости пикселя модифицированного изображения либо равно значению яркости пикселя контейнера, либо изменяется на единицу с вероятностью ~1/2. Для поиска следов внедрения был предложен метод анализа закономерностей в частотах появления «соседних» значений яркости. Такие пары значений («Pair of Values») различаются только значением НЗБ. Значение яркости, двоичное представление которого заканчивается нулевым битом, назовем «левым» (*L*), а соседнее с ним значение яркости, двоичное представление которого заканчивается единичным битом, – «правым» (*R*). Пусть цветовая гамма исходного контейнера включает 8 цветов. Следовательно, при внедрении сообщения в НЗБ цветовой компоненты пикселов необходимо исследовать статистические характеристики в 4 парах номеров цвета. На рисунке 1.12 показано типичное распределение пар значений для «пустого» изображения (слева) и изображения, содержащего внедренную информацию (справа).



*Рисунок 1.12 - Изменение распределения пар значений при внедрении ЦВЗ*

На диаграмме хорошо видно, что вероятности появления левых и правых номеров цвета в естественных контейнерах существенно различаются между собой во всех парах, а в изображении с внедренной информацией эти вероятности равны. Это является признаком наличия скрываемой информации. Заметим, что среднее значение вероятностей для каждой пары в модифицированном изображении не изменилось по сравнению с исходным контейнером (показано на рисунке 1.12 пунктирной линией).

При замещении битами внедряемого сообщения младших битов яркостной компоненты пикселов контейнера изображения проявляются аналогичные статистические различия. Степень различия между вероятностными распределениями элементов естественных контейнеров и изображений с внедренной ин формацией может быть использована для оценки вероятности наличия внедренной информации в изображении. Данную вероятность удобно определять с использованием критерия согласия хи-квадрат. Согласно критерию, определяется, насколько распределение исследуемой последовательности близко к характерному для изображений с внедренной информацией распределению. В исследуемой последовательности подсчитывается, сколько раз *i* ее элемент *x* принял рассматриваемые значения при общем количестве элементов *k*.

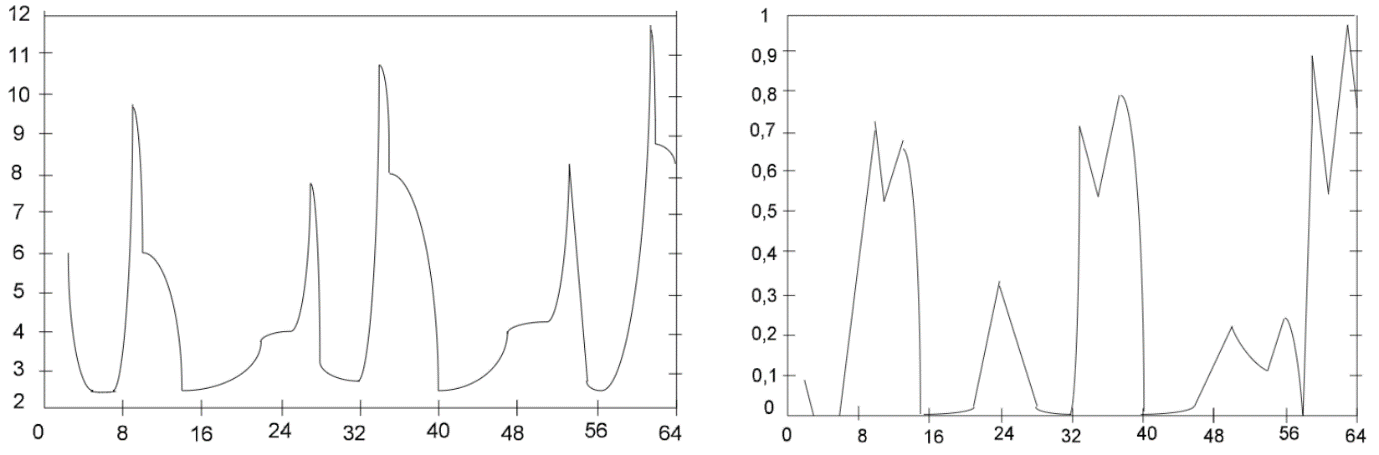
### 1.2.5. Использование для стеганоанализа НЗБ вейвлет декомпозиции

В работе представлен еще один эффективный способ обнаружения стеганографии в НЗБ при отсутствии информации об используемом алгоритме стеганографии. В этом случае преимущество получают методы универсального стеганоанализа, построенные на основе классификации признаков.

