УДК004.056

Студ. А. А. Сущеня
Науч. рук. проф. П. П. Урбанович
(кафедра информационных систем и технологий, БГТУ)

ПРОГРАММНОЕ СРЕДСТВО СТЕГАНОГРАФИЧЕСКОГО ПРЕОБРАЗОВАНИЯ ТЕКСТОВ-КОНТЕЙНЕРОВ НА ОСНОВЕ ЯЗЫКА РАЗМЕТКИ XML

Стеганография − это наука о скрытой передаче информации при условии, что сам факт передачи сохраняется в тайне. Главная задача – сделать так, чтобы человек, анализирующий сообщение, не подозревал, что внутри передаваемых данных, внешне не представляющих ценности, содержится скрытая информация

Компьютерная стеганография позволяет скрывать информацию внутри медиафайлов, то есть файлов, содержащих изображение, звук или текст [1]. Такие файлы называются файлами-контейнерами. В качестве контейнера, содержащего информацию, можно использовать документ, основанный на языке XML [2].

XML (eXtensible Markup Language) — это язык описания документов, во многом похожий на язык разметки гипертекста HTML, но гораздо более универсальный. XML используется для конструирования Web-страниц. XML рекомендован Консорциумом Всемирной паутины и фактически представляет собой свод общих синтаксических правил. XML — текстовый формат, предназначенный для хранения структурированных данных, для обмена информацией между программами, а также для создания на его основе более специализированных языков разметки.

Достоинства XML. XML — язык разметки, позволяющий отобразить двоичные данные в текст, читаемый человеком и анализируемый компьютером; XML поддерживает Юникод; в формате XML могут быть описаны такие структуры данных как записи, списки и деревья; XML имеет строго определённый синтаксис и требования к анализу, что позволяет ему оставаться простым, эффективным и непротиворечивым.

Недостатки XML. Синтаксис XML избыточен. Размер XML-документа существенно больше бинарного представления тех же данных. Размер XML-документа существенно больше, чем документа в альтернативных текстовых форматах передачи данных (например, JSON, YAML) и особенно в форматах данных, оптимизированных для конкретного случая использования.

Зачастую XML используется скорее в качестве языка разметки, а не формата данных. При описании внешнего вида документа, как правило, используются атрибуты, что подразумевает наличие большого числа кавычек. Эта особенность позволяет при помощи определенного алгоритма разместить в файле XML информацию, никак не влияющую на семантику документа [3-4].

Ввиду того, что синтаксис XML избыточен, можно рассматривать данный формат в качестве стеганографического контейнера.

Формат DOCX представляет собой модернизированную версию формата DOC, причем по сравнению со своим предшественником этот формат гораздо более популярен и доступен. В отличие от файлов DOC, формат DOCX не является расширенным файловым форматом. Он представляет собой файл-архив. Формат файла основан на Open XML и использует сжатие по алгоритму *ZIP* для уменьшения размера файла.

Исходя из того, что DOCX файл является ZIP архивом с XML документами, можно использовать данный формат для осаждения тайной информации. В качестве примера для демонстрации возможности осаждения тайной информации в DOCX контейнер, было создано программное средство MarkupStego.

Программное средство реализовано при помощи интерфейса программирования приложений Windows Forms на языке C#.

Основные сущности для реализации внедрения\извлечения информации: XMLFile, DOCXFile, FileManager, Embedder, Extracter (рисунок 1).

XMLFile – в своем составе имеет свойство File, представляющее собой загруженный файл XML, а также метод GetContainerCapacity, позволяющий в зависимости от языка сообщения вычислять размер контейнера; DOCXFile – сущность представляющая собой абстракцию над документом Microsoft Office Word; FileManager – предназначенный для манипулирования загрузкой документов, находящихся в файловой системе копьютера; Embedder – сущность реализующая алгоритм внедрения информации в XML-документ. Метод MakeBinaryString конвертирует сообщение в двоичный код. Метод EmbedMessage, последовательно проходя по сообщению и ставя в соответствие очередную пару кавычек, изменяет ее при необходимости.Extracter при помощи ExtractMessage извлекает бинарную последовательность из прочитанного документа, пока сообщение не закончится (проверяет IsEndOfMessage). После извлечения, использовав метод RestoreMessage, из бинарной последовательности получаем исходное сообщение.

