УДК 004.056.5 **https://doi.org/10.33082/td.2018.2-3.19**

УСОВЕРШЕНСТВОВАНИЕ МЕТОДОВ СТЕГОАНАЛИЗА ДЛЯ ОПРЕДЕЛЕНИЯ СТЕГОВЛОЖЕНИЙ В АУДИОФАЙЛАХ

С.Н. Григоренко
ст. преподаватель кафедры Компьютеризированных систем управления
Института компьютерных систем
Д.А. Лысь
студентка кафедры Компьютеризированных систем управления
Института компьютерных систем

Одесский национальный политехнический университет

Аннотация. *В работе рассматривается усовершенствованная методика и разработанное для ее реализации программное обеспечение, что позволяет обнаруживать вложения, сделанные с помощью алгоритма LSB, в некоторые виды аудиофайлов формата WAVE любой битности и частоты дискретизации, а также определять наличие стеговложений в определенных видах аудиофайлов. Методика тестирования основана на методах определения стеговложений с помощью частотного метода и метода сжатия.*

Ключевые слова: *аудиостегоанализ, WAVE-файлы, LSB-методы, сжатие данных, оцифровка, стенография*.

ВДОСКОНАЛЕННЯ МЕТОДІВ СТЕГОАНАЛІЗУ ДЛЯ ВИЗНАЧЕННЯ СТЕГОВКЛАДЕНЬ В АУДІОФАЙЛАХ

С.М. Григоренко
ст. викладач кафедри Комп'ютеризованих систем управління
Інституту комп'ютерних систем
Д.А. Лись
студентка кафедри Комп'ютеризованих систем управління
Інституту комп'ютерних систем

Одеський національний політехнічний університет

Анотація. *В роботі розглядається вдосконалена методика і розроблене для її реалізації програмне забезпечення, що дозволяє виявляти вкладення, зроблені за допомогою алгоритму LSB, в деякі види аудіофайлів формату WAVE будь-якої битности і частоти дискретизації, а також визначатиме наявність стеговкладень в певних видах аудіофайлів. Методика тестування заснована на методах визначення стеговкладень за допомогою частотного методу і методу стиснення.*

Ключові слова: *аудіостегоаналіз, WAVE-файли, LSB-методи, стискування даних, оцифровка, стенографія.*

UDC 004.056.5

IMPROVEMENT OF STEGOANALYSIS METHODS FOR DETERMINING TRAINING IN AUDIO FILES

Grigorenko S.M.
Senior lecturer of the Department of Computerized control systems
of the Institute of computer systems
Lys D.A.
Student of the Department of Computerized control systems
of the Institute of computer systems

Odessa national polytechnic university

Abstract. *The paper considers the improved methodology and the software developed for its implementation, which makes it possible to detect attachments, made with the help of the LSB algorithm, in some types of audio files of the WAVE format of any bit rate and sampling frequency. It also provides us with the possibility to determine the presence of stego attachments in certain types of audio files. The testing procedure is based on methods of determining stego attachments using the frequency method and compression method. The suggested methods of searching for attachments in file service areas and developed software in Embarcadero Delphi XE5 environment effectively reveal the fact of the implementation of the message created by the majority of existing programs. The software was also tested on a large number of files without attachments. As a result a large number of false positives were revealed. The reason for this was damaged structure of MP3 files (corrupted frames were taken for possible stego attachments).*

Keywords: *audio stegoanalysis, WAVE-files, LSB-methods, compression of data, digitization, steganography.*

Введение. В настоящее время в связи с бурным развитием компьютерных технологий появилось новое направление стеганографии – компьютерная (или цифровая) стеганография, которая направлена на встраивание сообщений в различные типы файлов (текстовых, графических, аудио, видео и др.). В связи с возрастанием роли глобальных компьютерных сетей цифровая стеганография приобретает большую значимость.

