УДК 004.021

ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ СИСТЕМ ОБРАБОТКИ ДАННЫХ ПУТЕМ ПОИСКА СТЕГАНОГРАФИЧЕСКИХ ВЛОЖЕНИЙ В МЕТАДАННЫХ АУДИОФАЙЛОВ

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

В работе рассмотрен алгоритм обнаружения стеганографических вложений в целях пресечения скрытых каналов передачи (утечки) информации в цифровых системах обработки данных. Проведен подробный анализ алгоритма создания канала с большой пропускной способностью – формирование стего в кадрах ID3-тегов файлов формата «.mp3». Для реализации стеганографического алгоритма, выбрана программа FoxSecret v1.0, имеющая русскоязычный интуитивно понятный интерфейс и свободно распространяемая в сети Интернет. На основе теста хи-квадрат проведено сравнение статистики распределения младших бит файлов со стеганографическими вложениями и вложениями, выполненными без использования алгоритмов сокрытия информации. Приведены результаты анализа кадров тега, разбитых на интервалы различной длины. Приведен пример анализа одиночного файла, имеющего одновременно как нескрываемые («легальные»), так и стеганографические вложения. Определены численные значения разброса статистик, лежащие в основе принятия решения о наличии скрытого канала передачи данных.

Ключевые слова: стеганография, защита информации, теги ID3v2.3, статистический анализ, формат MP3, FoxSecret v1.0, HxD-Hex-редактор

MAINTENANCE INFORMATION SECURITY OF DATA HANDLING SYSTEMS BY SEARCH STEGANOGRAFICHESKY INVESTMENTS IN META DATA OF AUDIOFILES

Apsaljamova R.D., Dushkin A.V., Kravchenko A.S., Panychev S.N., Saharov S.L.
Federal state educational institution of higher education Voronezh institute of the Russian Federal Penitentiary Service, Voronezh, e-mail: kr_and@inbox.ru

In this paper we consider steganography detection algorithm investments in order to prevent hidden channels (leakage) of digital information processing systems. A detailed analysis of the algorithm, create a channel with large bandwidth – shaping stego in ID3-tags file format frames «.mp3». To implement the steganographic algorithm selected FoxSecret v1.0 program with a Russian intuitive interface and freely available on the Internet. Based on the chi-square test compared the statistical distribution of significant bits files with steganographic attachments and attachments without the use of information hiding algorithms. Results tag analysis frames, divided into intervals of different lengths. An example of analysis of a single file having both a unconcealed («legal») and embedding steganographic. Numerical values scatter the statistics underlying the decision on the presence of hidden data channel.

Keywords: steganography, data protection, ID3v2.3 tags, statistical analysis, MP3 format, FoxSecret v1.0, HxD-Hex-Editor

В настоящее время широкое применение средств вычислительной техники и многообразие способов обработки информации характерны для большинства учреждений, организаций и предприятий.

Существующие меры защиты информации используют большое количество средств обнаружения несанкционированных воздействий и реагирования на угрозы. В распоряжении персонала, обеспечивающего безопасную обработку информации, как правило, имеется широкий набор свободного программного обеспечения, которое даже при недостаточном функционале позволяет провести предварительный (предшествующий выбору профессиональных программ) мониторинг и анализ и принять решение о наличии угроз обрабатываемой информации. Однако применительно к области сравнительно нового направления – стеганографии – как правило, отсутствует информации о характере воздействий и происходящих процессах, что не дает возможности определения степени опасности и приводит к неадекватному реагированию системы защиты информации [3].

В распоряжение сетевых администраторов целесообразно включить алгоритмы и/или программы первичного анализа файлов, которые помогут в выборе мер по пресечению стеганографичеких каналов передачи данных силами самого администратора или с привлечением платных программ и услуг сторонних разработчиков.

В работе рассмотрена возможность принятия решения о наличии нелегальных стеговложений на основе статистического анализа при внедрении информации в метаданные файлов формата MP3 – в теги ID3v2.3. Выбор формата файлов осуществлялся из следующих соображений:

– применение стеганографичесих алгоритмов и атак на стего-системы в печатных и интернет-изданиях рассматривает-

ся, как правило, на примере графических файлов, и злоумышленник может ожидать, что именно эти файлы будут подвергаться первоочередному анализу при поиске стего;

– по сравнению с видео, аудио-файлы используются чаще и имеют простую структуру;

– MP3 получил наибольшее распространение среди форматов аудио (благодаря наилучшему соотношению «размер-качество») и в достаточном количестве имеется на любом компьютере, как в составе программного обеспечения, так и в личных и служебных библиотеках пользователей.

