УДК 003.26

СТЕГАНОГРАФИЧЕСКАЯ СИСТЕМА НА ОСНОВЕ КОМБИНИРОВАНИЯ МЕТОДОВ ДЛЯ ЭЛЕКТРОННОГО ТЕКСТОВОГО ДОКУМЕНТА

Блинова Е. А., Сущеня А.А.
Белорусский государственный технологический университет

В настоящее время является актуальной задача передачи скрытых сообщений в открытых источниках или размещения скрытых меток в открытых данных для подтверждения авторства на эти данные. Методы, реализующие такое скрытие, называются стеганографическими методами, данные, в которых размещаются скрытые сообщения, — стеганографическими контейнерами, а сами скрытые сообщения — стегосообщениями. Под стеганографическим ключом понимается место и порядок скрытия сообщения в открытых данных. Стеганографическая система объединяет все вышеперечисленное. Математическая модель стеганографической системы может быть представлена в следующем виде:

$$S = Emb(C, M, K), \quad (1)$$

$$M = Ext(S, K), \quad (2)$$

где C — множество всех контейнеров, K — множество стеганографических ключей, M — множество скрытых сообщений, S — множество стего, $Emb()$ и $Ext()$ – функции встраивания скрытого сообщения в файл-контейнер и извлечения из файла-контейнера соответственно.

Для скрытия информации или осаждения скрытых меток используются различные виды файлов-контейнеров: текстовые документы в разнообразных форматах, изображения, звук, видео. Для каждого типа файлов-контейнеров разработаны разнообразные методы, комбинирующие стандартные синтаксические методы текстовой стеганографии и методы, основанные на специфических свойствах документа-контейнера, например, осаждение скрытой информации в метаданных изображения или особенностях форматирования текста электронного текстового документа.

Основными направлениями применения стеганографических методов являются внесение различных стеганографических меток в каждую копию электронного документа (Digital Fingerprint), внесение одинаковых стеганографических меток во все копии документа (Watermaking) и скрытая передача и хранение данных. Следует отметить, что при стеганографическом преобразовании данные не шифруются, однако, часто предполагается, что скрытое сообщение может быть предварительно зашифровано криптографическими методами для дополнительной защиты данных.

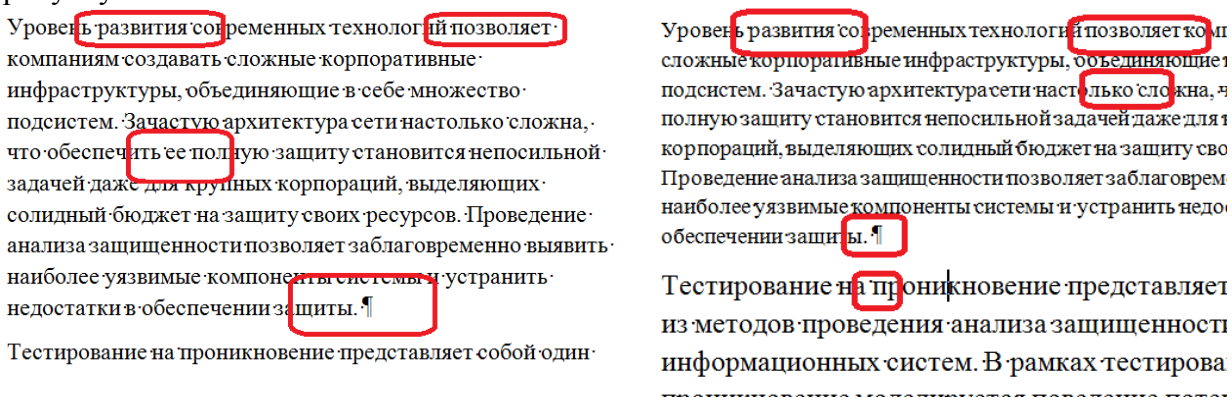
В данном докладе рассматривается совместное применение двух различных стеганографических методов осаждения скрытой информации в электронных документах Microsoft Word формата .DOCX. В связи с широким распространением, электронные документы формата Microsoft Office часто используются в качестве файлов-контейнеров. Для них применяются методы, которые используют наравне с классическими методами текстовой стеганографии методы, свойственные контейнеру, такие как формат и смещение текста, размещение диакритических знаков, наличие истории редактирования и прочей служебной информации, что позволяет добиться увеличения скрытности и пропускной способности [1].

