МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теоретических основ компьютерной безопасности и криптографии

# Стеганография и стегоанализ графических файлов

АВТОРЕФЕРАТ
дипломной работы

студента 6 курса 631 группы
специальности 10.05.01 Компьютерная безопасность
факультета компьютерных наук и информационных технологий
Ущаповского Алексея Николаевича

Научный руководитель
доцент _____ И. Ю. Юрин
22.01.2022 г.

Заведующий кафедрой
д. ф.-м. н., доцент _____ М. Б. Абросимов
22.01.2022 г.

Саратов 2022

# ВВЕДЕНИЕ

Развитие компьютерных методов обработки информации дало возможность значительно повысить уровень обеспечения информационной безопасности. Существенных преимуществ в этом направлении удалось достичь с помощью применения современных криптографических методов. Но в целом ряде задач информационной безопасности их оказывается недостаточно, так как они не делают возможным скрыть факт присутствия или передачи информации. К числу эффективных современных методов защиты информации относятся методы компьютерной стеганографии.

Стеганография является наукой, уходящей корнями в древние времена. В наши дни из-за бурного развития вычислительной техники и новых каналов передачи информации возникли новые стеганографические методы, в основу которых легли особенности представления информации в компьютерных файлах, вычислительных сетях и т.п. Методы современной компьютерной стеганографии используются в области военной и правительственной связи, защиты авторских прав, для решения задач обеспечения информационной безопасности.

Вместе с созданием новых стеганографических методов скрытия информации не менее важной является разработка алгоритмов современного стегоанализа.

Основная задача стегоанализа — установление факта присутствия в контейнере скрытой информации. В ходе решения данной задачи стегоаналитиком применяются разные алгоритмы анализа. Но при попытке автоматизации процесса на множестве контейнеров появляется проблема выбора алгоритма анализа, так как реализации разных алгоритмов могут давать противоречащие результаты, что предопределено неравенством вероятностей возникновения ошибок распознавания для этих реализаций. Также немалую сложность дает вопрос выбора исходных данных для анализа. Уменьшение объема анализируемых данных потенциального контейнера ведет к увеличению вероятности возникновения ошибки первого рода (выявление скрытой

информации в пустом контейнере), что также повышает вероятность возникновения ошибок второго рода (прием заполненного контейнера за пустой). Переработка правил выборки анализируемых данных приводит к увеличению числа ошибок распознавания.

В настоящее время для решения задачи стегоанализа и дальнейшего извлечения скрытой информации нужен комплексный подход, который позволит выделить на множестве результаты, способствующие минимизации вероятности ошибки второго рода при установленном уровне вероятности ошибки первого рода.

Актуальной проблемой безопасности современных компьютерных сетей является борьба со скрытой передачей информации. Наиболее значимой является разработка алгоритмов и программ современного стегоанализа.

Целью дипломной работы являются рассмотрение существующих методов стеганографии и стегоанализа, а также практическая реализация гистограммного анализа для графических файлов форматов BMP, PNG, JPEG.

Задачи дипломной работы:

- изучить графические форматы BMP, PNG, JPEG. BMP будет рассмотрен для сравнения с другими форматами;
- рассмотреть основные методы скрытия информации (стеганография) и анализа ее присутствия (стегоанализ);
- проанализировать существующие программы стегоанализа;
- разработать собственную программу для выявления скрытой информации в форматах BMP, PNG, JPEG с помощью гистограммного анализа.

Дипломная работа состоит из введения, 5 разделов, заключения, списка использованных источников и 1 приложения. Общий объем работы — 79 страниц, из них страниц 56 — основное содержание, включая 8 рисунков и 21 таблицу, список использованных источников из 32 наименований.

## КРАТКОЕ СОДЕРЖАНИЕ

Первый раздел дипломной работы «Стеганография и стегоанализ графических файлов» содержит описание графических форматов BMP, PNG, JPEG.