Структурно тег файла MP3 состоит из общего заголовка и вложенных кадров с собственными заголовками. В спецификации тегов [7] описаны несколько десятков типов кадров, в которые может быть записана текстовая, цифровая и графическая информация (изображения обложки, название произведений, имена исполнителей и композиторов, тексты произведений, комментарии к файлу и т.д.). Такую информацию можно считать легальной (или официальной). Как правило, она считывается стандартным программным обеспечением: отображается в свойствах файлов или в окне проигрывателя, доступна программам чтения-записи тегов.

Нелегальное скрытое вложение (стего) возможно благодаря алгоритму защиты от ошибок, который при чтении тегов игнорирует кадры с нечитаемым содержимым [7]. Алгоритм создания стего достаточно прост: необходимо сформировать нечитаемый кадр (используя заголовок, отсутствующий в спецификации) и, при необходимости, исправить байты общего заголовка, содержащие информацию о размере тега. Нечитаемость кадра достигается использованием шифрования, которое затрудняет (или исключает) чтение информации при обнаружении нелегального вложения. Кроме простоты реализации, такой алгоритм обладает еще одним существенным достоинством – независимостью от размера контейнера [1] и большой пропускной способностью скрытого канала. Теоретический верхний предел размера вложения – 28-значное двоичное число (размер тега записывается в общем заголовке в байтах № 6–9, старший бит в этих байтах не используется).

Показанный алгоритм создания стего изменяет общий размер файла, однако маловероятно получить для сравнительного анализа одновременно и файл со стего и файл-оригинал. Использование шифрования изменяет статистические свойства распределения наименьшего значащего бита [6], что позволяет реализовать атаку для обнаружения стего. В [2, 8] приведены результаты вычислений значений (по критерию χ^2) для распределения пустых и заполненных контейнеров.

Интерес представляет возможность проведения подобной атаки по кадрам ID3-тегов, и сравнение результатов для файлов, содержащих как обычную информацию (изображение и текст), так и стеговложения. Анализ будет проведен для младшего разряда каждого байта. В общем виде выражение для получения значений статистики имеет вид [5]:

$$\chi^2 = \frac{(p_0 - p)^2}{p} + \frac{(p_1 - p)^2}{p},$$

где p_0, p_1 – вероятности появления соответственно «0» и «1» в младшем бите, определяемая как отношение количества «0» (или «1») к общему числу проверенных битов;
p – ожидаемая вероятность появления «0» или «1», в ходе вычислений принято равной $p = 0{,}5$, что соответствует проверке гипотезы о совпадении распределения элементов выборки с равномерным (равновероятным).

Для проведения статистического анализа были выбраны 15 файлов с пустыми заголовками – теги ID3v2.3 (объемом около 1000 байт, заполненных нулями). Продолжительность воспроизведения – минимальная, т.к. анализу подлежат только заголовки.

В качестве вложений подготовлены 15 файлов с расширением *.jpg* (размеры от 70 до 160 Кбайт) и 15 файлов с расширением *.txt* (размеры от 50 до 120 Кбайт, информация для записей в каждый файл выбиралась случайным образом из различных текстовых файлов фрагментами по 1–2 страницы с общим объемом 15–25 страниц в каждом текстовом файле-вложении).

Предварительные исследования показали, что для получения достоверных данных о порогах принятия решений необходимо провести анализ не менее 10–15 Кбайт вложений. Размер вложений выбран с избытком, с учётом сжатия информации, реализуемого стего-программами. Анализу подлежат 20 Кбайт файла с вложениями. Собственные заголовки файлов-вложений проверены и имеют только минимум служебной информации (необходимой для корректного чтения этих файлов официальными программами).

Программой Mp3TagTools v1.2 проведена запись информации в стандартные кадры тегов заголовков и сформированы две группы по 15 файлов mp3:

1) файлы с рисунком, который отображается при воспроизведении стандартным плеером, как обложка альбома (кадр APIC);

2) файлы с текстом, полностью скопированным из файлов *.txt* и внесенным в стандартный кадр COMM (кадр комментария, содержание которого отображается при просмотре свойств файла) [7].

С помощью программ реализации стеганографических алгоритмов FoxSecret v.1, также сформированы две группы по 15 файлов *.mp3* с той же информацией, что и в предыдущих группах, только вложенной в кадр с неизвестным заголовком. Эти кадры игнорируются программами просмотра тегов и стандартными плеерами. Все четыре группы файлов *.mp3* воспроизведены стандартным плеером без ошибок.

Следует отметить одну особенность программы FoxSecret v1.0: при внесении изменений в общий размер тега (байты 6–9) запись осуществляется с ошибкой – непосредственным переводом в двоичный формат размера тега (с учётом стего-вложения).