Основной проблемой при применении стеганографических методов для осаждения скрытой информации в электронных документах Microsoft Word формата .DOCX с использованием специфических методов, свойственных контейнеру, является проблема разрушения скрытого сообщения после изменения форматирования текста. В связи с этим, предлагается использовать два стеганографических метода, использующих различные свойства контейнера, для взаимного контроля друг друга. Один метод использует изменения межстрочного расстояния для неотображаемых символов, а второй — особенность описания электронного документа Microsoft Word формата .DOCX в формате XML.

Метод изменения межстрочного расстояния, или line-shift coding, успешно применялся для маркирования технической документации для предотвращения утечек со стороны допу-

щенных к ней специалистов. В его стандартной реализации предлагалось скрывать стегосообщение в изменении высоты межстрочных интервалов, причем для каждой копии документа выбирался свой набор межстрочных интервалов, что позволяло выявить источник несанкционированных копий. Однако такой метод имеет несколько существенных недостатков: он обладает малой пропускной способностью и может быть выявлен для электронного документа путем изменения параметров размера и начертания шрифта. Была предложена модификация стеганографического метода изменения межстрочного расстояния электронного документа, заключающаяся в том, чтобы производить смещение не всей строки, а только неотображаемых символов (пробелов, табуляций, знаков переноса строки, неразрывных пробелов, абзацев и т.д.) [2-3]. В качестве редактора электронных текстовых документов использовался редактор Microsoft Word 2010, однако изменение высоты строки, как для полной строки, так и для отдельных символов существует и в других текстовых редакторах. В Microsoft Word такое смещение производится как *Шрифт/Интервал/Смещение*.

На рисунке 1а изображен текст со смещением некоторых неотображаемых символов – пробелов и знаков абзаца, красным выделены отдельные места, в которых произведено изменение межстрочного расстояния. Часть позиций текста, в которые производится осаждение, при этом не отмечена, и можно попытаться их найти самостоятельно или обратиться к рисунку 1б.



а *б*

Рисунок 1 а – Текст со смещенными неотображаемыми символами;
б – Текст со смещенными неотображаемыми символами
с изменением начертания и размера шрифта

Отметим, что, как видно из рисунка 1б, изменение начертания и размера шрифта не влияют на отображение электронного текста из-за особенностей реализации контейнера. Также отметим, что стандартными средствами текстового редактора различные высоты смещения символов текста не определяются, в отличие от других свойств формата (размера, начертания и пр.), и могут быть определены только визуально, либо программно.

Второй метод, использующий особенности описания электронного документа Microsoft Word формата .DOCX в формате XML, состоит в следующем. Файл формата .DOCX не является расширенным файловым форматом, а представляет собой архив. Формат файла основан на Open XML, подробно описанный в стандарте ECMA-376: Office Open XML File Formats, и использует сжатие по алгоритму ZIP для уменьшения размера файла. Данный архив содержит два типа файлов — файлы формата XML с расширениями xml и rels и медиафайлы, например, изображения. Логически файл состоит из трех видов элементов: типов, частей и связей. Типы — это список сущностей, встречающихся в документе, например, типов медиафайлов или частей документов, части — это отдельные части документа, для каждой части документа создан отдельный файл формата XML. Между частями документа устанавливаются связи. Таким образом, можно сказать, что файл формата .DOCX представляет собой набор сжатых файлов формата XML, причем все текстовое содержимое электронного документа Microsoft Word формата .DOCX находится в одном XML файле, а именно в document.xml. Файл document.xml представляет собой XML файл в элементной форме, где каждому элементу обычно соответствует один атрибут. Теги начинаются с «w:» и обозначают:

- <w:document> — тег свойства документа, указываются пространства имен, используемые при построении XML файла;
- <w:body> — тег тела документа, является корневым тегом для частей документа;
- <w:p> — тег абзаца документа, где указываются свойства абзаца, такие как выравнивание, абзацные отступы и т.д.;
- <w:r> — тег фрагмента текста, для которого указываются особенности форматирваония данного участка текста, такие как размер шрифта, высота межстрочного интервала, цвет и т.д.;
- <w:t> — тег текста, в котором содержится текст части документа.

