УДК 004.021

НЕКОТОРЫЕ АСПЕКТЫ ЗАЩИТЫ ИНФОРМАЦИИ СИСТЕМЫ ОБРАБОТКИ МУЛЬТИМЕДИЙНЫХ ДАННЫХ

Апсалямова Р.Д., Долматова Я.Г., Душкин А.В., Панычев С.Н., Сахаров С.Л.
ФКОУ ВО «Воронежский институт» ФСИН России, Воронеж, e-mail: a_dushkin@mail.ru

В работе рассмотрена методика анализа аудиофайлов формата WAV с целью поиска стеганографических вложений, внедренных в аудиоданные методом наименьших значащих бит. Для реализации стеганографического алгоритма выбрана программа FoxSecret v1.0, имеющая русскоязычный интуитивно понятный интерфейс и свободно распространяемая в сети Интернет. Приведена краткая описательная модель действий злоумышленника. Подробно рассмотрена методика анализа на основе оценки математического ожидания распределения младших бит с последующей оценкой эксцесса распределения полученных значений. Приведены примеры анализа группы файлов, показана последовательность действий при выборе численных значений порога обнаружения. Приведена методика углубленного анализа одиночного файла по паузам в прямой речи, рекомендованная для файлов с большим разбросом характеристик пустых контейнеров. В условиях отсутствия файла оригинала, в качестве опорных значений используются параметры первых бит, имеющих в пустом контейнере корреляцию с младшими (нулевыми) разрядами.

Ключевые слова: стеганография, защита информации, наименьшие значащие биты, статистический анализ, порог обнаружения, FoxSecret v1.0, Hex-редактор

SOME ASPECTS OF THE PROTECTION OF MULTIMEDIA INFORMATION PROCESSING SYSTEM

Apsalyamova R.D., Dolmatova Ya.G., Dushkin A.V., Panychev S.N., Sakharov S.L.
Federal State Educational Institution of Higher Education Voronezh Institute of the Russian Federal Penitentiary Service, Voronezh, e-mail: a_dushkin@mail.ru

The paper considers the method of analysis of WAV format audio files to search for steganography attachments embedded in the audio data by the least significant bits. To implement the steganographic algorithm selected FoxSecret v1.0 program with a Russian intuitive interface and freely available on the Internet. The short descriptive model of actions of the malefactor is resulted. The technique of the analysis on the basis of an estimation of a population mean of distribution of younger bats with the subsequent estimation of an excess of distribution of the received values is in detail considered. Examples of the analysis of group of files are resulted, the sequence of actions is shown at a choice of numerical values of a threshold of detection. The technique of the profound analysis of a single file on pauses in the direct speech, recommended for files with a wide spacing of characteristics of empty containers is resulted. In the conditions of absence of a file of the original, as basic values parameters of the first bats having in the empty container correlation with younger (zero) categories are used.

Keywords: steganography, data protection, the least significant bits, statistical analysis, detection threshold, FoxSecret v1.0, Hex-Editor

Информатизация общества, основанная на современных инфокоммуникационных технологиях, способствует повсеместному внедрению электронного делопроизводства, расширению инфраструктуры обеспечения функционирования и развития системы передачи и обработки данных, предоставляет пользователям информационные ресурсы. Возможность доступа к сетям связи общего пользования выводит на передний план вопросы обеспечения безопасности данных в государственных и частных организациях и учреждениях.

В настоящее время все большую опасность приобретают способы скрытой передачи информации под видом стандартных файлов, при помощи методов стеганографии [5, 6]. При организации скрытого канала в цифровых сетях связи и передаче сравнительно небольших объемов данных (от нескольких десятков до нескольких сотен килобайт) для модели действий нарушителя могут быть приняты следующие допущения:

– отправка скрытых сообщений осуществляется в переписке (электронной почтой), в виде приложений из небольших групп по 5–10 файлов или отдельных файлов;

– используется способ повышающий скрытность канала, при котором ограничен размер стеговложения, но не изменяется общий размер файла – внедрение в наименьшие значащие биты (НЗБ);

– размер контейнера не должен вызывать подозрений у должностных лиц, отвечающих за информационную безопасность и контролирующих служебный документооборот (уменьшение размера контейнера возможно за счет снижения частоты дискретизации и «глубины» звучания).