В работе универсальных методов стеганоанализа можно выделить этап получения векторов признаков для стегоконтейнера, удовлетворяющих требованиям репрезентативности и контекстной независимости, и их последующую классификацию.

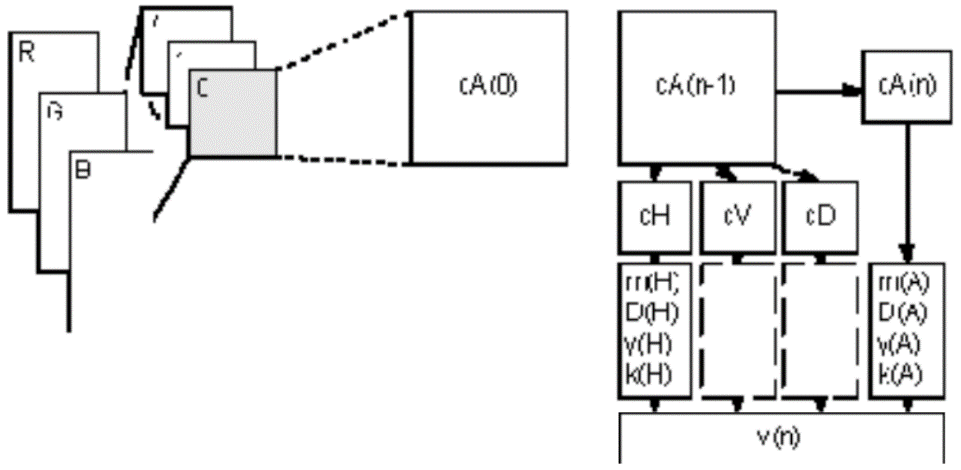
В ранее опубликованной работе автора вектор признаков изображения формировался применением последовательной вейвлет-декомпозиции исходного изображения и вычислением статистических оценок коэффициентов.

При последующем анализе обучающей и тестовой выборок была установлена существенная зависимость оценок статистических характеристик от контекстно обусловленного уровня дисперсии оценок и отсутствие характерных особенностей, присущих классам стегоконтейнеров (рисунок 1.13).



*Рисунок 1.13 - Значения математического ожидания (а) и дисперсии (б) элементов вектора признаков, соответствующие незаполненным и заполненным стегоконтейнерам*

В представленной модификации метода для обеспечения контекстной независимости вейвлет преобразование первого уровня применяется к битовой плоскости анализируемого цветового канала. Далее производится последовательная четырехуровневая декомпозиция низкочастотной полосы в соответствии с рисунком 1.14.



а б

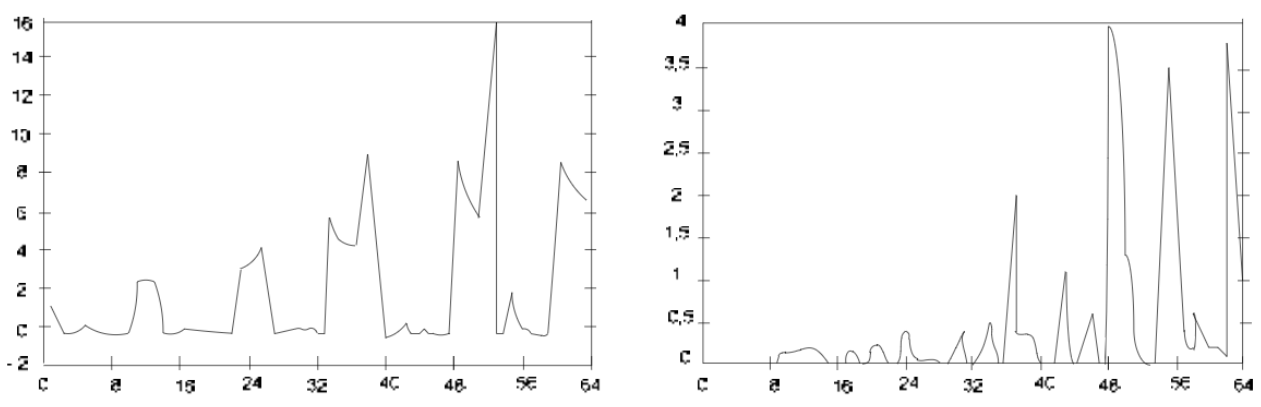
*Рисунок 1.14 - Получение битовой плоскости для последующей декомпозиции (а) и формирование вектора признаков v на n-этапе (б)*

Каждый уровень декомпозиции строится на основе низкочастотной субполосы предыдущего уровня. После проведения декомпозиции на фиксированную глубину создается статистическая модель на основе оценок математического ожидания, дисперсии, асимметрии и эксцесса для каждой из субполос всех уровней разложения, включая низкочастотные.