Теги <w:p> и <w:r> содержат вложенный тег <w:s**ection**Pr> для описания особенностей форматирования именно этого участка. Например, тег описания свойств абзаца <w:pPr> содержит в себе вложенный тег описания интерлиньяжа <w:spacing w:lineRule="exact" w:line="360"/>, который обозначает, что высота межстрочного интервала задана точно и составляет 18 пунктов, так как параметр «w:line» измеряется в двадцатых долях пункта. Для описания форматирования отдельных символов используется тег <w:rPr>. Например, в следующей конструкции <w:rPr> <w:sz w:val="28"/> <w:szCs w:val="28"/> </w:rPr> параметр <w:sz w:val="28"/> измеряется в ½ пункта и в данном случае указывает, что кегль данного участка текста равен 14 пунктам, а параметр <w:szCs w:val="28"/> используется для отображения специфических шрифтов, например арабского.

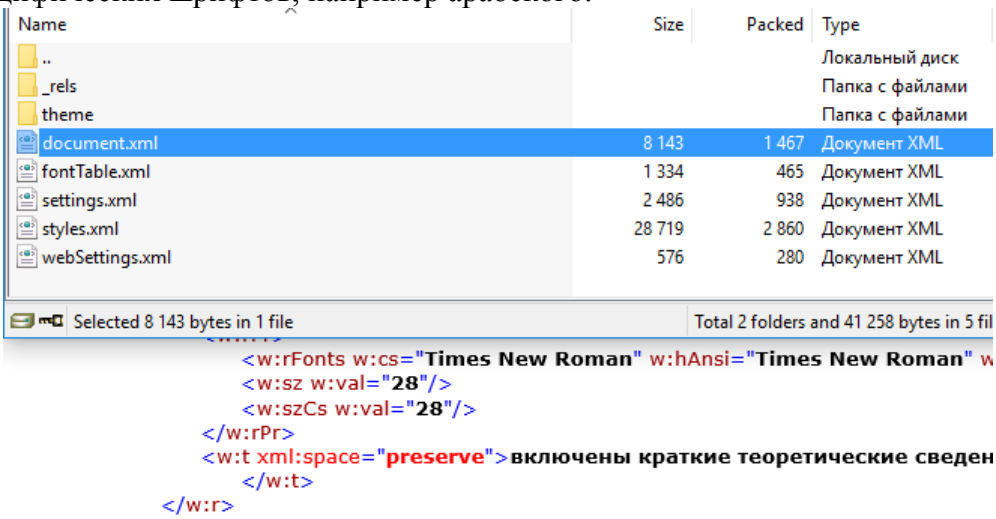


Рисунок 2 – Структура документа формата .DOCX

Известно, что интерпретация XML документа допускает различный регистр тегов и порядок следования атрибутов. Кроме того, XML документ безразличен к типу кавычек — для обрамления значений атрибутов могут использоваться как двойные, так и одинарные кавычки, причем при визуальном анализе документа Microsoft Word со стороны пользователя никаких отличий видно не будет. Был предложен стеганографический метод замены типа кавычек с двойных на одинарные в XML документе [4-8].

Таким образом, существуют два метода, каждый из которых позволяет осадить некоторое скрытое сообщение, используя особенности формата файла электронного документа. Будем использовать один из методов для осаждения скрытой информации в файл контейнер, а второй — для контроля целостности файла контейнера. Для осаждения сообщения будем приводить его к виду бинарной последовательности. Выбор метода осаждения выполняется исходя из емкости контейнера, которую можно рассчитать следующим образом. Подсчитывается количество неотображаемых символов в файле контейнере. Визуально незаметное смещение может производиться в диапазоне +/-1 пункт, что дает 6 бит скрытого сообщения на 3 неотображаемых символа. Для метода замены типа кавычек подсчитывается количество пар кавычек в файле document.xml, одна пара кавычек соответствует 1 биту скрытого сообщения.

Алгоритм осаждения скрытого сообщения в файл контейнер изображен на рисунке 3.

Обозначим метод изменения межстрочного расстояния для неотображаемых символов как S, а метод замены типа кавычек с двойных на одинарные — Q.