BMP (Bitmap Picture) — формат хранения растровых изображений, созданный компанией Microsoft.

Данные в BMP представляют собой 3 базовых блока разного размера:

1) заголовок из структур BITMAPFILEHEADER и BITMAPINFO. Последняя включает в себя информационные поля; опциональные битовые маски для извлечения значений цветовых каналов; опциональную таблицу цветов;

2) опциональный цветовой профиль;

3) пиксельные данные.

Пиксельные данные могут располагаться в произвольной позиции в файле (она указана в структуре BITMAPFILEHEADER).

PNG (Portable Network Graphics) — растровый графический формат, поддерживающий сжатие без потерь данных.

Данные в PNG представляют собой следующие блоки:

- заголовок файла;
- фрагменты (chunks).

Фрагмент IDAT содержит фактические данные изображения. Пиксели всегда упаковываются в строки без ненужных битов между пикселями. Пиксели размером меньше байта упакованы в байты с крайним левым пикселем в битах старшего порядка байта, крайним правым в битах младшего порядка. Разрешенные разрядности и типы пикселей ограничены.

Строки всегда начинаются с границ байтов. Когда пиксели имеют менее 8 бит, а ширина строки сканирования неравномерно делится на количество пикселей на байт, младшие биты в последнем байте каждой строки сканирования теряются впустую.

Основой JPEG формата является набор маркеров. Первый байт каждого маркера имеет значение 0xFF. Следом за ним идет второй байт, определяющий тип маркера. Стоит заметить, что маркеры, которые определяют структуру файла JPEG, не могут иметь подмаркеров.

В таблице 1 указаны маркеры, встречающиеся практически всегда в любом файле формата JPEG и требующие обязательной обработки.

Таблица 1 — маркеры формата JPEG, требующие обязательной обработки

| Тип маркера | Идентификатор | Обозначение стандарта | Определение |
|---|---|---|---|
| **$SOF_0$** | 0xC0 | Baseline DCT | Начало кадра, базовый метод |
| **$SOF_1$** | 0xC1 | Extended sequential DCT | Начало кадра, расширенный последовательный метод |
| **$SOF_2$** | 0xC2 | Progressive DCT | Начало кадра, прогрессивный метод |
| **DHT** | 0xC4 | Define Huffman table(s) | Определение таблиц Хаффмана |
| **SOI** | 0xD8 | Start of image | Начало изображения |
| **EOI** | 0xD9 | End of image | Конец изображения |
| **SOS** | 0xDA | Start of scan | Начало сканирования |
| **DQT** | 0xDB | Define quantization table(s) | Определение таблиц квантования |

Процесс сжатия по алгоритму JPEG содержит следующие этапы:

- преобразование изображения в оптимальное цветовое пространство;
- субдискретизация компонентов цветности путем усреднения групп пикселей;
- использование дискретных косинус-преобразований для уменьшения избыточности данных изображения;
- квантование каждого блока коэффициентов ДКП с применением весовых функций, оптимизированных с учетом визуального восприятия человеком;
- кодирование результирующих коэффициентов (данных изображения) с помощью алгоритма Хаффмана для избавления от избыточности информации.

Стоит обратить внимание на то, что декодирование JPEG осуществляется в обратном порядке.

Второй раздел работы посвящен стеганографии, основным понятиям и методам скрытия информации.

Стеганография — наука о скрытии передаваемых данных в некотором контейнере таким образом, чтобы спрятать сам факт передачи информации.

Одним из основных методов скрытия информации является LSB-метод, заключающийся в скрытии сообщения в младших битах данных. Для BMP или PNG изображений сообщение прячут в пиксельных данных или палитре. Для JPEG изображений существует метод, базирующийся на скрытии данных в коэффициентах дискретного косинусного преобразования (ДКП). Он представляет собой разновидность LSB-метода. Этот алгоритм скрытия данных заключается в изменении величин коэффициентов ДКП. Стоит отметить, что коэффициенты «0» и «1» неизменны — встраивание сообщения в них не представляется возможным.

