УДК 004.056

Ахрамеева К.А., кандидат технических наук,
доцент кафедры «Защищенные системы связи»
Санкт-Петербургский Государственный Университет
Телекоммуникаций,
Российская Федерация, г. Санкт-Петербург
Юркин Д.В., кандидат технических наук,
доцент кафедры «Защищенные системы связи»
Санкт-Петербургский Государственный Университет
Телекоммуникаций,
Российская Федерация, г. Санкт-Петербург
Герлинг Е.Ю., кандидат технических наук,
доцент кафедры «Защищенные системы связи»
Санкт-Петербургский Государственный Университет
Телекоммуникаций,
Российская Федерация, г. Санкт-Петербург
Седельников Д.А.
студент
4 курс, факультет «Инфокоммуникационные сети и системы»
Санкт-Петербургский Государственный Университет
Телекоммуникаций,
Российская Федерация, г. Санкт-Петербург

АНАЛИЗ ФОРМАТОВ, ИСПОЛЬЗУЮЩИХ СТЕГАНОГРАФИЮ В ЭЛЕКТРОННЫХ КНИГАХ

Аннотация: *В статье рассматривается возможность использования форматов файлов, использующихся в электронных книгах, в качестве стеганографических объектов. Рассмотрена структура наиболее*

распространенных форматов текстовых документов. Предложены методы вложения дополнительной информации в анализируемые файлы.

Ключевые слова: *стеганография, pdf, fb2, DjVu.*

Annotation: *The article considers the possibility of using file formats used in e-books as steganographic objects. The structure of the most common formats of text documents is considered. Methods of embedding additional information in the analyzed files are proposed.*

Key words: *steganography, pdf, fb2, DjVu.*

Современные технологии проникли во все сферы жизни человека: от мощнейших компьютеров до бытовой техники. Чтение книг является одним из самых неотъемлемых видов досуга, соответственно технологический процесс не обошёл и его. Большое распространение получили электронные книги, которые стали вытеснять бумажные аналоги. Наделенные большими преимуществами, такими как возможность чтения в любое время суток на любых девайсах, возможность быстрого доступа к ним и оперативного обмена ими между пользователями, электронные книги оказались отличной заменой обычным бумажным книгам. Более того, современные электронные книги могут быть представлены в различных форматах, что позволяет открывать и использовать их в любых доступных системах и на любых устройствах. Именно возможности изменения их без вреда для изначально содержащейся в них информации и скрытной передачи данных через исходный код файлов и стороннее программное обеспечение сделали электронные книги интересным объектом для исследования относительно использование методов стеганографии.

Наиболее распространенным методом для передачи скрытой информации в текстовых файлах являются методы лингвистической стеганографии [1]. Лингвистическая стеганография предполагает вложение информации в текстовые документы, представленные на любом языке, таким образом, чтоб содержание основного текста не исказилось. Однако, помимо лингвистических методов стеганографии, представляет интерес сам формат файлов. Рассмотрим

на примере некоторых форматов электронных книг вложение дополнительной информации, как в саму структуру файла, так и при модификации исходного текста (лингвистическая стеганография).

1. Portable Document Format (PDF) – открытый формат электронных документов, предназначенный для представления полиграфической продукции в электронном виде.

PDF-документ образуется объектами разных типов: логические переменные, числа (целые и дробные), строки, массивы, словари, потоки, комментарии.

Структура PDF выглядит следующим образом: заголовок, объекты (**obj** данные), xref-таблица, информация об объектах, с которых необходимо начать чтение файлов.

PDF-формат является независимым от платформы форматом. Текст и изображения внутри файлов PDF отображаются одинаково на любой платформе. PDF документ состоит из множества объектов, которые определяют внешний вид и функциональность документа. Способ отображения объектов контролируется определенными командами внутри объекта, называемыми операторами. Спецификация PDF определяет множество операторов для управления отображением текста. Файлы PDF обычно сжимаются для экономии места на диске. Сжатие PDF файла не влияет на защищенность служебной информации внутри файла и не является препятствием для ее извлечения.

Анализ PDF стандарта позволяет предложить следующие методы сокрытия дополнительной информации:

- Каждый объект чередовать определенным способом, тем самым меняя структуру документа. В данном способе предполагается изменять структуру документа, не меняя содержимое. Если существует **n** объектов, то имеется **n!** различных комбинаций, следовательно, возможно передать не более $\mathbf{\log_2(n!)}$ бит данных.

- Использование межсимвольных и пробельных интервалов. Например, изменяя обычные пробельные символы в A0 для кодирования единицы и

оставляя в неизменности обычные пробельные символы для вложения нуля [2]. Либо изменяя ширину символа A0 до нуля позволяет произвести вставку любого количество таких символов в документ без опасения, что данные изменения будут визуально видны в результирующем документе. Таким образом, между двумя любыми символами в документе встраивается несколько неразрывных пробелов нулевой длины, количество, которых кодирует необходимый ASCII символ. [3].