После формирования векторов признаков производится нормирование элементов *v\*i,j* векторов в соответствии с выражением

,

На рисунке 1.15 приведены графики значений математического ожидания и дисперсии элементов вектора признаков, соответствующие незаполненным и заполненным стегоконтейнерам после процедуры логарифмирования. Дисперсия информативных признаков начальных уровней понизилась при увеличившейся разности признаков конечных уровней.



а б

*Рисунок 1.15 - Значения математического ожидания (а) и дисперсии (б) элементов вектора признаков, полученного модифицированным методом декомпозиции*

Классификация нормированных векторов производится на основе трехслойной нейросети прямого распространения с обратным распространением ошибок. Первый слой содержит 80 нейронов с линейной функцией активации. Второй и третий слои содержат по 80 и 2 нейрона с тангенциальными и линейными функциями активации соответственно.

Обучение нейросети производилось на основе базы изображений, содержащей 150 растровых 24битовых изображений формата BMP 256×256 пикселов, имеющих аналоговую природу. Формирование стегоконтейнеров из изображений производилось внедрением псевдослучайной битовой последовательности в НЗБ плоскости цветовых каналов. Для оценки чувствительности метода были получены векторы признаков, из которых были сформированы обучающее множество (по 100 векторов) и тестовое множество (50 векторов). В процессе обучения на вход нейросети последовательно предъявлялись векторы, принадлежащие первому и второму множествам. Векторы *ti* целевого множества *T* были определены в соответствии с условием:

где *xi* – вектор признаков изображения из обучающего множества, предъявляемый нейросети; *Ce* – класс векторов-признаков изображений, не содержащих встроенных данных; *Ch* – класс векторов-признаков изображений, содержащих встроенные данные.

Результаты проведенных экспериментов (табл. 1.2) показали, что данный метод обнаруживает факт наличия внедренной информации при использовании свыше 2 % от максимальной емкости плоскости наименее значимых битов в полноцветных изображениях размерности 256×256 пикселов. Достигаемая точность классификации для объема внедренной в цветовой канал информации 2 % от максимума соответствует 80 %.

Таблица 1.2 - Зависимость точности классификации от объема внедренной информации, %

|  |  |  |  |
| --- | --- | --- | --- |
| Встроенная информация | Верное обнаружение | Ошибка 1-ого рода | Ошибка 2-ого рода |
| 100-30 | 100 | 0 | 0 |
| 10 | 98 | 2 | 0 |
| 5 | 98 | 1 | 1 |
| 2 | 80 | 12 | 8 |

В качестве проверки на возможность использования предлагаемым методом особенностей форматов файлов было произведено обучение нейросети на обнаружение внедренной информации в 100 стегоконтейнерах, являющихся изображениями формата BMP и полученных конвертированием из формата JPEG. При этом был зафиксирован уровень ложных обнаружений 2 и 0 % пропусков. Необходимо отметить, что использование контейнеров, исходные версии которых были получены декомпрессией форматов с потерей качества (JPEG, MPEG) без последующей обработки, ведет к снижению скрытности стеганографического канала и обнаруживается как методами универсального стеганоанализа, так и аналитическими методами.

Таким образом, применение рассмотренного метода стеганоанализа, основанного на последовательной вейвлет-декомпозиции битовой плоскости НЗБ-битов с последующей нейросетевой классификацией признаков, позволяет обнаруживать изменения свыше 1 % от максимальной емкости битовой плоскости полноцветных контейнеров-изображений размерности с 256×256 пикселов, а также эффективно использовать метод для обнаружения информации, скрытой в контейнерах, полученных преобразованием из формата JPEG.