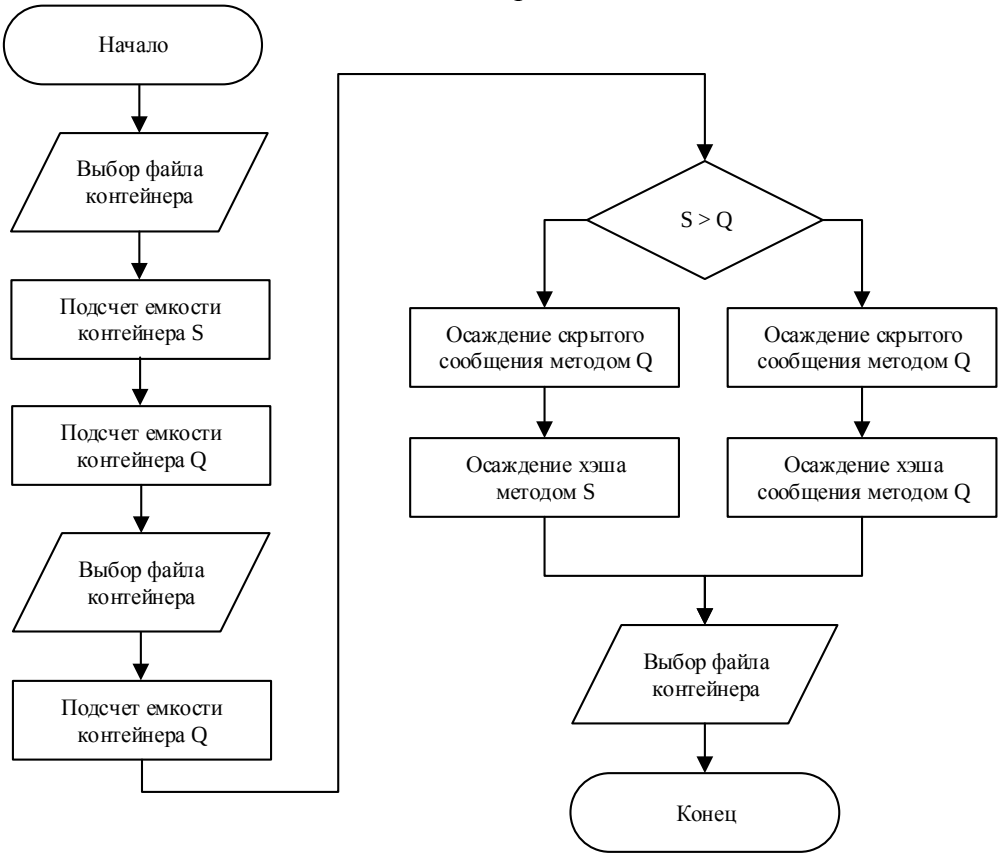


Рисунок 3 – Алгоритм осаждения скрытого сообщения в документ формата .DOCX

Для реализации осаждения скрытых сообщений в документы формата .DOCX был разработан программный продукт SpaceQuoteStego.

SpaceQuoteStego позволяет создавать стеганографические контейнеры на основе электронных документов формата .DOCX. В этом программном средстве реализован вышеописанный алгоритм осаждения скрытого сообщения в электронных документах формата .DOCX с некоторыми ограничениями: из неотображаемых символов рассматриваются только пробелы, доступно осаждение только буквенно-цифровых комбинаций и пробелов, хеширование производится по методу MD5 [9].