В третьем разделе рассмотрен стегоанализ, основные понятия и несколько разновидностей атак на графические файлы.

Стегоанализ — наука об определении факта внедрения информации в контейнер. Для обнаружения скрытой информации используются стеганографические атаки.

Аналитик пытается взломать стеганографическую систему, то есть найти факт передачи сообщения, извлечь сообщение и или изменить сообщение, или запретить пересылку сообщения.

Одной из статистических атак на файлы является гистограммный анализ. Данный алгоритм представляет собой атаку на основании известного заполненного LSB-методом стегоконтейнера.

Равномерное встраивание информации уменьшает разницу между частотами распределения соседних цветов, различающихся в наименьшем бите. Помимо этого, было определено, что в процессе скрытия LSB-методом сумма распределения частот соседних пар будет постоянной. На этих утверждениях и основывается алгоритм анализа с применением критерия $\chi^2$.

Если стегоконтейнер не является изображением с индексацией цветов, а JPEG-изображением, то вместо индексов цвета для анализа берут коэффициенты ДКП.

Четвертый раздел содержит описание и тестирование программ стегоанализа, а также ход подготовки данных для проведения исследований.

В ходе работы были рассмотрены программы стегоанализа. Для проведения исследования было отобрано чуть более 1000 «пустых» файлов для каждого из 3 форматов: BMP, PNG, JPEG. Далее каждый файл форматов BMP и PNG был обработан стеганографической утилитой Stegosuite, а файлы формата JPEG - программой JPHIDE. С их помощью были созданы стегоконтейнеры с заполнением от 10 до 100% от их объема. Рассмотрим несколько программ стегоанализа подробнее.

StegDetect является open-source программой, разработанной компанией Niels Provos. Данная утилита представляет собой автоматизированный инструмент для обнаружения стеганографического содержимого в изображениях. Она способна обнаруживать несколько различных стеганографических методов для встраивания информации в изображения JPEG.

StegExpose — это мультиплатформенный инструмент стегоанализа с открытым исходным кодом, специализирующийся на обнаружении LSB-стеганографии в изображениях без потерь, таких как PNG и BMP. Он имеет интерфейс командной строки и предназначен для массового анализа изображений, предоставляя при этом возможности создания отчетов с помощью CSV-файлов и более точной настройки параметров детектирования.

Пятый раздел посвящен программной реализации выявления стеганографии с помощью гистограммного анализа, в результате которой была написана программа StegAnalyze, способная обнаружить информацию, скрытую в файлах формата BMP, PNG, JPEG. Запуск программы производится из командной строки: java -jar StegAnalyze.jar [-h] -f [file/folder] -t [threshold] -o [outputFile].

Найдя графический файл формата BMP или PNG, программа производит следующие действия:

1) программа получает RGB представление графического файла. Также из этого файла программа извлекает точки, не относящиеся к однородной области. Однородными областями будем считать совокупность расположенных рядом точек, цвет которых не отличается более чем на 1 в RGB представлении;

2) далее утилита производит подсчет частот появления цветов и атаку хи-квадрат, которые можно разбить на следующие этапы:

- берется очередное значение цвета и пересчитываются частоты появления;
- рассчитываются ожидаемые (среднее арифметическое частот соседних значений) и фактические (первое из соседних значений) частоты для атаки хи-квадрат;
- полученные ожидаемые и фактические частоты атакуются методом хи-квадрат;
- результат атаки складывается с результатом, полученным от предыдущего значения;

3) анализируя максимальную из полученных сумм, программа делает вывод о заполнении контейнера.

Найдя графический файл формата JPEG, программа производит практически те же действия, только вместо значений цветов анализу подвергаются коэффициенты ДКП. При расчете пропускаются значения «0» и «1».