Для анализа в работе выбраны файлы звукового формата, т.к. применение стеганографических алгоритмов и атак на стегосистемы в печатных и интернет-изданиях рассматривается, как правило, на примере графических файлов и злоумышленник может ожидать, что именно эти файлы будут подвергаться первоочередному анализу при поиске стего. Кроме того, по сравнению с видео, аудио-файлы используются чаще и имеют простую структуру.

Предварительный анализ показал, что распределение НЗБ зависит от вида записанной информации:

– музыкальные произведения представляют собой непрерывный звуковой поток, в котором паузы (зоны молчания) являются редким исключением;

– речевые сообщения (запись совещаний, выступлений, аудиописьма, аудио-уроки и инструкции и т.д.) включают многочисленные паузы разной длительности (между словами, предложениями и т.д.);

– некий промежуточный вариант: речевые сообщения на фоне естественных негромких шумов (шум улицы, работающие в помещении радио или телевизор).

Запись скрытого сообщения может иметь разную плотность (количество измененных младших бит на интервал в 100 или 1000 байт). Запись может производиться в один из нескольких каналов или последовательно в байты независимо от количества каналов. Кроме того, скрытые вложения могут быть распределены по всему файлу равномерно, что снижает плотность и затрудняет обнаружение.

Многообразие контейнеров и способов (алгоритмов и программ) размещения стеговложений дает основание предполагать, что для достоверного обнаружения скрытых каналов требуется отдельно рассматривать каждую комбинацию «пустой контейнер – способ записи».

Среди программ реализующих стеганографические алгоритмы методом НЗБ, обращает на себя внимание FoxSecret, свободно распространяемая в глобальной сети. В сравнении с аналогичными программами у нее есть еще ряд преимуществ:

– русскоязычный интерфейс, более простой и понятный, чем в OpenPuff или StegoMagic;

– плотность записи выше, чем в OpenPuff;

– в отличие от Stools возможно использование файлов с глубиной звучания 8 бит на сэмпл (что необходимо для снижения размера контейнера).

Так как в учебной и научной литературе, например в [3, 4], посвященной стеганографии и атакам на скрытые вложения, методика только упоминается с приведением результатов без описания алгоритмов работы с аудиофайлами, то для поиска возможных решений был проведён ряд статистических тестов. Целями исследования являлось обнаружение скрытого стеганографического вложения и определение численных критериев принятия решения. Кроме того, показанную ранее возможность поиска вложений в заголовках файлов [1] следует дополнить алгоритмами поиска стего в аудиоданных.

С учетом приведенных выше предположений и допущений, для проведения анализа предварительно были подготовлены две группы пустых контейнеров, отличающихся частотой дискретизации – 44100 и 22050 Гц. Остальные характеристики: формат WAV, один канал (монофонический сигнал), глубина (точность) звучания – 8 бит на сэмпл.

В качестве скрытого вложения с помощью программы FoxSecret в наименьшие значащие биты файла помещался текстовый документ размером 100Кбайт. При этом тип вложения несущественен, так как программой проводится предварительное сжатие и шифрование (в данном случае по алгоритму Blowfish). В качестве дополнения к ключу пользователя (даже если ключ «пустой») программой автоматически добавляются данные о создании/изменении файла. Поэтому даже одно вложение (один и тот же файл) размещенное в одинаковых контейнерах (копии файла, отличающиеся только временем создания) имеет характер псевдослучайных не повторяющих друг друга битовых последовательностей без взаимной корреляции, что определено требованиями к системам шифрования [7]. Анализ бинарного кода пустых и заполненных контейнеров с использованием HxD (hex-редактор, версия 1.7.7.0) показал, что в заполненных контейнерах изменению подвергались младшие биты в байтах № 40–346740. Следует отметить, что Fox Secret не распределяет вложение равномерно по всему файлу и осуществляет запись с начала файла с большой плотностью – до 50 % измененных байт.

Статистические характеристики распределения значений нулевых и первых разрядов байтов файлов получены с помощью системы компьютерной алгебры MathCAD 15, позволяющей реализовать вычисление с использованием встроенной системы программирования и наглядно представить результаты на графиках.