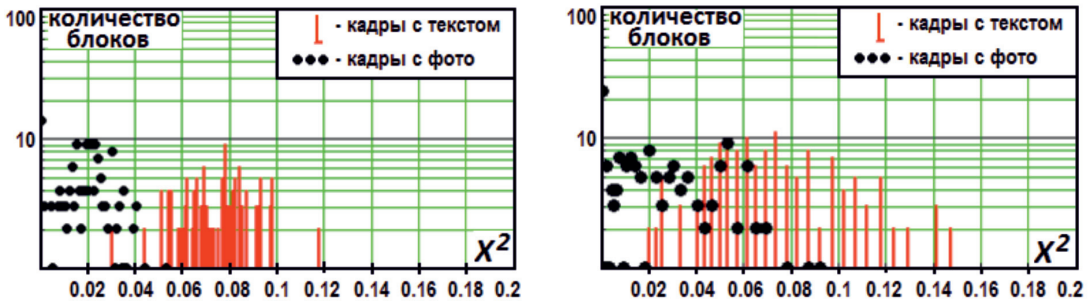
Однако согласно спецификации [7] в байтах №№ 6–9 старший бит всегда «0» и при чтении отбрасывается, при записи алгоритм должен быть обратным. Несоответствие алгоритму спецификации приводит к ошибке, заметной при достаточно больших стего-вложениях и при наличии файла-оригинала – воспроизведение файла начинается не с первой секунды, чем больше вложение, тем больше пропуск в начале воспроизведения (до 30–40 с при вложениях до 1–1,5 Мб). Содержание музыкальных кадров при этом остается неизменным, а исправление байтов №№ 6–9 в соответствии с алгоритмом спецификации (с помощью любого Hex-редактора) полностью устраняет ошибку. Выявленный недостаток является признаком первой версии FoxSecret и нехарактерен для других программ, реализующих стеганографические алгоритмы, поэтому не может быть использован в качестве признака наличия стего-контейнера (легко предположить, что подобная ошибка возможна в любом редакторе тегов).

При проведении расчетов 20 Кбайт информации, записанной в заголовках, разбивались на блоки по 250, 500 и 1000 байт и значение статистики рассчитывалось для каждого блока. Такой подход позволяет получить значения статистики на различных отрезках анализируемого файла. При анализе не учитывались различия в количестве блоков разной длины для одного и того же объема информации, что допустимо для проведения первичной обработки на основе общего характера распределения без расчета вероятностных значений.

Для блоков 1000 байт значительная часть значений статистики получена в диапазоне $\chi^2 = 0 \div 1$ (рис. 1, а), кроме того, проявляются особенности распределения официальных вложений: смещение к началу отсчета значений статистики для графических данных. При уменьшении длины блока различия текстовых и графических вложений становятся менее заметными, диапазон значений статистики расширяется (для блоков длиной 250 байт почти все результаты получены в диапазоне $\chi^2 = 0 \div 0{,}15$, рис. 1, б).

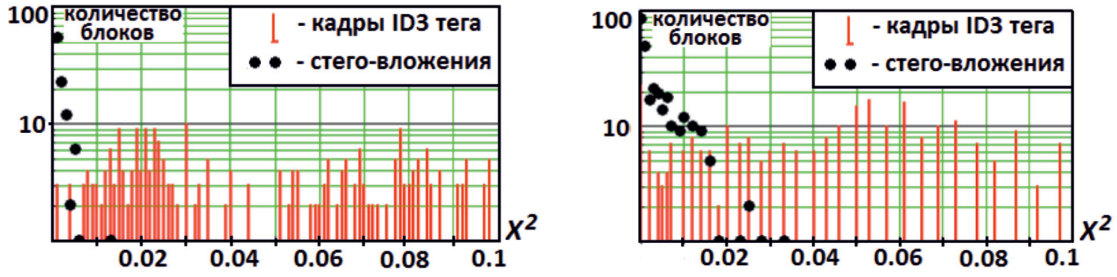
Особенности распределения стего показаны на рис. 2 в сравнении с суммарной статистикой официальных вложений. Для блоков длиной 1000 байт характерны значения χ^2 до 0,015 (рис. 2, а), более 95 % отсчетов не превышают значений $\chi^2 = 0{,}01$. При сокращении размера блока до 250 байт аналогичным 95-процентным порогом можно считать значение 0,025 (рис. 2, б).

Подобный характер распределения обусловлен использованием криптографических алгоритмов, которые, как отмечено в [2], применяются совместно со стего для дополнительного шифрования и имеют близкое к равномерному распределение значений младшего бита.