### 1.2.6. Стеганоанализ НЗБ на основе статистики числа переходов

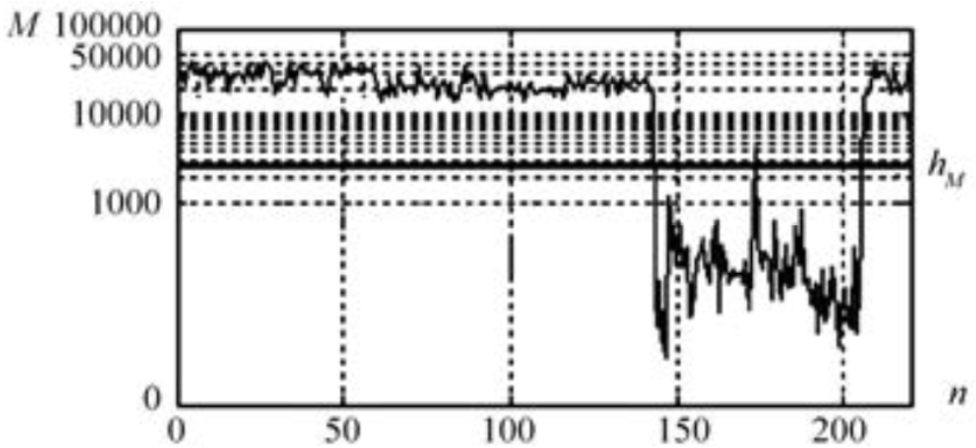
Между младшими битами имеются существенные корреляционные связи, которые могут нарушаться при внедрении данных. Зависимость между битами в соответствующих байтовых разрядах отсчетов имеет марковский характер. При этом параметры зависимости определяются номером разряда. Число переходов в потоке НЗБ из 0 в 0, из 0 в 1, из 1 в 0, из 1 в 1 различно для файла-контейнера и файла, содержащего встроенную информацию. При стегоскрытии данных с некоторой вероятностью изменяется вид закона распределения НЗБ. Как правило, распределение младших битов приобретает случайный характер, соответственно частоты переходов в потоке НЗБ отсчетов звукового файла будут примерно одинаковы в тех областях файла, где содержится скрытая информация.

Статистический критерий для оценки частот переходов битовых значений можно сформулировать следующим образом: анализируемый файл разбивается на *K* блоков одинаковой длины, выбирается некоторое пороговое значение *hM*. Далее вычисляются значения статистик

*,*

*j = 1, 2, …, k*

*μi* – количество переходов в потоке НЗБ из *i* в *j* (*i*, *j* = 0,1). Если, *j*<*hM*, считается, что в *j*-м блоке содержится скрытая информация. На рисунке 1.12 представлен результат применения о пи санного метода. На практике данный критерий способен выявлять стеганографическое скрытие при последовательной модификации младших битов, когда наблюдается резкий скачок статистики, *jM* в случае же псевдослучайного выбора отсчетов сигнала для внедрения в них стегоданных данный статистический критерий не способен выявить факт информационного скрытия.

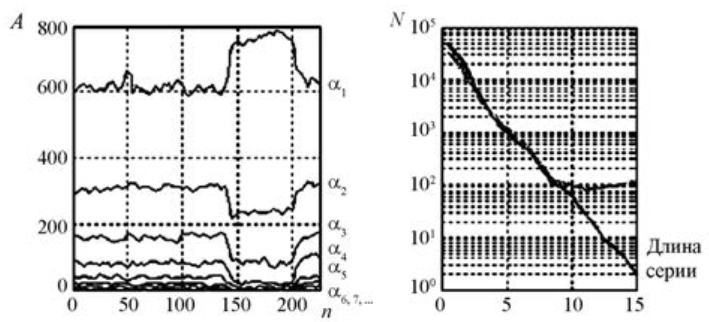


*Рисунок 1.12 - Статистический критерий оценки частот битовых переходов НЗБ звукового файла: по оси X – номера блоков, на которые разбивается файл, по оси Y – значение статистики, hM – пороговое значение*

Как показали результаты экспериментов, данный критерий можно использовать для формирования векторов входных воздействий нейронной сети. При отсутствии стеганографического скрытия значения *jM* будут колебаться относительно некоторого среднего значения и это не приведет к возрастанию числа ложных тревог, а резкое изменение статистик *jM* отразится на выходных данных нейросети, что и будет зафиксировано.

### 1.2.7. Стеганоанализ НЗБ на основе статистики длин битовых серий

Под *α* серией длины *n* понимается последовательность, состоящая из одинаковых бит вида *α α ... α*, где *α*∈ {0,1}, *n* = 1, 2, ..., *N*. Количество *α* серий большой длины является более устойчивой характеристикой, чем число битовых переходов в младших разряда *х*. На рисунке 1.13,а приведены графики частот появления 0-х *α* серий различной длины (для 1-х аналогично) в зависимости от их расположения в файле-контейнере. По оси абсцисс отложены номера блоков, на которые разбивается звуковой файл, по оси – ординат частоты α серий для каждого блока. Видно, что в областях файла, содержащих скрытые данные, увеличиваются частоты для α серий единичной длины и одновременно происходит уменьшение количества α серий длины больше 1. На рисунке 1.13,б представлены кривые распределения длин серий 0 бит для незаполненного (серая кривая) и заполненного контейнера (черная кривая).