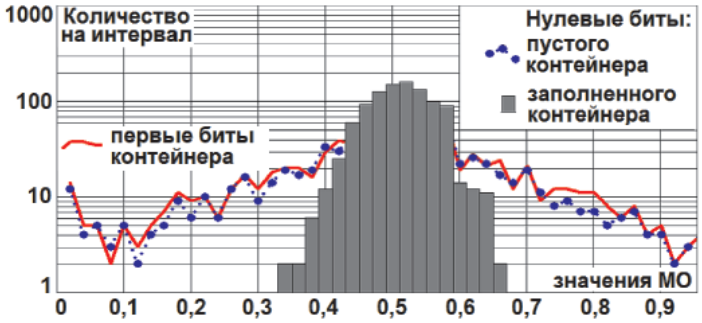

Рис. 1. Распределение МО в блоках по 100 бит для участка файла от 100 до 200 Кбайт

Оценка математического ожидания (МО) проводилась (в соответствии с методикой, изложенной в [2]) для блоков по 100 бит, последовательно считываемых из разрядов байтов файла. Распределение нулевых и первых разрядов байтов пустого контейнера (файла-оригинала) практически не отличаются и имеют разброс значений МО по всему диапазону, со средним значение около 0,5 и огибающей близкой к гауссовской (рис. 1). Для заполненного контейнера на участке размещения стеговложения характерно примерно двукратное сужение разброса крайних значений.

Полученные результаты могут быть интерпретированы как вероятность попадания на заданный интервал (участок гистограммы) с распределением близким к гауссовскому. Так же представляется целесообразным проверить возможность использования первых бит для нормировки результатов. Это предположение основано на том, что первые и нулевые биты пустого контейнера имеют коррелированные параметры распределения, что отмечено в [3]. Кроме того, первые биты не подвергаются изменению при стеганографическом вложении методом НЗБ и могут служить индикатором изменения распределений нулевых разрядов. При проведении дальнейших расчетов, МО нулевых бит пустого и заполненного контейнера так же разбивались на блоки по 100 значений и для каждого блока оценивались параметры распределения. Из рис. 1 очевидно, что в качестве оцениваемого параметра распределения могут быть использованы: разница между максимальным и минимальным значением, дисперсия и эксцесс. Результаты, полученные для всех трех параметров, в целом аналогичны. Однако далее приведены результаты только для эксцесса, т.к. оценка этого параметра дает наибольший разброс значений, что удобно для анализа и определения порога обнаружения.

Результаты, нормированные к параметрам распределения первых бит (на тех же участках файла, что и младшие биты), показаны на рис. 2 как среднее значение для 10 файлов. Увеличение количества исследуемых файлов приводит только к сглаживанию кривых, практически без изменения оцениваемых значений. Основные значения для пустого контейнера расположены в достаточно узком диапазоне 0,8–1,0 с малым разбросом, без существенных выбросов или отклонений. Для участков файла со стеговложением полученные значения лежат в диапазоне < 0,01 при частоте дискретизации 44100 Гц и при частоте дискретизации 22050 Гц имеют больший разброс, но не превышают уровня 0,1. Ход кривых может отличаться для различных групп файлов. Однако очевидно, что кратные отличия позволяют сформировать порог обнаружения в достаточно широких пределах 0,1–0,8. Решение о наличии стеговложения при отсутствии файлов оригиналов (пустых контейнеров) может приниматься по количеству пересечений порога на участках файла. В общем случае порог обнаружения должен определяться из анализа групп файлов (пустых контейнеров), наиболее часто используемых в сети организации (учреждения).

Однако в практической работе наиболее вероятна проверка не группы, а единичных файлов. На рис. 3 приведены характеристики файла, в котором наличие и интервал расположения скрытого вложения не могут быть уверенно определены, независимо от частоты дискретизации. Но при этом характеристики пустого контейнера (этот же файл без стего) могут привести к ложному решению о наличии вложения, распределенного по всему файлу. Если приоритетным является пресечение утечки информации, то ошибочное решение будет являться «платой» за возможность обнаружения скрытого канала передачи данных, а также вынуждает перейти от периодического выборочного контроля к постоянному наблюдению за выявленным каналом.

