УДК 004.56+003.26

М.Г. Савельева, асп. (БГТУ, г. Минск)

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СИНТАКСИЧЕСКИХ МЕТОДОВ ТЕКСТОВОЙ СТЕГАНОГРАФИИ

Стеганография – это наука о способах передачи скрытой информации, при которых скрытый канал организуется на базе и внутри открытого канала с применением особенностей восприятия информации.

Исследования и разработки в области стеганографии становятся всё более популярными в современном информационном обществе наряду с широким использованием цифровых форматов мультимедиа и существующими проблемами управления цифровыми ресурсами и контроля использования прав собственности на компьютерные файлы.

Стеганография, использующая текстовые контейнеры для скрытия данных, называется текстовой [1]. При скрытии информации используются допущения при расположении и количестве символов в тексте, не учитываемые при прочтении человеком и компьютерном анализе файла. Методы текстовой стеганографии можно разделить на три группы: синтаксические, случайная и статическая генерация и лингвистические [2].

Синтаксические методы основаны на использовании особенностей пунктуации, аббревиатуры и сокращения. К синтаксическим методам относят также методы, основанные на изменении стиля и структуры предложения без заметного искажения исходной смысловой нагрузки. Недостатками можно считать лёгкую обнаруживаемость и высокую вероятность разрушения скрытого сообщения при повторном наборе текста или использовании более сложных текстовых редакторов [3].

В данной работе были рассмотрены синтаксические методы текстовой стеганографии, использующие в качестве контейнера документы, созданные в программе Microsoft Word, в связи с их широким распространением. Для них применяются методы, которые используют наравне с классическими методами текстовой стеганографии и методы, свойственные контейнеру, такие как особое форматирование и смещение текста, наличие истории редактирования и прочей служебной информации, что позволяет добиться увеличения скрытности и пропускной способности. К таким методам относятся следующие:

– Line Shift Coding – метод использующий изменение расстояния между строками электронного текста.

– Word Shift Coding – метод использующий изменение расстояния между словами в одной строке электронного текста.

– Feature Coding – методы, использующие внесение специфических изменений в шрифты (начертания отдельных букв). Примеры включают: удлинение или укорочение конечной части конкретных символов, таких как h, d. b; изменение размера точки в символах, таких как i, j и т. д. Классификация методов представлена на рисунке 1. Некоторые из методов описаны ниже:

- Luminance Modulation Coding – метод использующий изменение модуляции яркости символов.
- Метод невидимых символов заключается в встраивании секретного сообщения в цвет невидимых символов (пробелы, табуляции, новые строки) в формате RGB.
- Метод подчеркивания символов для MS Word. Подчеркивание символов добавляет невидимые стили подчеркивания к символам, поэтому каждый символ может нести 8 секретных битов.
- Метод масштабирования символов для MS Word. Шкала текстовых символов по умолчанию составляет 100%. Можно скрыть 1 бит на символ, используя 99%-ную шкалу или 101%-ную.

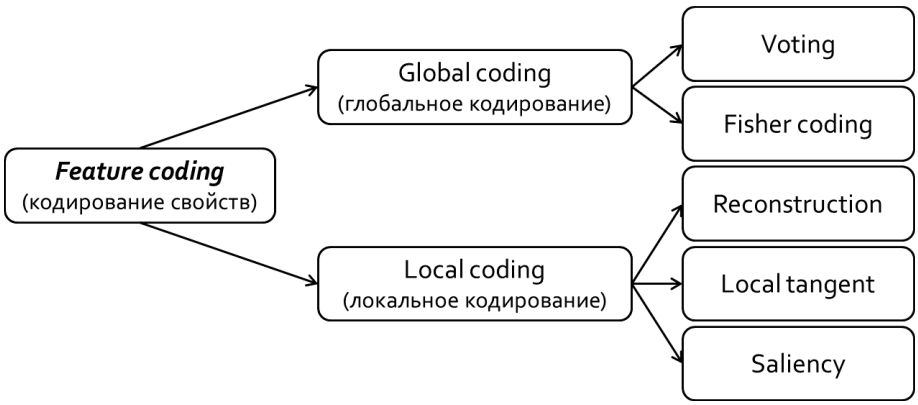

Рисунок 1 – Методы группы Feature Coding

– Property Coding методы используют свойства объектов документа, отличные от символов (например, границы абзаца), в качестве носителя секретной информации:

- Метод границ абзаца использует возможность добавить левую и правую границы к абзацу и раскрасить их цветами, представленными компонентами (R, G, B), где R, G, B > 249 не может быть отличен от белого цвета (255, 255, 255) для человеческого глаза.
- Метод границ предложения использует возможность добавления левых и правых границ к предложению с цветами, не заметными пользователю [4].

