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

Анализ аудиофайлов формата mp3 на наличие сокрытой информации

АВТОРЕФЕРАТ
дипломной работы

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

Научный руководитель
доцент, к.п.н. _____ А. С. Гераськин
23.01.2021 г.

Заведующий кафедрой
д. ф.-м. н., доцент _____ М. Б. Абросимов
23.01.2021 г.

Саратов 2021

ВВЕДЕНИЕ

В связи с развитием и распространением компьютерных технологий, одним из самых удобных и часто используемых видов хранения информации является цифровое представление. Аудиофайлы являются одним из примеров такой формы. Возможность хранения на любом электронном носителе памяти, небольшой размер, а также возросшая популярность передачи через сеть Интернет делают аудиофайлы особенно привлекательными для использования в качестве стегоконтейнера.

К примеру, необходимо передать секретное сообщение между секретными базами. Для того чтобы злоумышленник не смог перехватить его и легко прочитать, могут быть использованы программы для встраивания сообщения в аудиофайл с музыкой или другого рода содержанием. Если преступнику всё-таки удастся захватить файл, то без специального программного обеспечения он не сможет понять, где спрятано сообщение, и как его извлечь из аудиофайла.

Однако и со своей стороны необходимо уметь проанализировать подобный перехваченный файл. На основе имеющегося элемента используют различные методы и признаки для проверки файла на наличие сокрытой информации, которые будут рассмотрены в данной работе. Необходимо отметить, что эта область является актуальной для исследования, поскольку для сокрытия сообщения, кроме стандартных методов, могут разрабатываться и вводиться в эксплуатацию новые и ещё не изученные прежде возможности. Это позволяет создавать и использовать новые алгоритмы для получения более точных сведений.

Целью дипломной работы является анализ аудиофайла на наличие сокрытой информации.

Задачи дипломной работы:

- изучение теоретических основ стеганографии;
- рассмотрение методов сокрытия информации в аудиофайлах;

- рассмотрение методов анализа аудиофайла на наличие сокрытой информации;
- создание программного продукта, реализующего анализ аудиофайла на наличие сокрытой информации;
- тестирование программного продукта.

Дипломная работа состоит из введения, 4 разделов, заключения, списка использованных источников и 1 приложения. Общий объем работы – 53 страницы, из них 41 страниц – основное содержание, включая 22 рисунка, список использованных источников из 21 наименования.

КРАТКОЕ СОДЕРЖАНИЕ

В первой главе «Теоретические основы стеганографии» рассматриваются основные понятия стеганографии. Стеганография один из эффективных способов сокрытия секретных данных. Описание стегосистемы и ее составляющих необходимо для понимания как формируется скрытый канал передачи информации. Стегосистема состоит из следующих компонентов: скрываемая информация, контейнер, кодирующее устройство, контейнер с информацией, декодирующее устройство, скрываемая информация в соответствии с рисунком 1.

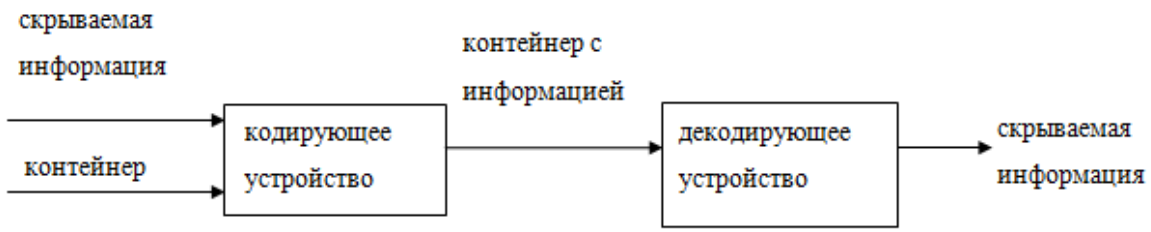

Рисунок 1 – Модель стеганографической системы

Скрываемая информация встраивается в объем цифрового контейнера, при этом в самом контейнере невозможно определить его начало и конец. Информация встраивается в контейнер с помощью прямого стеганографического преобразования, ее также можно восстановить путем обратного стеганографического преобразования в соответствии с рисунком 2.