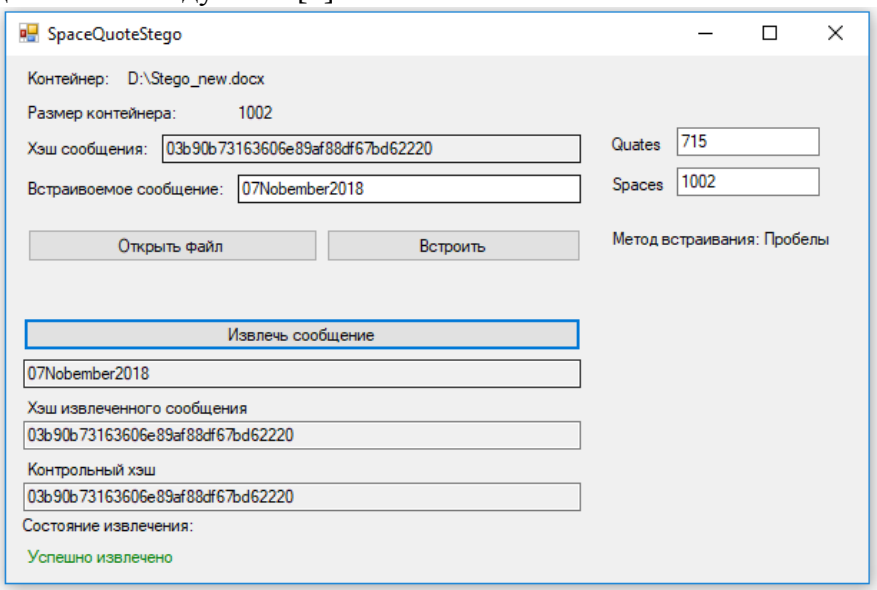


Рисунок 4 – Интерфейс программного средства SpaceQuoteStego

В докладе рассмотрена стеганографическая система комбинированного применения

двух стеганографических методов, основанная на различных свойствах электронного текстового документа Microsoft Word формата .DOCX. Каждый из методов может быть использован либо для осаждения данных, либо для хранения значения хэша данных, чем осуществляется проверка целостности сообщения при внесении изменений в файл контейнер. Одним из методов является метод изменения межстрочного расстояния для неотображаемых символов, второй использует особенности описания электронного документа в формате XML. Разработано приложение, позволяющее создавать стеганографические контейнеры из электронных документов с использованием данного подхода, что может быть применено для скрытой передачи и хранения данных и подтверждения права собственности на информацию, представленную в цифровом виде.

Литература

1. Урбанович, П. П. Защита информации методами криптографии, стеганографии и обфускации: учеб.-метод. пособие / П. П. Урбанович. – Минск: БГТУ, 2016. – 220 с.
2. Блинова, Е.А. Стеганографический метод на основе изменения межстрочного расстояния неотображаемых символов строк электронного текстового документа// Материалы 80 конференции профессорско-преподавательского состава БГТУ . – Минск. – 2016. – с. 11.
3. Блинова, Е.А. Стеганографический метод на основе изменения междустрочного расстояния неотображаемых символов строк электронного текстового документа// Труды БГТУ. Сер. Физико-мат. науки и информатика № 6. — Минск: БГТУ. — 2016. — С.166-169.
4. Сущеня, А. А. Стеганографическое преобразование текстов-контейнеров на основе языков разметки/ А. А. Сущеня // 68-я научно-техническая конференция учащихся, студентов и магистрантов, 17-22 апреля, Минск: сборник научных работ: в 4 ч. Ч. 4 / Белорусский государственный технологический университет. — Минск: БГТУ, 2017. — С. 145-149.
5. Сущеня, А.А. Способ стеганографического осаждения информации в документ с расширением .DOCX / А. А. Сущеня // XXI Республиканская научная конференция студентов и аспирантов, 19–21 марта, Гомель: сборник научных работ / Гомельский государственный университет имени Ф. Скорины. – С. 303-304.
6. Сущеня, А.А. Идея и архитектура веб-приложения, использующего в качестве стеганографического контейнера документы формата DOCX / А. А. Сущеня // Международная научно-практическая конференция, 14–18 мая, Минск: сборник научных работ / Белорусский государственный университет. – С. 170.
7. Сущеня, А.А., Блинова Е.А., Урбанович П.П. Модификация стеганографического метода изменения междустрочного расстояния электронного документа // Технические средства защиты информации: Тезисы докладов XVI Белорусско-российской научно-технической конференции, 5 июня 2018 г., Минск. Минск: БГУИР, 2018. – С 90-91.
8. Сущеня, А. А. Программное средство стеганографического преобразования текстов-контейнеров на основе языка разметки XML / А. А. Сущеня // 69-я научно-техническая конференция учащихся, студентов и магистрантов, 2-13 апреля, Минск : сборник научных работ: в 4 ч. Ч. 4 / Белорусский государственный технологический университет. - Минск : БГТУ, 2018. - С. 81-84.
9. WhiteSpaceStego [Электронный ресурс]: https://github.com/bntdeep/WhiteSpaceStego Дата доступа: 07.11.2018.