Для анализа используем документ MS Word в качестве документа-контейнера. Документ описан в таблице 1.

Таблица 1 – Характеристики трех документов-контейнеров

	Документ
Страницы	3
Слова	1025
Знаки без пробелов	7259
Знаки с пробелами	8274
Символы	8309
Абзацы	35
Строки	135
Предложения	65
Размер (бит)	141584

Таблица 2 – Сравнение максимального количества встроенных битов/символов в некоторых из описанных методов

	Документ
Метод Luminance Modulation Coding	7 259
Метод невидимых символов	8309 – 7259 = 1 050
Метод подчеркивания символов	8274 × 8 = 66 192
Метод масштабирования символов	8 274
Метод границ абзаца	35 × 28 = 980
Метод границ предложения	65 × 7 = 455

Из таблицы 2 видно, что метод подчеркивания символов имеет наибольшую ёмкость встраивания, за ним следует метод масштабирования символов, а наименьшую ёмкость встраивания имеет метод границ предложения. Это объясняется тем, что метод подчеркивания символов встраивает до 8 бит в 1 символ, метод масштабирования символов встраивает 1 бит в 1 символ, а метод границ предложения встраивает до 7 бит в одно предложение, которое включает в себя в среднем 125 символов.

Некоторые методы стеганографии текста, такие как кодирование со сдвигом строк, кодирование со сдвигом слов и кодирование с модуляцией яркости надежны для печати и сканирования документов, но большинство имеет низкую ёмкость встраивания. Другие методы, имеют более высокую ёмкость встраивания, но менее или совсем не устойчивы к печати и сканированию документов. Кодирование свойств относится ко второй группе, и оно совсем не устойчиво к пе-

чати и сканированию документов. Кодирование свойств устойчиво к действиям по сохранению, а также имеет меньшие накладные расходы.

Скрытый текст с кодировкой свойств может быть изменен или уничтожен путем редактирования текста. Наличие методов подчеркивания символов, границ абзаца и границ предложений можно легко обнаружить, если кто-то намеренно изменит цвет фона документа, в результате чего границы и подчеркивание станут видимыми. Методы масштабирования символов и изменения яркости устойчив к такого рода атакам.

Кодирование свойств не совсем подходит для приложений защиты авторских прав, где требуется надежное сокрытие данных, поскольку злоумышленник всегда может использовать оптическое распознавание символов для полного удаления скрытых данных.

Таким образом после сравнительного анализа можно сделать вывод, что из всех рассмотренных методов оптимальным вариантом для обеспечения тайны переписки и защиты авторских прав документа по совокупности факторов (ёмкость встраивания, устойчивость к печати и сканированию документов, некоторые виды атак, в частности изменение фона документа) является метод модуляции яркости (метод Luminance Modulation Coding).

ЛИТЕРАТУРА

1. Information Hiding Techniques for Steganography and Digital Watermarking / Ed. Stefan Katzenbeisser, Fabien A. P. Petitcolas. – London: Artech House, Inc., 2000. – P. 110-113.

2. Bennett, K. Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text / K. Bennett // cERIAS Tech Report. – 2004. – 13 p.

3. Урбанович, П. П. Защита информации методами криптографии, стеганографии и обфускации: учеб.-метод. пособие / П. П. Урбанович. – Минск: БГТУ, 2016. – 220 с.

4. Stojanov, I. A New Property Coding in Text Steganography of Microsoft Word Documents / I. Stojanov, A. Mileva, I. Stojanovic // In Securware 2014: The Eighth International Conference on Emerging Security Information, Systems and Technologies. – 2014. – P. 25-30.