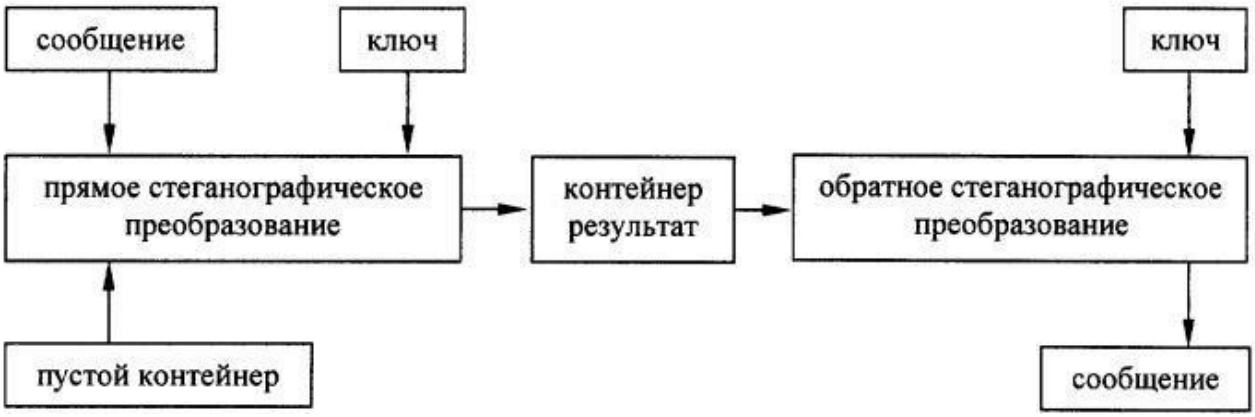

Рисунок 2 – Структура стеганографического преобразования

С помощью стегоключа предопределяется секретный алгоритм, который определяет порядок внесения сообщения в контейнер. Тип ключа определяет существование двух типов стеганосистем: с секретным ключом и с открытым ключом. Также можно определить стеганографическую стойкость – это определяется способностью противостоять попыткам нарушителя разрушить, исказить, удалить скрытно передаваемые сообщения, а также способность подтвердить или опровергнуть подлинность скрытно передаваемой информации. Существует процедура обнаружения факта сокрытия информация, которая называется стегоанализом. Также определяются основные положения и принципы современной компьютерной стеганографии

Во второй главе «Стеганографические методы» рассматриваются методы стеганографии в аудио. Глава начинается с описания того, каким основным требованиям должна отвечать система сокрытия данных в аудиофайле:

- быть стойкой к повсеместно используемым алгоритмам сжатия с потерями;
- не вносить в сигнал воспринимаемые человеческим слухом искажения;
- не вносить заметных изменений в статистику контейнера.

Далее идёт четыре подраздела, описывающие различные алгоритмы для встраивания секретного сообщения в файл. Для каждого метода определяются его преимущества и недостатки использования. Стоит отметить, что наиболее достойным и устойчивым к атакам является метод фазового кодирования, т.к. человеческое ухо воспринимает не сами значения фазы, а только их разность, что позволяет скрыть необходимую информацию в файле. Также заслуживающим внимания можно отметить метод расширения спектра, поскольку основным достоинством данного алгоритма является возможность передачи небольших объемов информации при высоком уровне устойчивости к искажениям.

В третьей главе «Методы анализа аудиофайла на наличие сокрытой информации» приводится метод, с помощью которого можно произвести анализ

аудиофайла на наличие сокрытой информации. Для детектирования секретной информации в аудиофайле необходимо проанализировать распределение значений квадрантов фаз гармоник. В случае если во фрейм было встроено сообщение, будет заметно преобладание некоторых квадрантов. Преимуществом данного метода является то, что при определённого порога преобладания для рейтинга пары квадрантов, обнаруживаются все вложения, созданные большинством существующих программ. В то же самое время его недостатком является ложное срабатывание на поврежденные файлы. Для предотвращения этого необходимо анализировать сами подозрительные данные, что является не менее трудной задачей. Далее речь идет о характерных признаках аудиофайлов, которые можно определить, как математические алгоритмы, благодаря которым можно извлекать полезную информацию из сигнала. Признаки извлекаются из временной или частотной области. Для временной области аудио признаки извлекаются на двух уровнях: краткосрочный (фреймовый) и долгосрочный. Перед вычислением любого признака в частотной области, сигнал должен быть преобразован в частотную область с помощью дискретного преобразования Фурье. Обратное преобразование Фурье можно использовать для преобразования сигнала обратно во временную область, но обычно не требуется в контексте извлечения признаков. Все признаки описаны в двух подразделах рассматриваемой главы. Приведено описание признаков, их формулы, а также преимущества и недостатки.