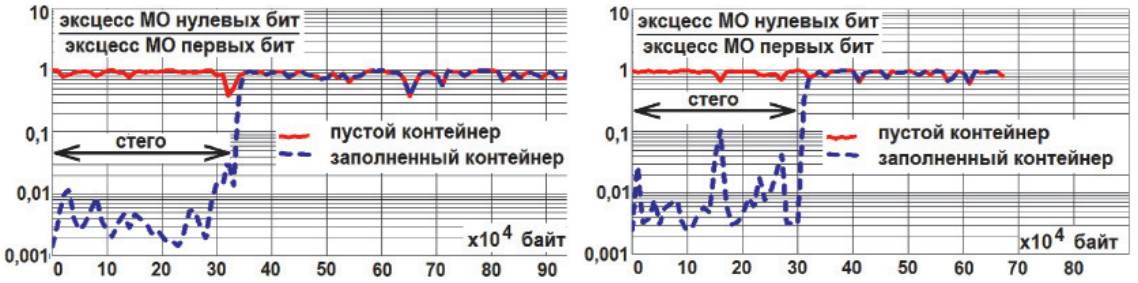

Рис. 2. Нормированный эксцесс распределения МО контейнеров с разными частотами дискретизации (среднее значение для групп из 10 файлов): а – 44100 Гц; б – 22050 Гц

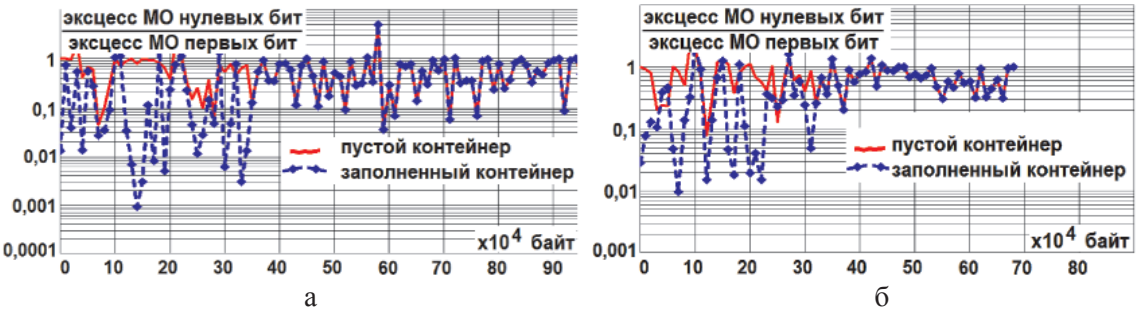

Рис. 3. Эксцесс распределения МО по одиночному файлу с разными частотами дискретизации (пример со значительным разбросом параметров): а – 44100 Гц; б – 22050 Гц

Следует учитывать, что качество записи зависит от оборудования и условия записи могут сильно изменяться. Сравнение кривых для пустого и заполненного контейнера на рис. 3 показывает достаточно большой разброс значений параметров распределения младших бит (с перекрытием диапазонов значений). В отсутствии пустого контейнера, для сравнения, анализ такого файла может привести не только к ложному обнаружению, но и к пропуску скрытого сообщения и утечке информации – одной из наиболее опасных угроз для автоматизированных системы обработки данных.

Неопределенность в принятии решений характерна для файлов речевых сообщений с негромкими фоновыми шумами, хорошо различимыми как при воспроизведении прямой речи, так и в паузах между словами и предложениями. Причина возникновения шумов: низкое качество (техническое состояние) записывающего оборудования или посторонние звуки (непрофессиональная запись или запись в плохих условиях). Для файла, характеристики которого показаны на рис. 3, при воспроизведении слышны постоянное негромкое потрескивание (шумы записывающего оборудования), периодический шорох одежды и негромкий шум перекладываемых на столе предметов, а в паузах отчетливо слышно дыхание диктора. В то же время на основе детального анализа сделано предположение, что при описываемой методике основной вклад в обнаружение стего в речевых сообщениях вносится именно за счет характеристик распределения младших бит в паузах между словами и предложениями.