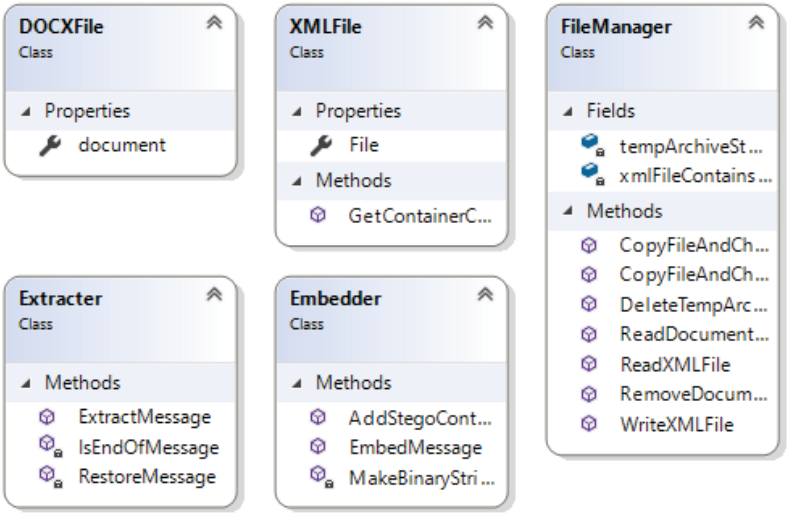

Рисунок 1 – Классы, реализующие основную логику приложения

Перед началом осаждения необходимо указать язык сообщения. Данная возможность была введена для рационального использования размера контейнера. Так как для осаждения русского символа требуется больше бит, то нет смысла тратить место впустую, если сообщение на английском.

После того, как контейнер был загружен в приложение, становиться доступной информация о его емкости. Осаждение информации, превышающей емкость контейнера не возможно. Внедряемое в контейнер сообщение вводиться в поле «Встраиваемое сообщение». После завершения ввода сообщения, для внедрения информации необходимо нажать кнопку «Встроить».

Для извлечения внедренного сообщения следует нажать на кнопку «Извлечь сообщение», после чего в файловой системе необходимо указать файл-контейнер содержаний сообщение (рисунок 2).

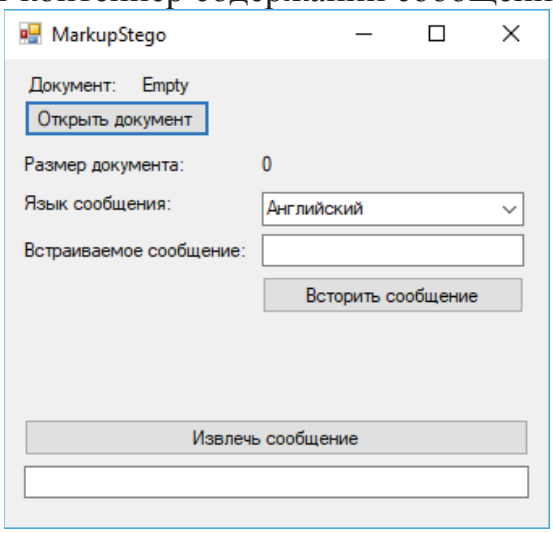

Рисунок 2 – Интерфейс приложения MarkupStego

ЛИТЕРАТУРА

1. Урбанович П.П. Защита информации методами криптографии, стеганографии и обфускации/ П.П. Урбанович. – Минск: БГТУ. – 2016. – 220 с.

2. Сущеня, А. А. Стеганографическое преобразование текстов-контейнеров на основе языков разметки / А. А. Сущеня // 68-я научно-техническая конференция учащихся, студентов и магистрантов БГТУ, 17-22 апреля, Минск: сборник научных работ: в 4 ч., Ч.4/ Белорусский государственный технологический университет. – Минск : БГТУ, 2017. – С. 145-149.

3. Интернет-портал [Электронный ресурс]/ Стеганография в XML, HTML: метод изменения порядка следования атрибутов в файлах с разметкой. – Режим доступа: http://www.nestego.ru/ 2012/05/xml-html.html. – Дата доступа: 20.03.1017.

4. Интернет-портал [Электронный ресурс]/ Текстовая стеганография. Метод хвостовых пробелов. – Режим доступа: http://www.nestego.ru/2012/05/blog-post_03.html. – Дата доступа: 20.03.1017.

УДК004.056

Студ. О.С. Михоленко
Науч. рук. проф. П. П. Урбанович
(кафедра информационных систем и технологий, БГТУ)

ПРОГРАММНОЕ СРЕДСТВО ДЛЯ АНАЛИЗА ПРОСТРАНСТВЕННО-ГЕОМЕТРИЧЕСКИХ ПАРАМЕТРОВ ТЕКСТОВЫХ ДОКУМЕНТОВ

Одним из способов решения проблемы защиты передаваемой информации является стеганография. Стеганография – способ передачи или хранения информации с учётом сохранения в тайне самого факта такой передачи (хранения). Информация, которую необходимо передать защищенно (скрыть), помещается (осаждается) в контейнер – данные (изображения, текст, видеофайлы и т.д.), используемые для осаждения информации. Само же осаждаемое сообщение при этом называется стегосообщением.

Большую часть передаваемой информации составляет текст, а наиболее распространенным форматом, используемым для представления текстовых данных, является DOCX. Следовательно, в качестве контейнера для осаждения информации данный формат подходит больше всего.