В четвёртой главе «Создание программного продукта для анализа аудиофайла на наличие сокрытой информации» описывается разработка и тестирование программного продукта для анализа аудиофайла на предмет сокрытой информации. При запуске программы открывается окно, показанное на примере рисунка 6, в меню которого имеются вкладки «Внедрение сообщения» и «Извлечение сообщения» для сокрытия и извлечения секретного сообщения из аудиофайла соответственно.

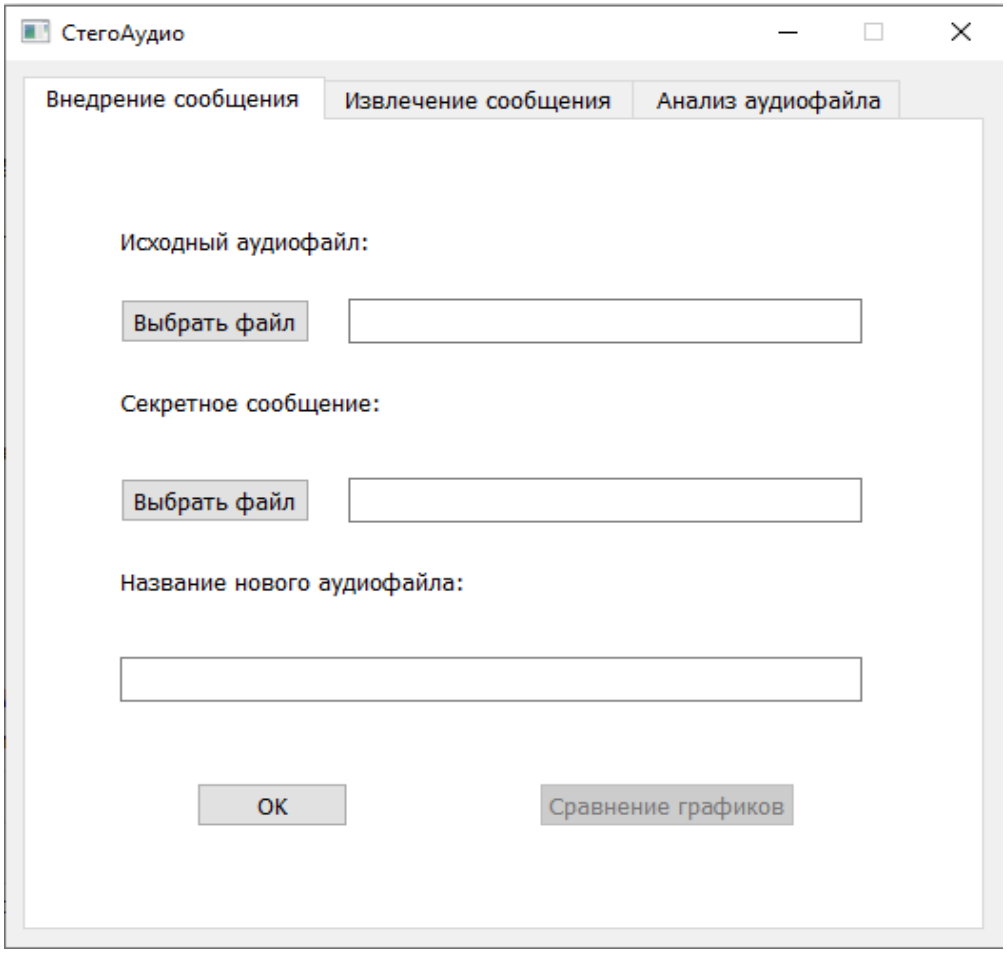

Рисунок 6 – Главное окно программы

Глава о создании программного продукта состоит из двух подразделов. В первом подразделе реализуется метод фазового кодирования. Приведен пример работы программы в виде скриншотов и описания к ним. Пошагово расписаны действия, приводящие к сокрытию и извлечению сообщения. Следует отметить, что в некоторых случаях возможно небольшое искажение, связанное с тем, что обратное преобразование Фурье не восстанавливается в точности первоначальные значения фаз и возможны небольшие отклонения по сравнению с исходными значениями фаз. Во втором подразделе описывается, как будет производиться анализ аудиофайла на наличие сокрытой информации, а также пошагово расписаны действия, приводящие к этому. Для определения возможных изменений в аудиофайле достаточно, чтобы хотя бы один из признаков (спектральный центроид или спад) нарушал определенные условия в соответствии с рисунками 21 и 22.

