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

Стегоанализ видеофайлов

АВТОРЕФЕРАТ
дипломной работы

студента 6 курса 631 группы
специальности 10.05.01 Компьютерная безопасность
факультета компьютерных наук и информационных технологий
Ковалева Михаила Сергеевича

Научный руководитель
доцент _____ И. Ю. Юрин
21.01.2023 г.

Заведующий кафедрой
д. ф.-м. н., доцент _____ М. Б. Абросимов
21.01.2023 г.

Саратов 2023

ВВЕДЕНИЕ

Стеганография – это наука о скрытой передаче данных. Информация встраивается в различные цифровые носители информации, такие как изображения, аудио и видео, таким образом, чтобы внешний вид контейнера после встраивания не отличался от оригинала. Основной целью стеганографии является сокрытие факта обмена информацией.

С ростом популярности обмена видео файлов через Интернет, видео стало лучшим средством для встраивания стеганографии. Исследования показали, что в 2021 году видео-трафик составил 82% от всего интернет-трафика. Помимо роста популярности, видеофайлы, в качестве контейнера для сокрытия информации, имеют еще одно преимущество перед другими типами медиафайлов – наибольший объем. Таким образом в видеофайлах можно сокрыть наибольшее количество информации.

Видеофайл имеет различные компоненты, пригодные для встраивания скрытых сообщений: вектора движения, коэффициенты дискретного косинусного преобразования (ДКП), метаданные. В отличии от методов сокрытия в коэффициентах ДКП, которые схожи с методами стеганографии в изображениях, алгоритмы встраивания информации в вектора движения ориентированы только на стеганографию в видеофайлах.

Методы стегоанализа были разработаны для обнаружения существования скрытых сообщений в цифровых носителях информации. В качестве критерия для оценки эффективности работы методов стегоанализа используется вероятность обнаружения скрытого в контейнере сообщения.

Целью данной работы является изучение способов стеганографии и алгоритмов стегоанализа в видеофайлах, реализация алгоритма стегоанализа для обнаружения скрытой информации в видеофайлах для операционной системы Windows.

Для достижения данной цели, необходимо решить задачи:

- провести анализ методов стеганографии;
- провести анализ существующих алгоритмов стегоанализа;
- реализовать программу, определяющую наличие скрытой информации в видеофайлах.

В работе рассматривается статистический метод стегоанализа. Данный метод нацелен на обнаружение скрытых данных в векторах движения. Алгоритм подсчитывает совместное распределение разниц значений векторов движений в некоторой области. Анализ полученного распределения выявляет стеганографию, так как при изменении векторов движения неизбежно меняется совместное распределение разницы значений соседних векторов.

Дипломная работа состоит из введения, 4 разделов, заключения, списка использованных источников и 1 приложения. Общий объем работы – 63 страницы, из них 44 страниц – основное содержание, включая 25 рисунков и 2 таблицы, список использованных источников из 20 наименований.

КРАТКОЕ СОДЕРЖАНИЕ

В первом разделе дипломной работы рассматривается формат видеофайлов MPEG. Данный раздел содержит 4 подраздела.

$MPEG$ (Motion Picture Experts Group) – это группа экспертов, которая разрабатывает, обновляет и поддерживает стандарты сжатия цифровой аудио и видео информации. Группа MPEG разработала множество стандартов, наиболее важными из которых являются MPEG-1, MPEG-2 и MPEG-4. MPEG-1 – самый первый стандарт, который был разработан для сжатия аудио и видео объектов с последующей записью на компакт-диск. Стандарт MPEG-2 используется в телевизионном вещании и в DVD видео. Стандарт MPEG-4 обладает более высоким коэффициентом сжатия по сравнению с MPEG-2 и позволяет оперировать объектами (изображения, трёхмерные модели, текстовые данные).