а б

*Рисунок 1.13 - а – график и частот появления 0-х α-серий различной длины (для 1-х аналогично) в зависимости от их расположения в файле контейнере: по оси X – номера блоков, на которые разбивается звуковой файл, по оси Y – частоты αсерий для каждого блока; б – кривые распределения длин серий 0 бит для незаполненного (серая кривая) и заполненного контейнера (черная кривая)*

На практике данный критерий устойчиво показывает статистически значимые различия в случае слабо выраженной марковской зависимости в НЗБ контейнера. В то же время он не позволяет фиксировать факт стеганографического скрытия при незначительных объемах встраиваемых данных и их «рассеивании » по всему звуковому файлу.

### 1.2.8. Критерии информативности системы признаков

Одной из важнейших задач, возникающих при стеганоанализе изображений, является задача выбора признаков или исходного описания анализируемых изображений. Ниже рассматриваются вопросы, связанные со способами оценки качества той или иной системы признаков. При исследовании этих вопросов устанавливается связь между понятиями достаточной статистики, количества информации и вероятностью правильного распознавания.

Все признаки можно разделить на первичные и вторичные. Первичные признаки представляют собой результаты некоторых прямых измерений на изображении, т. е. являются функционалами от дискретного распределения яркости *v*(*x*, *y*) каждого цвета. Множество прямых измерений можно считать заданным. Можно считать, что для изображений это множество совпадает с множеством линейных функционалов от распределения яркостей с неотрицательными весовыми функциями для каждой цветовой составляющей. Неотрицательность весовой функции означает, что в любой точке изображения при увеличении яркости результат измерений не может уменьшаться.

Возникает вопрос, нельзя ли получить более экономное описание изображения с помощью вторичных признаков, число которых или общий объем информации будет меньше, а вероятность правильной классификации *pn* не уменьшится. При этом необходимо принимать во внимание сложность вычисления вторичных признаков, а также сложность решающей функции, для которой вторичные признаки будут аргументами. Но это означает минимизацию всего алгоритма классификации в целом. Что при таком подходе считать вторичными признаками и как их выбирать остается неясным. Поэтому необходимо ввести дополнительный критерий качества вторичных признаков и выбирать систему c помощью этого критерия. При этом желательно показать связь этого критерия с вероятностью ошибки классификации и со сложностью алгоритма классификации в целом. При разработке словаря признаков имеется ряд ограничений:

1. в словарь могут быть включены только те признаки, относительно которых может быть получена априорная информация, достаточная для описания классов на языке этих признаков;
2. с другой стороны, в словарь нецелесообразно включать все признаки, о которых есть априорная информация, так как они могут быть малоинформативными;
3. сложность вычисления признаков.

В ряде частных случаев можно показать монотонное упрощение алгоритма стеганоанализа при преобразовании признаков, так как число признаков уменьшается, сложность вычисления признака остается неизменной, а решающее правило с использованием меньшего числа признаков упрощается. При этом необходимо уметь оценивать достижимую вероятность правильной классификации для заданной системы признаков и ее изменение для того или иного преобразования признаков.

В основу критерия информативности системы признаков целесообразно положить понятие достаточной статистики. При этом необходимо исследовать взаимосвязь различных критериев, в частности критериев условной энтропии и вероятности ошибки.

Переходя от первичных признаков к вторичным, мы осуществляем преобразование исходного описания *v* в некоторое другое описание *u*. Преобразованное описание представляет собой функцию от исходного описания, или статистику. При этом многим различным значениям *v* сопоставляется одно значение *u*, за счет чего и происходит упрощение. Зная только преобразованные описания *u* не всегда можно различать исходные описания *v*, их породившие.

Рассмотрим условия, при которых такое объединение различных описаний vне приводит к ухудшению классификации, т. е. не увеличивает риска. Рассмотрим случай байесовского риска. Для сохранения функции риска необходимо и достаточно, чтобы значениям *v1* и *v2*, которым соответствуют различные решения , соответствовали бы также различные значения статистики . Так как выбор оптимального решения зависит от вида матрицы штрафов, то одно и то же преобразование может удовлетворять указанному требованию при одной матрице штрафов и не удовлетворять при другой.