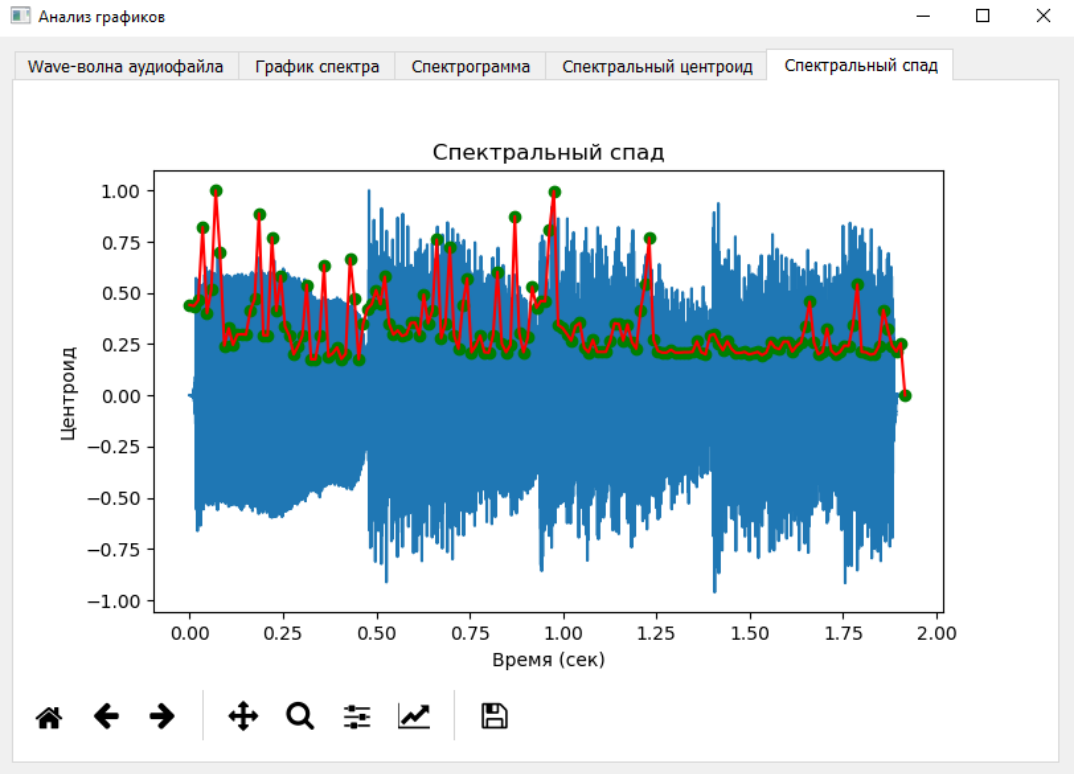

Рисунок 21 – График спектрального спада для аудиофайла с секретным сообщением

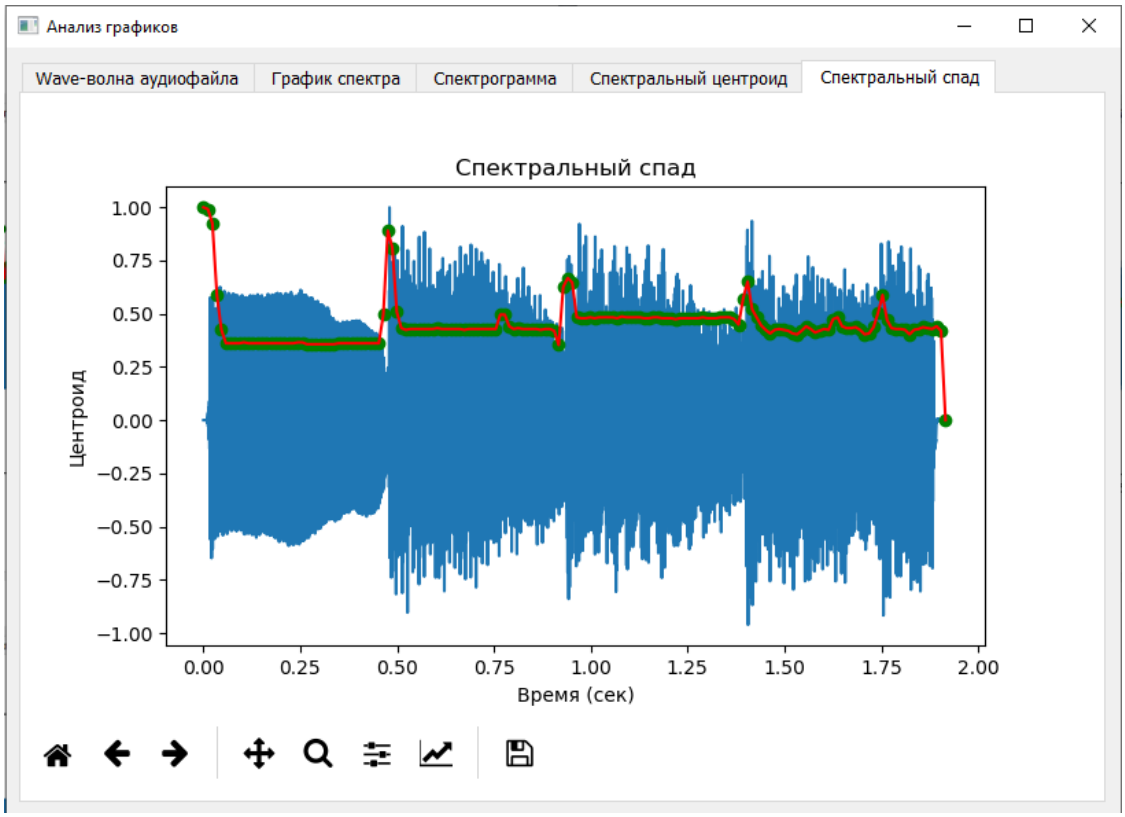

Рисунок 22 – График спектрального спада для аудиофайла с секретным сообщением

Для аудиофайла с секретным сообщением, также можно заметить, что график ведёт себя скачкообразно, имеются резкие переходы от больших значений к меньшим и наоборот, что может говорить о возможных изменениях в области данных аудиофайла. Чтобы правильно выявлять подозрительные изменения в значениях спектрального центроида и спада, необходимо понять насколько большой должна быть разницы между двумя соседними величинами. Также для каждого из признаков приведено сравнение аудиофайлов с внедренным секретным сообщением и без него.

Программный продукт был протестирован для 60 аудиофайлов формата mp3, 30 аудиофайлов с внедрённым секретным сообщением и 30 обычных (без секретного сообщения). Тестирование можно считать успешно пройденным. Точность определения наличия сокрытой информации составляет 90-95%. При этом можно заметить, что существует небольшая погрешность в результатах работы программы. Для 2 из 30 аудиофайлов с секретным сообщением программный продукт не смог сделать правильные выводы. Предположительно это связано с тем, что в данных аудиофайлах наблюдались небольшие изменения, которые были трудноразличимы разработанным методом.

ЗАКЛЮЧЕНИЕ