Высокая степень сжатия в стандартах MPEG достигается с помощью методов сжатия информации с потерями. В данном стандарте для передачи пиксели кадров кодируются в цветовой схеме YC_bC_r с использованием цветовой субдискретизации в формате 4:2:0, а так же временная модель и вектора движения для компенсации движения между кадрами. Эти методы устраняют значительное количество межкадровой и внутрикадровой избыточности. Модель сжатия MPEG приведена на рисунке 1.

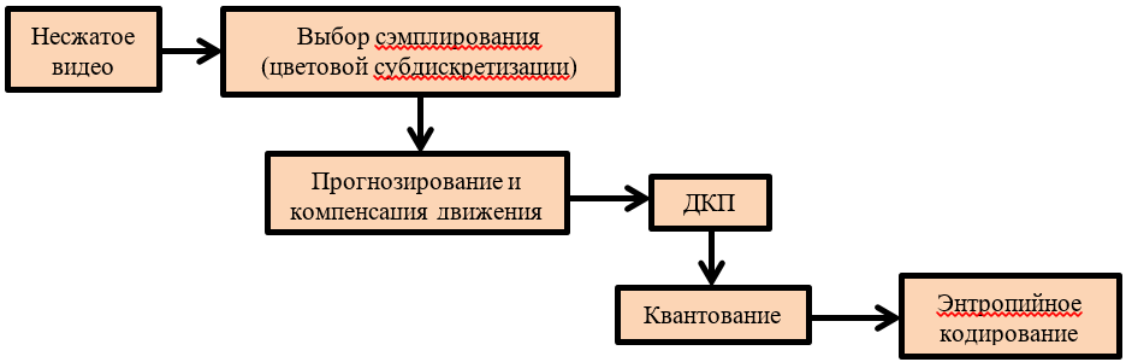

Рисунок 1 – Модель сжатия $MPEG$

Во втором разделе рассматриваются основные положения стеганографии, требования к методам стеганографии, виды стегоконтейнеров, области применения стеганографии, а также способы стеганографии в видеофайлах.

Стеганография – наука о скрытой передаче данных.

Так как большинство коммуникаций на сегодняшний день происходят в электронном виде, были достигнуты успехи в использовании цифровых мультимедийных сигналов в качестве транспортного средства для скрытой передачи информации. Эти сигналы, типичными представителями которых являются аудио, видео и изображения, являются контейнером.

Процесс заполнения контейнера информацией, которую требуется сокрыть, можно описать схемой, представленной на рисунке 2:

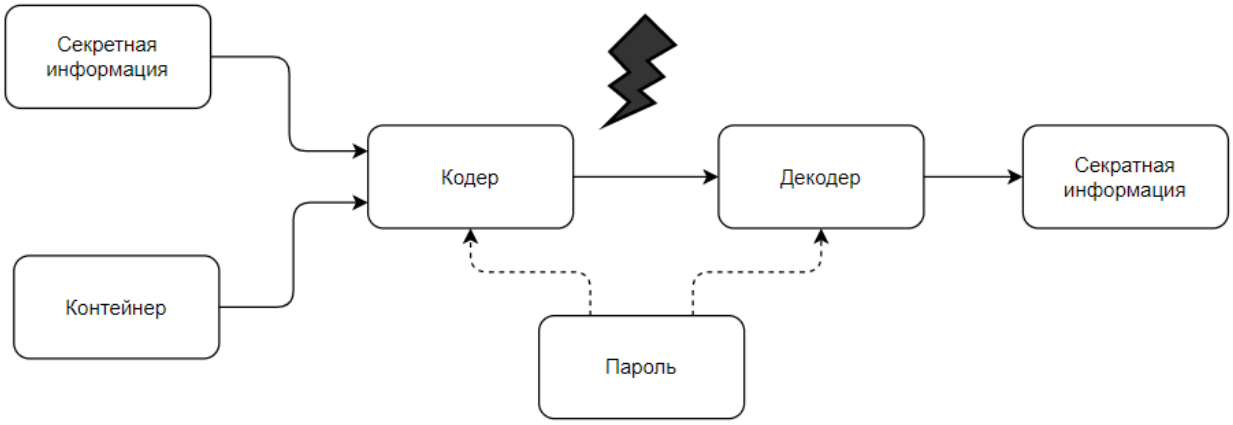

Рисунок 2 – Структурная схема стеганографии

Можно выделить 4 требования к методам стеганографии:

1) Прозрачность – пустой контейнер и контейнер с информацией неразличимы не вооруженным глазом, то есть человек, посмотрев на изображение или прослушав аудиозапись не сможет достоверно узнать о сокрытой в данном контейнере информации.
2) Устойчивость к случайным (или злонамеренным) искажениям в канале связи. В процессе передачи звук, изображение или другой контейнер может претерпевать различные преобразования, такие как: уменьшение

или увеличения объема, преобразование в другой формат, а также сжатие, с использованием алгоритмов сжатия с потерями.

3) Отсутствие маркировки (в худшем случае секретность маркировки).
4) Знание нарушителем факта наличия сообщения в каком-либо контейнере не должно помочь ему при обнаружении сообщений в других контейнерах.

В третьем разделе рассматривается понятие стегоанализа, атаки на системы скрытой передачи сообщений, а также понятие классификатора.

Стегоанализ – это наука о выявлении факта передачи скрытой информации в анализируемом сообщении. Методы стегоанализа так же могут использоваться для извлечения скрытых данных из стеганографического сообщения.

Можно выделить 2 основных метода атак на стеганограмму:

1) пассивные методы;
2) активные методы.

Пассивные методы могут лишь обнаружить факт сокрытия информации. Если существует возможность выявить факт наличия скрытой информации, то стеганографический алгоритм считается нестойким. Пассивные методы не позволяют принимающей стороне узнать о факте вмешательства в процесс передачи информации.

Активные методы стегоанализа более обширны. С помощью активных методов можно удалять или разрушать сообщение, скрытое в контейнере. В этом случае, принимающая сторона может узнать о факте вмешательства в процесс передачи информации.

Так же к активным методам стегоанализа относят создание ложных или поддельных стегоконтейнеров.

Обнаружения скрытых данных достаточно, чтобы помешать самой цели стеганографии, даже если секретное сообщение не извлечено. Обнаружение обычно осуществляется путем выявления некоторых характерных особенностей контейнера, которые изменены скрытыми данными.

Различают 2 основных подхода к стегоанализу:

1) Визуальные подход: визуальный анализ контейнера, например, рассмотрение каждого из 24-х однобитных компонент RGB-изображения и пытка найти визуальную разницу в этих однобитных изображениях.
2) Статистический подход: основная гипотеза данного подхода состоит в том, что некоторые статистические характеристики контейнера могут быть изменены в процессе встраивания. Следовательно, суть большинства стеганоаналитических методов заключается в разработке функции для выявления этих незначительных искажений. Обычно используется метод классификации по образцу, при котором извлекается набор признаков из контейнера без встроенных данных и контейнера, содержащего стеганограмму. Затем выбирается классификатор, который обучается отличать стегообъекты от обычных.

В работе для стегоанализа видеофайлов выбран пассивный статистический алгоритм, который рассчитывает совместную функцию распределения разности компонент соседних векторов движения. Пример соседних векторов движения показан на рисунке 3.