Выбор оптимальных байесовских решений при любой матрице штрафов производится на основе апостериорных вероятностей *p*(*k*/*v*) или *p*(*k*/*u*). Поэтому для того, чтобы найти оптимальные решения при любой матрице штрафов, достаточно различать значения *v*, которым соответствуют различные апостериорные распределения. Статистика *u* позволяет получить при любой матрице штрафов то же значение функции риска, что и исходное описание *v*, если объединяются только такие значения *v*, для которых апостериорные распределения тождественны. При таком преобразовании описаний апостериорное распределение сохраняется: *p*(*k*/*u*(*v*) = *p*(*k*/*v*), *k* = 1, 2, …, *n*. В этом случае статистика *u*(*v*) называется достаточной статистикой.

В общем случае, когда некоторое исходное описание *v* тем или иным способом преобразуется в другое описание *u*, мерой недостаточности служит разность условных энтропий:

где *K* – общее число классов. Эта величина может быть проинтерпретирована как потеря информации о классе *k* при преобразовании описания *v* в новое описание *u*.

Известна теорема, утверждающая, что равенство условных энтропий является необходимым и достаточным условием достаточности статистики.

Различимость двух классов сигналов на основании данного описания v характеризуется величиной, называемой дивергенцией, или расстоянием между гипотезами о принадлежности к классу. Дивергенция имеет вид:

*.*

Для неразличимых по v классов, когда *,* дивергенция равна нулю. В случае же безошибочно разделимых классов, когда для любого описан и я, по крайней мере, одна из вероятностей равна нулю, дивергенция бесконечна, т.е. выражение для дивергенции терпит разрыв. Известно, что когда дивергенции *divv* и *divu* конечны, то их равенство является необходимым и достаточным условием того, что статистика u является до статочной по отношению к v.

Дивергенция, введенная Кульбаком для двух классов, может быть обобщена на любое число классов. Воспользуемся в качестве необходимой и достаточной статистики совокупностью отношений вероятностей *p*(*v/k*) к одной из них, например к первой: , *k* = 1, 2, …, *n*.

Такая совокупность может рассматриваться как обобщение отношения правдоподобия на случай n классов. Можно показать, что выражение для дивергенции, симметричное относительно всех классов, выглядит следующим образом:

Недостаток использования дивергенции Кульбака как меры недостаточности состоит в том, что она непригодна, когда одна из вероятностей равна нулю.

Как было указано, изменение условной энтропии класса при переходе от одного описания к другому служит мерой недостаточности последнего и характеризует снижение надежности стегоалгоритма. Однако более целесообразной характеристикой надежности является вероятность ошибки.

Под вероятностью ошибки для данного набора признаков или описания *v* можно понимать ту минимальную вероятность ложного решения, которую может обеспечить оптимальный классификатор, оперирующий описанием *v*. Поскольку вероятность ошибки – это лишь частное значение функции риска, за энтропией остается важная роль величины, характеризующей недостаточность описания.

Функциональной зависимости между энтропией и вероятностью ошибки нет. Однако при данной энтропии вероятность ошибки может изменяться только в определенных пределах. Точно также при заданной вероятности ошибки энтропия лежит в определенных границах. Эти границы определяются приведенными ниже выражениями.

Максимальная при данной вероятности ошибки *P* энтропия определяется зависимостью

Минимальное значение частной энтропии при фиксированном значении vопределяется выражением:

где целое число m такое, что

Подводя итог, мы проанализировали классификацию методов стеганоанализа и стеганоанализ наименьших значащих бит. Вышеупомянутая информация ляжет в основу тестовой части.

Список используемых источников

1. АВТОР??Стеганографические системы. Цифровые водяные знаки: учебно-методическое пособие / В. Г. Грибунин, В. Е. Костюлков, А. П. Мартынов [и др.]; под редакцией В. Г. Грибунин. – Саратов: Российский федеральный ядерный центр – ВНИИЭФ, 2016. – 210 с.

-URL: <http://www.iprbookshop.ru/60865.html> (дата обращения 26.10.2020). – Текст: электронный.

1. ГОСТ 7.32-2017. Отчет о научно-исследовательской работе. Структура и правила оформления. –URL: <http://docs.cntd.ru/document/1200157208> (дата обращения: 26.10.2020). –Текст: электронный.
2. ГОСТ 19.701-90. Единая система программной документации. Схемы алгоритмов, программ, данных и систем. –URL: <http://www.gostrf.com/normadata/1/4294848/4294848992.pdf> (дата обращения: 26.10.2020). –Текст: электронный.