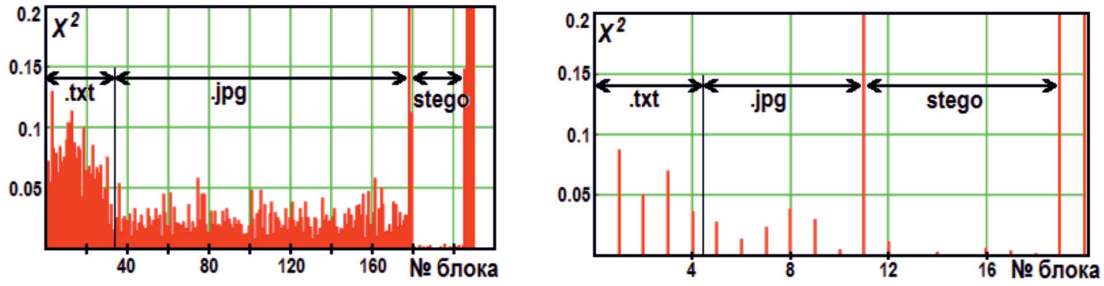
а) *б)*

Рис. 1. Статистика значений распределения χ^2 для официальных вложений



а) *б)*

Рис. 2. Сравнение статистики χ^2 для официальных вложений и стего



а) *б)*

Рис. 3. Распределение значений χ^2 с комплексным вложением в теге

Вполне логичным предположением следует считать, что использование стеганографических вложений не имеет массового характера, а наиболее вероятным будет наличие единичных файлов, передаваемых по сети или хранящихся на жестком диске компьютера. Исходя из этого предположения, был проведен анализ одного файла, в ID3 тегах которого одновременно присутствовали одно стеганографическое вложение и два официальных – текстовое и графическое. Расположение вложений в заголовке определялось программами записи автоматически. Результаты показаны на рис. 3. Правый всплеск значений получен для участка файла с записью непосредственно звуковой информации и не относящийся к рассматриваемой области тега. Предшествующий ему всплеск – область одинаковых (нулевых) байт, внесенных автоматически программой записи кадра тега. Реальные значения χ^2 для этих всплесков превышают значения 0,5 и на рисунке обрезаны по уровню 0,2 (для более наглядного представления остальных результатов). Такие всплески характеризуют, как правило, границы тега или его кадров. В ходе дальнейшего анализа они не учитывались, т.к. появление подобных выбросов значений не является стабильным, а алгоритм формирования этих участков файла требует отдельного рассмотрения.

Анализ одного тега с достаточно объемными кадрами (общий размер около 205 Кбайт) блоками по 1000 байт позволяет выделить три интервала с распределением, характерным для стеганографического и официальных вложений (рис. 3, а). Для тегов с вложениями вложений по 5–8 Кбайт (размер тега 18 Кбайт) подобный анализ возможен, но в общем случае количество отсчетов может быть недостаточным для принятия решения о наличии и характере вложений (рис. 3, б).

Анализ коротких кадров целесообразно проводить блоками меньших размеров. При сокращении размеров блоков до 250 байт, общий характер распределения полученных результатов сохраняется. Для оценки статистики распределения младшего разряда необходимо проверять участок файла размером не менее 2–2,5 Кбайт.

Полученные характеристики распределения для разрешенных способов внедрения (вложения) информации могут быть использованы в качестве математической модели части пустого контейнера [4]. Примерно 10-кратные различия статистики рас-

пределений значений χ^2 для официальных и стеганографических вложений позволяют реализовать первичную сортировку файлов и выделение предполагаемого стего, использующего дополнительное криптографическое шифрование.

Ответственность за обеспечение безопасности информации лежит, как правило, на администраторе сети, который будет определять значения порога обнаружения (в зависимости от важности информации, уровней доступа пользователей и т.д.) и возможные меры по пресечению нарушений: удаление/перезапись подозрительных участков файлов или их сохранение для последующего детального анализа.

Список литературы

1. Барсуков В.С. Стеганографический камуфляж в джунглях интернета [Текст] / В.С. Барсуков // Специальная техника. – 2005. – № 5. – С. 31–37.

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

3. Душкин А.В. Обоснование метода рационального комплексирования разнородных признаков несанкционированных воздействий на информационные системы. // Безопасность информационных технологий. – 2011. – № 1. – С. 38–43.

4. Душкин А.В., Кравченко А.С., Новосельцев В.И., Смоленцева Т.Е. Сумин В.И. Математические модели и информационные процессы управления сложным объектом: Монография. – Воронеж: Научная книга, 2014. – 125 с. – ISBN 978-5-4446-0512-7.

5. Ластивка И.А. К вопросу о проверке параметрических статистических гипотез в схемах Бернулли [Текст] / И.А. Ластивка // Молодой ученый. – 2014. – № 6. – С. 19–23.

6. Fridrich J., Du R., Long M. Steganalysis of LSB encoding in color images // ICME, 2000.

7. Nilsson M. ID3v2.3 [Электронный ресурс]: Informal Standard Document. / M. Nilsson. 1999. Режим доступа: http://id3.org/id3v2.3.0, свободный (дата обращения: 20.06.2016).

8. Provos N. Defending Against on Statistical Steganalysis // Proceeding of the 10 USENIX Security Symposium. – 2001. – P. 323–335.