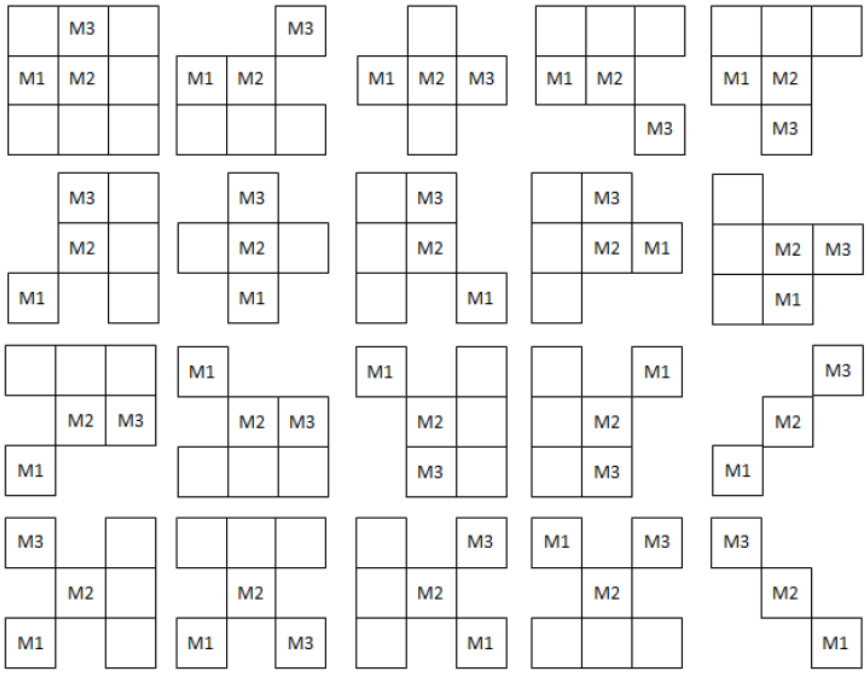

Рисунок 3 – 20 возможных комбинаций центрального макроблока и двух соседних макроблоков

Одновременно рассматривая 3 последовательно идущих кадра, можно получить 168 значений разности компонент вектора движения, из которых будет формироваться выборка для расчета совместной функции распределения. Значения, полученные из совместной функции распределения, будут формировать вектор признаков.

После того, как вышеупомянутый вектор был получен, с его помощью необходимо определить наличие или отсутствие скрытой информации в видеофайле. Эту задачу можно свести к задаче классификации.

Задача классификации состоит в определении к какому классу из, как минимум, двух изначально известных относится данный объект. Обычно таким объектом является вектор в n-мерном вещественном пространстве . Координаты вектора описывают отдельные аттрибуты объекта. Если классов всего два (есть стеганограмма / нет стеганографммы), то задача называется бинарной классификацией.

Набор векторов извлекается из видео без стеганограммы и со стеганограммой. Каждый вектор помечен для использования в качестве учебного примера для настройки классификатора.

Для классификации используется метод Random Forest.

Четвертый раздел посвящён программной реализации алгоритма поиска сокрытых данных в видеофайле. Программа для анализа написана на следующих языках программирования:

1) C++ с использованием библиотеки ffmpeg для низкоуровневой работы с видеофайлом в формате $MPEG$;
2) Python с использованием библиотеки sklearn для классификации векторов;
3) Java – статистические рассчеты, а так же пользовательский интерфейс, реализованный с помощью библиотеки JavaFX.

Результаты работы программы будем оценивать с использованием трех величин:

1) TP – вероятность определения наличия стеганограммы в видеофайле со встроенными данными;
2) TN – вероятность корректного определения файлов без стеганографии;
3) AR – точность определения обоих случаев.

Экспериментально полученные численные результаты представлены в таблице 1:

Таблица 1 – Результаты работы программы

BPF	TP, %	TN, %	AR, %
1500	91,4	89,7	90,55
1000	85,5	85,4	85,45
500	74,5	69,3	71,9

ЗАКЛЮЧЕНИЕ

В ходе работы была реализована программа для обнаружения стеганографии на основе векторов движения в видеофайлах формата MPEG.

В программе использовался алгоритм стегоанализа с использованием совместной функции распределения разности компонент соседних векторов движения. Одновременно рассматривая 3 последовательно идущих кадра, можно получить 168 значений разности компонент вектора движения, из которых будет формироваться выборка для расчета совместной функции распределения. Значения, полученные из совместной функции распределения, будут формировать вектор признаков, который будет подан на вход классификатору.

Для классификации использовался метод Random Forest, который устанавливал принадлежность вектора признаков к одному из двух классов: со скрытыми данными или без скрытых данных.

Поставленные задачи полностью решены.

Программный комплекс может применяться в учебных целях, а также в прикладных задачах, связанных с анализом стегосистем.