Анализ основных достижений и литературы. Анализ источников сети Internet позволяет сделать вывод, что в настоящий момент цифровая стеганография используется для следующего:

1. Скрытая передача сообщений, используемая для различных целей;
2. Защита конфиденциальной информации от несанкционированного доступа;
3. Преодоление систем мониторинга и управления сетевыми ресурсами;

4. Камуфлирование ПО;
5. Защита авторского права на определенные виды интеллектуальной собственности.

В настоящее время разрабатываются новые методы компьютерной стеганографии, основанные на особенностях представления информации в цифровом виде. Часть этих методов использует модификацию палитры, неточность устройств оцифровки, избыточность аудио и видео файлов и др. подходы. Несмотря на бурное развитие стеганографических методов, в свободном доступе имеется недостаточно ПО для стеганографии в аудиофайлах. Проблема связана с тем, что методы вложения информации в аудиофайлы разных битностей несколько различны. В настоящее время не существует универсальных программных решений для работы с аудиофайлами разных битностей.

Цель статьи, постановка заданий. Целью работы является модификация методов стеганографии и стегоанализа на аудиофайлах.

Основные задачи: рассмотреть особенности сокрытия и обнаружения скрытых вложений в аудиофайлах, создать приложение для сокрытия данных в WAVE файлах, предложить и модифицировать методы обнаружения скрытых данных в аудиофайлах.

Основная часть. Для стеганографии на аудиофайлах и для рассмотренных методов стегоанализа разработано оригинальное программное обеспечение, реализованное в виде пакета Stegora WaveHide.

Формат WAVE был выбран из тех соображений, что он идеально подходит для реализации алгоритма LSB в силу своей избыточности. В области данных аудиофайлов формата WAVE хранятся несжатые и никоим образом не измененные данные, полученные напрямую с аналого-воцифрового преобразователя, поэтому реализовывать стеганографические алгоритмы на файлах данного типа несколько проще и понятнее.

Поскольку WAVE файлы имеют достаточно большой размер, они не используются для обмена в сети интернет и для хранения музыки на портативных устройствах (плееры, мобильные телефоны). WAVE файлы используются там, где необходимо сохранить первозданный вид файла высокого качества там, где нет ограничения на размер свободного дискового пространства.

Аудиофайлы формата WAVE характеризуются битностью и частотой дискретизации.

Метод LSB (Least Significant Bit, Младший Значащий Бит) является методом, использующим избыточность звуковых файлов. Как известно, младшие разряды цифровых отсчетов содержат очень мало полезной информации. Их заполнение дополнительной информацией практически не влияет на качество восприятия, что и обеспечивает возможность скрытия. У данной группы методов имеется ряд отличительных особенностей. Сначала рассмотрим негативные особенности. С изменением информации искажаются статистические характеристики цифровых потоков. Ввиду этого для снижения компрометирующих признаков требуется

коррекция статистических характеристик. К достоинствам можно отнести: возможность скрытой передачи большого объема информации, возможность защиты авторского права, скрытого изображения, товарной марки, регистрационных номеров и т.п.

В силу своей простоты и прозрачности реализации метод LSB в настоящее время широко применяется в стеганографии.

Для стеганографии на аудиофайлах и для рассмотренных методов стегоанализа разработано оригинальное программное обеспечение, реализованное в виде пакета Stegora WaveHide. Приложение выполняет стеговложения в аудиофайлы формата WAVE любой частоты дискретизации и битности со свойством многотомности, а также в приложении реализована собственная методика обнаружения таких вложений и метод обнаружения сообщений в аудиофайлах на основе методов сжатия. ПО разработано в среде Embarcadero Delphi XE5.

В программе в качестве стеганографического алгоритма был реализован алгоритм LSB. Как уже говорилось ранее суть данного алгоритма состоит в замене наименьших значащих битов аудиофайла битами сообщения. На вход программе подается файл-контейнер, файл-сообщение, а также ключевой файл. Файл-сообщение шифруется с помощью алгоритма EQXOR и ключевого файла.