Для проверки предположения заполненный контейнер с частотой дискретизации 22050 Гц и разделен на две части: прямая речь без пауз и только паузы между словами и предложениями. Анализ прямой речи не позволил выявить скрытое вложение, а в паузах (с сохраненными фоновыми шумами, продолжительность – около 8 с из 34 с общей длительности записи) стего обнаруживается по резкому изменению значений статистических характеристик (рис. 4). При указанном выше пороге принятия решения пределах 0,1–0,8 и предположении, что паузы в прямой речи расположены равномерно, можно сделать вывод, что вложение расположено в начале файла и занимает около 45% от общей длительности. Полученный результат является достаточно точным: анализ с использованием HxD (hex-редактора) показал, что из 771299 байт изменению подвергались нулевые разряды в байтах с № 46 по № 346741.

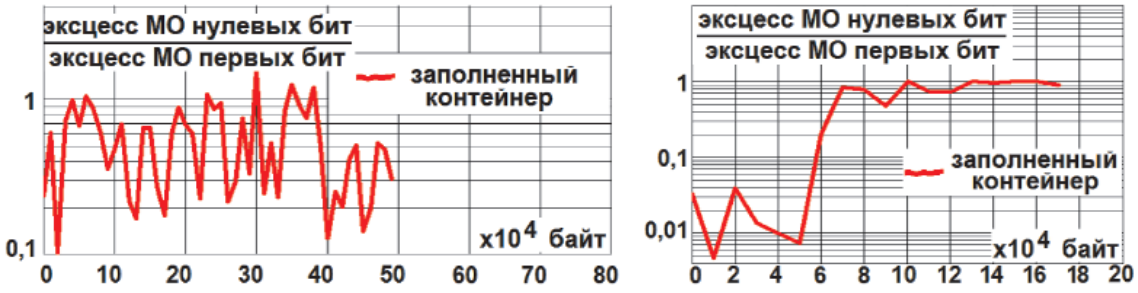

Рис. 4. Эксцесс распределения МО по одиночному файлу с частотой дискретизации 22050 Гц, при разделении прямой речи и пауз:
а – прямая речь; б – паузы

Таким образом, при выявлении скрытого канала передачи информации, реализованного стеганографическими алгоритмами методом НЗБ целесообразно определение МО нулевых бит файла на коротких интервалах с последующим анализом эксцесса распределения в группах полученных значений с нормировкой к эксцессу распределения первых бит. Порог обнаружения может устанавливаться исходя из предварительного анализа групп файлов по минимальному уровню значений для характеристик пустых контейнеров. При анализе отдельных файлов, в случае большого разброса полученных числовых характеристик (неоднозначности в принятии решения) возможно проведение дополнительного исследования по фрагментам файла – паузам в прямой речи.

Полученные результаты могут быть реализованы программно и применимы для анализа одного канала аудиозаписей речевых сообщений (совещаний) в формате WAV с глубиной (точностью) звучания – 8 бит на сэмпл.

Список литературы

1. Апсалямова Р.Д., Душкин А.В., Кравченко А.С., Панычев С.Н., Сахаров С.Л. Обеспечение информационной безопасности систем обработки данных путем поиска стеганографических вложений в метаданных аудиофайлов // Современные наукоемкие технологии. – 2016. – № 8–1. – С. 13–17.

2. Вентцель Е.С. Теория вероятностей: учеб. для вузов. – 6-е изд. стер. – М.: Высш. шк., 1999. – 576 с.

3. Грибунин В.Г. Цифровая стеганография / В.Г. Грибунин, И.Н. Оков, И.В. Туринцев – М.: СОЛОН-ПРЕСС, 2009. – 272 с.

4. Конахович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография. Теория и практика. – Киев: МК-Пресс, 2006. – 288 с, ил. ISBN 966-8806-06-9.

5. Николаев А. Проблемы выявления скрытой передачи информации по сетям // Information Security / Информационная безопасность. – 2009. – № 1. – С. 24–26.

6. Сердюк В. Современные технологии защиты от утечки конфиденциальной информации // Век качества. – 2005. – № 3. – С. 62–67.

7. Скляров Д.В. Искусство зашиты и взлома информации. — СПб.: БХВ-Петербург, 2004. – 288 с: ил. ISBN 5-94157-331-6.