- Инкрементальные обновления. PDF формат позволяет использовать инкрементальные обновления для хранения различных версий одного документа, то есть небольшие части документа, которые содержат изменения по сравнению с предыдущей версий. Например, можно встраивать данные[2]:

- изменяя текст видимым образом, затем записывать инкрементальное обновление, содержащее исходные данные PDF, поэтому измененный текст фактически не отображается;
- записывать инкрементальные обновления для объектов, которые не существуют в исходных данных, поэтому обновление не имеет никакого эффекта. Данные встроенные в значение объектов потока, используются в обновлении;
- записывать инкрементальные обновления с помощью заданной длины для нескольких объектов; следовательно, данные можно получить, прочитав раздел перекрестной ссылки обновления, который включает начальный адрес каждого обновленного объекта.

- Выравнивание текста и TJ оператор. В работе [4] утверждается, что для выравнивания текста (слева и справа) используется PDF writer, который генерирует случайные значения для TJ операторов. В таком случае можно скрыть данные в наименее значимых битах этих значений TJ оператора (в случае, когда значения TJ оператора являются случайными и не содержат шаблон). Оператор TJ используется для отображения текстовых строк в файле PDF. Он содержит массив строк и чисел, состоящих из символов и значений пространства, используемых между этими символами. Каждое значение

пространства между символами вычитается из текущей текстовой позиции, которая сдвигает соответствующую строку влево на это значение (или вправо, в случае отрицательного значения)[3].

2. FictionBook (FB2) – формат представления электронных версий книг в виде XML-документов, где каждый элемент книги описывается своими тегами. Стандарт призван обеспечить совместимость с любыми устройствами и форматами. XML позволяет легко создавать документы, готовые к непосредственному использованию и программной обработке (преобразованию, хранению, управлению) в любой среде. Документы, обычно имеющие расширение fb2, могут содержать структурную разметку основных элементов текста, некоторое количество информации о книге, а также вложения с двоичными файлами, в которых могут храниться иллюстрации, например обложка.

Возможные способы внедрения дополнительной информации:

- формат может содержать binary тег, внутри которого содержится изображение. В таком случае есть возможность произвести вложение дополнительной информации методами стеганографии для изображений. Однако, выбор подходящего метода, требует дополнительных исследований;

- лингвистические методы вложения, например, увеличение количества пробелов и иных символы (например, нижнее подчеркивание) за пределами тегов или внутри самих тегов.

3. DjVu – технология сжатия изображений с потерями, разработанная специально для хранения сканированных документов – книг, журналов, рукописей и прочее, где обилие формул, схем, рисунков и рукописных символов делает чрезвычайно трудоёмким их полноценное распознавание.

В DjVu используется алгоритм JB2, который ищет повторяющиеся символы и сохраняет их изображение только один раз. Соответственно имеется перечень возможных способов сокрытия:

- выбрать множество всех похожих символов и выбрать один с помощью хеш-стеганографии;

- выбирать два символа вместо одного. Первый символ считать передающим 0, а второй символ считать передающим 1. С помощью «чередования» можно передавать сокрытую информацию;

- скрывать данные внутри самой картинки, обозначающей символ в DjVu, например, с помощью метода с вложением в наименее значащие биты;

- использовать методы стеганографии, предполагающие вложение дополнительной информации в шумы сканера, изменение расстояний между строками и т.д.

Электронные книги сегодня получают все более широкое распространение. Поэтому ожидаемо, что форматы электронных книг будут использоваться для организации скрытой передачи информации методами стеганографии. Важно понимать какие алгоритмы могут быть использованы для вложения сообщений, как для совершенствования уже существующих методов, так и для разработки методов стегоанализа. Анализ наиболее распространенных форматов, использующихся в электронных книгах показал, что помимо методов лингвистической стеганографии, возможно использовать для вложения и особенности структуры формата файлов.

Использованные источники:

1. Коржик В.И., Небаева К.А., Герлинг Е.Ю., Догиль П.С., Федянин И.А. Цифровая стеганография и цифровые водяные знаки. Часть 1. Цифровая стеганография. [монография]: – СПбГУТ. – СПб., 2016 – 226 с.

2. Lee I-Shi. A new approach to covert communication via pdf files / I-Shi Lee, Wen-Hsiang Tsa // Signal Processing. – 2010. – P. 557–565.

3. Карачанская Е., Использование стеганографии для сокрытия сообщения внутри PDF-файлов. // Коношко К. // Евразийское Научное Объединение. С. 1-2.

4. Zhong Shangping. Data hiding in a kind of pdf texts for secret communication / Shangping Zhong, Xueqi Cheng, Tierui Chen // International Journal of Network Security. – 2007. –№ 4(1). – P. 17–26.