При тестировании базы файлов на предмет вложений с помощью частотного анализа для отслеживания поведения случайности младших бит необходимо упорядочить файлы по какому-либо признаку. В качестве такого признака было выбрано относительное количество нулевых байт в файле. Относительное количество нулевых байт определяется как отношение нулевых байт файла к общему количеству байт. В соответствии с этим признаком были упорядочены все 108 тестовых файлов. Для проверки гипотезы в исходную базу (кроме пустых контейнеров) были добавлены частично заполненные, а именно, во все файлы из базы были осуществлены стеговложения (при помощи разработанного ПО Stegora WaveHide) на 10 %, 50 % и 100 % от максимальной возможности рассматриваемого стегоконтейнера. В качестве теста был использован частотный побитовый тест пакета NIST, который оценивает соотношение между нулями и единицами в двоичной последовательности. Алгоритм этого теста заключается в следующем: файл рассматривается как битовая последовательность, при этом единица принимается за +1, ноль за 0. Считается сумма последовательности. Затем вычисляется статистика по формуле

$$S_{obs} = \frac{|S|}{\sqrt{n}}, \qquad (1)$$

где |S| – сумма последовательности;

n – количество элементов в последовательности. Вычисляется P-значение через дополнительную функцию ошибок

$$P_{value} = erfc(\frac{S_{obs}}{\sqrt{2}}) . \quad (2)$$

Дополнительная функция ошибок вычисляется так:

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt . \quad (3)$$

И если результат больше, чем 0,01, то последовательность признается случайной с уровнем доверия 99 %.

С помощью рассмотренного выше, всего 108 файлов наименьших значащих бит из базы были проверены на случайность. Результаты представлены на рис. 1, 2.

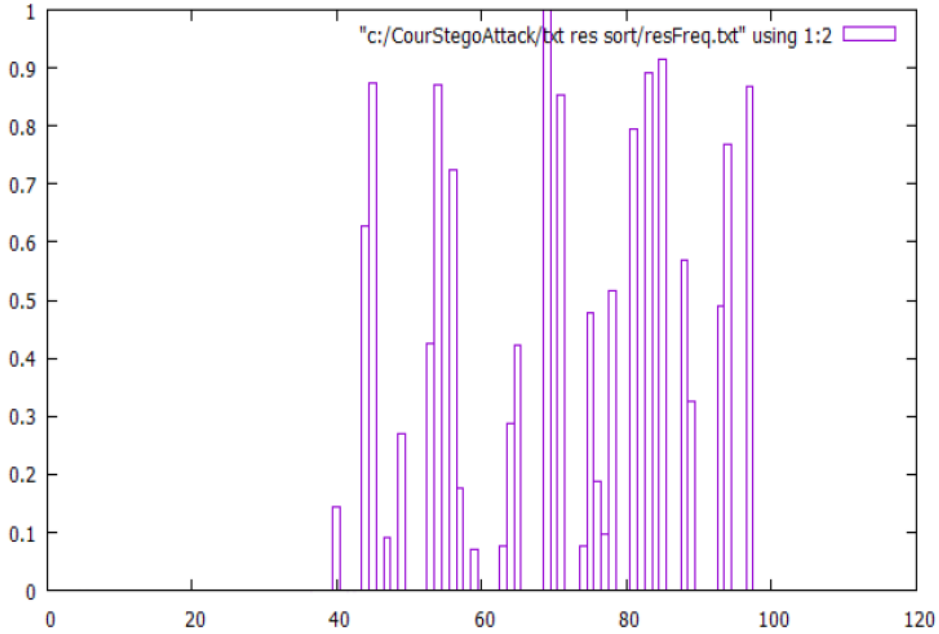

Рис. 1. Проверка на случайность последовательности LSB 108 «пустых» аудиофайлов

Как видно из рисунка, последовательности LSB у файлов с относительным количеством нулевых байт менее 0,08 (значение для 40 файла) неслучайны.