Полный код программы приведен в Приложении А.

```
MacBook-Pro-Aleksej:StegAnalyze_jar uschapovskiy$ java -jar StegAnalyze.jar -h

Пример запуска программы из командной строки: java -jar StegAnalyze.jar [-h] -f
[file/folder] -t [threshold] -o [outputFile]

Аргументы:
-h, отобразить сообщение с подсказкой
-f, файл/папка для проверки
-t, пороговый уровень, по умолчанию равен 0,20, должен быть не меньше 0 и не бол
ьше 1
-o, файл вывода, если аргумент не указан, то для вывода результатов работы прогр
аммы будет использован стандартный поток ввода/вывода, если указан, то будет соз
дан новый файл с указанным именем или перезаписан уже существующий
```

Рисунок 1 — возможности программы StegAnalyze.

Как можно увидеть из результатов тестирования StegAnalyze, представленных в таблице 2, утилита смогла обнаружить практически все заполненные стегоконтейнеры, также стоит отметить, что программа имеет хороший показатель «неверных» ответов для пустых контейнеров.

Таблица 2 — результаты тестирования программы StegAnalyze

| **Заполнение контейнера, %** | **Количество неверных решений** | | | | | |
|---|---|---|---|---|---|---|
| | **BMP** | | **PNG** | | **JPEG** | |
| | **Количество / Всего** | **%** | **Количество / Всего** | **%** | **Количество / Всего** | **%** |
| **0** | 112 / 1023 | 10.95% | 130 / 1021 | 12.73% | 103 / 1017 | 10.13% |
| **10** | 103 / 1023 | 10.07% | 126 / 1021 | 12.34% | 58 / 1017 | 5.7% |
| **20** | 68 / 1023 | 6.65% | 90 / 1021 | 8.81% | 37 / 1017 | 3.64% |
| **30** | 66 / 1023 | 6.45% | 89 / 1021 | 8.72% | 26 / 1017 | 2.56% |
| **40** | 54 / 1023 | 5.28% | 73 / 1021 | 7.15% | 23 / 1017 | 2.26% |
| **50** | 31 / 1023 | 3.03% | 38 / 1021 | 3.72% | 15 / 1017 | 1.47% |
| **60** | 19 / 1023 | 1.86% | 25 / 1021 | 2.45% | 12 / 1017 | 1.18% |
| **70** | 12 / 1023 | 1.17% | 17 / 1021 | 1.67% | 5 / 1017 | 0.49% |
| **80** | 0 / 1023 | 0% | 0 / 1021 | 0% | 0 / 1017 | 0% |
| **90** | 0 / 1023 | 0% | 0 / 1021 | 0% | 0 / 1017 | 0% |
| **100** | 0 / 1023 | 0% | 0 / 1021 | 0% | 0 / 1017 | 0% |

Сравнив разработанную программу с протестированными ранее, можно выделить следующие преимущества:

- работает с BMP, PNG, JPEG файлами;
- может быть запущена под любой ОС;
- имеет простой и понятный интерфейс на русском языке;
- предоставляет отличное качество выявления скрытой информации.

# ЗАКЛЮЧЕНИЕ

В современном мире стегоанализ представляет собой активно развивающееся направление в области информационной безопасности.

Одной из остро стоящих проблем развития современного общества является проблема обнаружения скрытой информации, выходом из этой ситуации служит применение данного направления.

На данный момент существует малое количество утилит, проводящих «качественный» стегоанализ графических файлов. Многие из них устарели и нуждаются в серьезных обновлениях, некоторые и вовсе исчезают из рабочего пространства. Поэтому данное направление требует развития и новых разработок в этой сфере.

В дипломной работе были рассмотрены графические форматы BMP, PNG, JPEG и их описания, стеганография и возможные методы скрытия информации, стегоанализ, проанализированы существующие программы стегоанализа, а также разработана программа для выявления скрытой информации в изображениях, указанных ранее форматов.