Информационная достаточность форматов хранения аудиоданных даёт большое количество мест для сокрытия информации в аудиофайлах. Несмотря на появление новых способов внедрения, основные методы, которые были рассмотренные в работе, являются наиболее используемыми на сегодняшний день. Анализ аудиофайла на наличие сокрытого сообщения – важный процесс, позволяющий оценить файл на вопрос внедрения какого-либо сообщения или человеком. При правильно выбранном способе детектирования, к примеру, можно определить есть ли в перехваченном файле какое-то секретное сообщение или нет.

В данной работе были рассмотрены основные теоретические понятия стеганографии, а также основные положения и принципы современной компьютерной стеганографии.

Также были рассмотрены различные стеганографические методы. Рассмотренные методы имеют свои достоинства и недостатки, но наиболее достойным и устойчивым к атакам является метод фазового кодирования, т.к. человеческое ухо воспринимает не сами значения фазы, а только их разность, что позволяет скрыть необходимую информацию в файле.

Также были определены и рассмотрены некоторые методы анализа аудиофайла на наличие сокрытого сообщения. Для рассмотренных методов были выявлены их преимущества и недостатки, которые могут тем или иным образом сказаться на полученном результате анализа.

В результате был разработан и протестирован программный продукт, реализующий предложенный способ анализа выбранного аудиофайла на наличие сокрытого сообщения. Стоит отметить, что точность полученных значений очень высока, что говорит о высокой точности определения предложенным способом детектирования сообщения в аудиофайле.