Установлено, что предложенная методика работает весьма эффективно (определяются даже 10 % вложения (рис. 1), но только на аудиофайлах с малым относительным количеством нулевых байт.

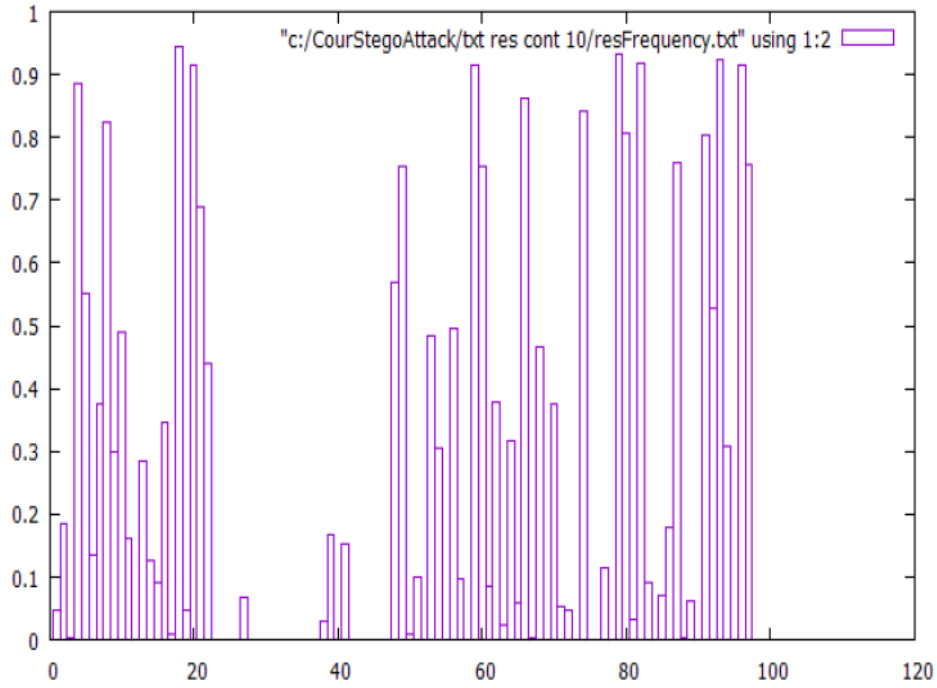

Рис. 2. Проверка на случайность последовательности LSB 108 аудиофайлов, заполненных на 10 % от максимальной возможности

Практическое использование методики для определения вложений в аудиофайл:

1. Берется проверяемый файл.
2. Определяется относительное число нулевых байт.
3. Если это число меньше, чем 0,0038 (подобрано эмпирически), то файл проверяется частотным тестом NIST.

Если тест показал, что последовательность младших битов файла является случайной (т.е. результат частотного теста больше, чем 0,01), то это с большой долей уверенности говорит о том, что в файле имеются стеговложения.

При тестировании базы файлов на предмет вложений с помощью метода, основанном на алгоритме сжатия, принято:

- виды исследуемых файлов максимально произвольные (представленные в базе);
- количество младших бит, заменяемых на случайные – 2 (как это было сделано в стеганографической части разработанного приложения);
- битность файлов от 8 до 32.

В разработанном программном обеспечении данный метод реализован следующим образом: В интерфейсе разработанного пакета Stegora WaveHide при нажатии на кнопку «Alternate Analysis» и вводе коэффициентов sigma и K пользователю требуется выбрать файл для анализа. Вначале файл разбивается на K равных частей. Затем производится сжатие каждой из частей при помощи методов из стандартной библиотеки zlib и высчитывается отношение размера сжатого куска файла к несжатому. После в изначальном файле два младших бита изменяются на случайные и процедура повторяется, высчитывается отношение размеров сжатого куска к несжатому. Для каждого куска рассчитывается параметр Δ.

Если количество кусков, параметр Δ для которых меньше, чем введенный коэффициент sigma, больше половины, то выносится решение о наличии в рассматриваемом файле стеговложений. В противном случае выносится решение об отсутствии таких вложений.

Тестирование базы файлов на предмет вложений с помощью метода, основанного на алгоритме сжатия Тестовая база из 108 аудиофайлов была проверена на стеговложения с помощью данного метода. На графиках ниже показаны значения параметра Δ. Сравнивались значения этого параметра для «пустых» и заполненных (с помощью разработанного ПО Stegora WaveHide) на 100 % файлов. Результаты представлены на графиках ниже. По оси абсцисс на графиках находятся номера тестируемых файлов, по оси ординат – полученное значение параметра Δ.

Как видно из результатов тестирования разработанного ПО, заполненный контейнер сжимается в среднем в два раза хуже. Следовательно, метод стегоанализа, основанный на алгоритмах сжатия можно использовать для определения наличия вложений в аудиофайлах. Но, поскольку параметр Δ различен для разных видов файлов, для эффективного применения данного метода необходимо классифицировать файлы и для каждого класса установить свое пороговое значение Δ.

Результаты исследования. Результаты тестирования базы из 108 различных аудиофайлов формата WAVE показали, что метод, основанный на алгоритмах сжатия требует дополнительной доработки в плане классификации аудиофайлов. С большой долей уверенности можно говорить о возможности применения данного метода для определения стеговложений в аудиофайлах, принадлежащих к одному классу.

Результаты тестирования этой же базы на основании двух методик показали следующее, что метод работает весьма эффективно для файлов, относительное количество нулевых байт для которых меньше, чем 0,038. Файлы, относящиеся к данному классу, отличаются большой информационной нагруженностью. К этому классу относятся шумы, записи музыкальных инструментов с большим количеством шумов (большое количество шума возникает из-за использования некачественных АЦП).

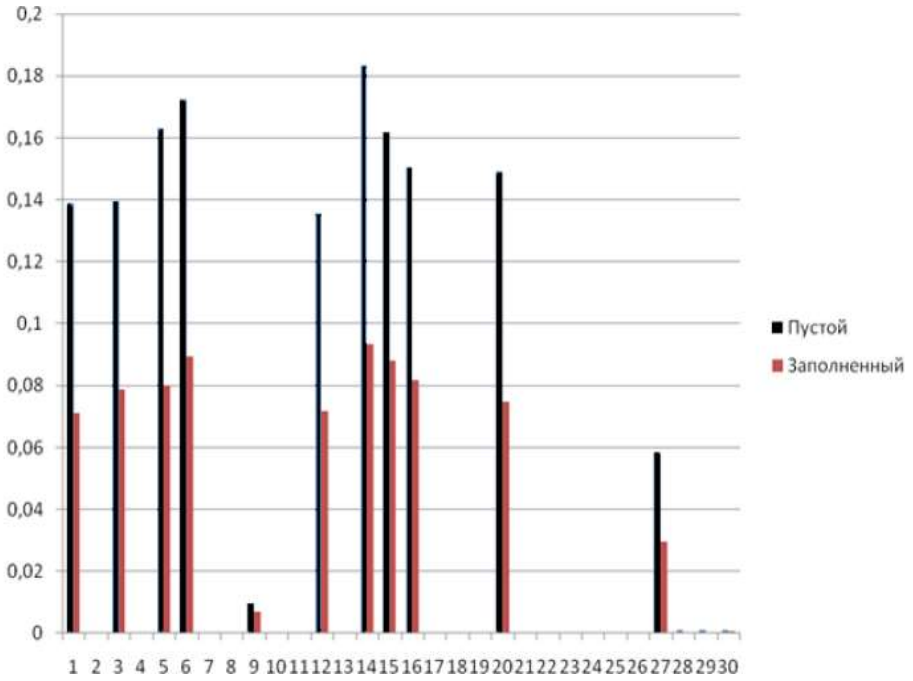

Рис. 3. Тестирование файлов с 1 по 30

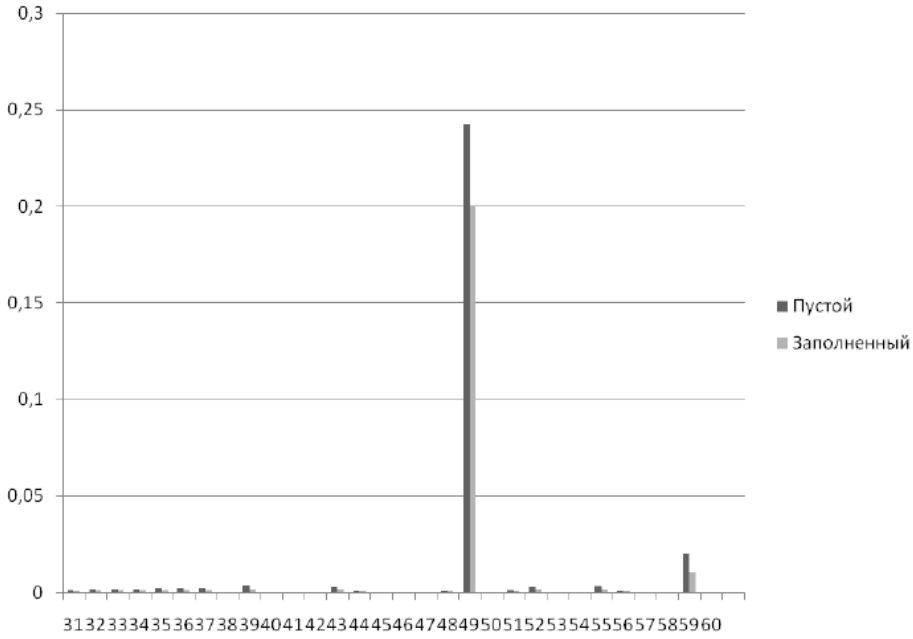

Рис. 4. Тестирование файлов с 31 по 60

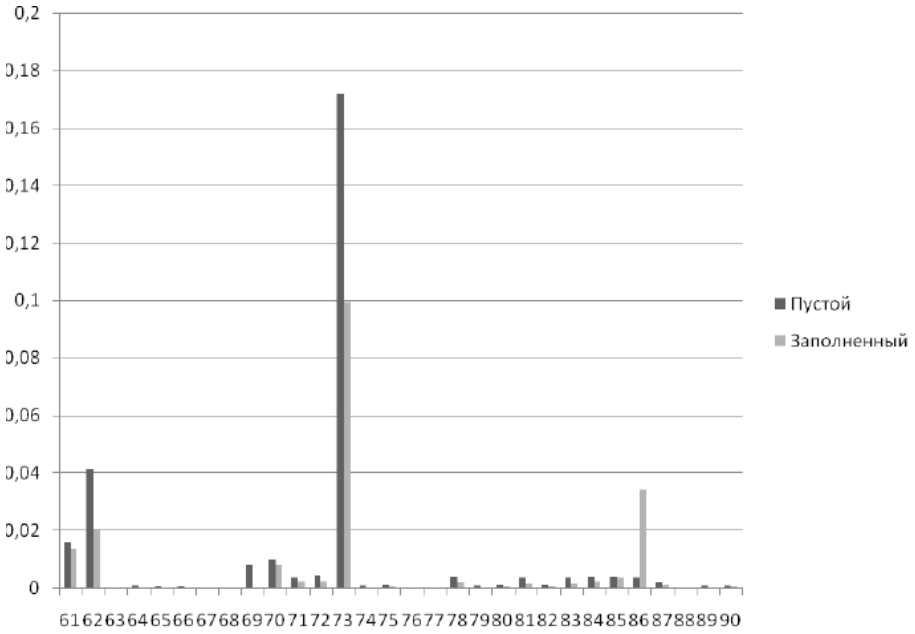

Рис. 5. Тестирование файлов с 61 по 90

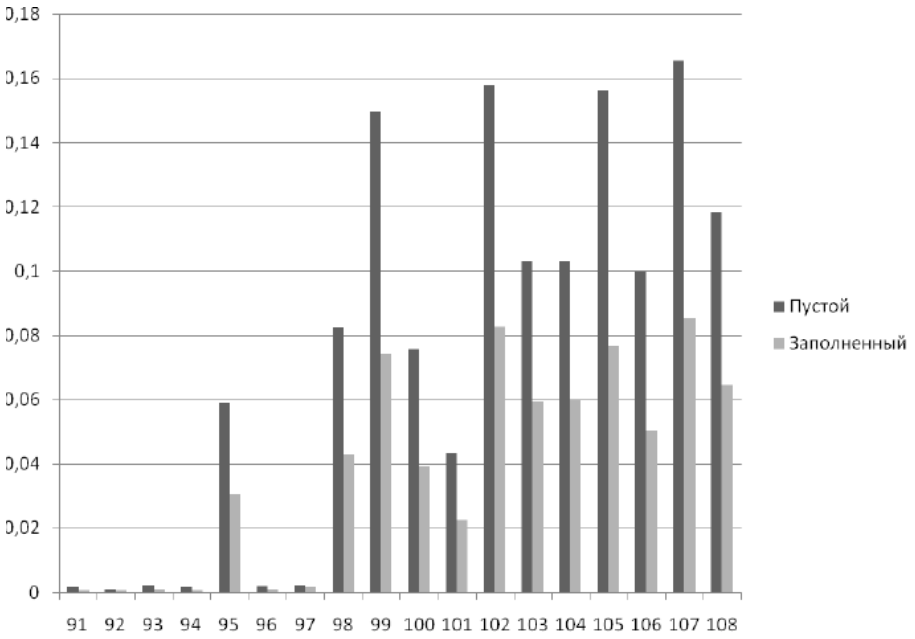

Рис. 6. Тестирование файлов с 91 по 108

Выводы. Информационная избыточность форматов хранения аудио данных предоставляет большое количество мест для сокрытия информации как в служебных областях файлов, так и непосредственно в области аудиоинформации. Появление большого количества различных методов и алгоритмов стеганографии в аудиофайлах порождает трудности для стегоанализа, т.к. требуется учитывать особенности работы каждого из этих методов. Предложенные методы поиска вложений в служебных областях файлов и разработанное ПО эффективно обнаруживает факт внедрения сообщения, созданного большинством существующих программ. В то же время его недостатком является ложное срабатывание на поврежденные файлы. Для предотвращения этого необходимо анализировать сами подозрительные данные, что является, несомненно, не менее трудной задачей, выходящей за рамки данной работы.

Разработанная методика позволяет обнаруживать вложения, сделанные с помощью алгоритма LSB, в аудиофайлы формата WAVE. В будущем планируется усовершенствовать данную методику и расширить ее на все виды аудиофайлов.

СПИСОК ЛІТЕРАТУРИ

1. *Барсуков В.С., Романцов А.П. Несколько слов о стеганографии // Специальная техника. – 1998. – № 4. – С. 25-26.*
2. *Быков С.Ф., Мотуз О.В. Основы стегоанализа // Защита информации. – 2000. – № 3.*
3. *Генне О.В. Основы стегоанализа // Защита информации. – 2000. – № 3. – С. 57-58.*
4. *Гурский Д.И. ActionScript 2: программирование во Flash MX / Д.И. Гурский. – 2004. – 860 с.*

Стаття надійшла до редакції 20.02.2018 р.