DOI 10.37882/2223–2966.2023.02.30

МЕТОДЫ СТЕГАНОГРАФИИ, РЕАЛИЗУЕМЫЕ В ТЕКСТОВЫХ ДОКУМЕНТАХ MS WORD ПРИ РАБОТЕ СО СТАНДАРТНЫМИ (НЕ СПЕЦИАЛИЗИРОВАННЫМИ) ПРОГРАММАМИ

STEGANOGRAPHY TECHNIQUE REALISED IN MS WORD DOCUMENTS IMPLEMENTED WITH STANDARD (NON-SPECIALISED) PROGRAMS

V. Mikhaelis
S. Mikhaelis

Summary. Most of modern research in the field of steganography either focuses on the concept of concealed information embedded in multimedia containers (images, audio and video files) of different formats, or is aimed at the use of telecommunication networks (network steganography). The present paper represents an initial stage of research in the field of textual steganography and considers modification of textual colour parameters as well as information retrieval from the container without the use of special software.

Keywords: textual steganography, text colour parameters modification.

Михаэлис Владимир Вячеславович
Кандидат педагогических наук, доцент, ФГБОУ ВО «Иркутский государственный университет путей сообщения»
mvv_1967@mail.ru
Михаэлис Светлана Ивановна
Кандидат педагогических наук, доцент, ФГБОУ ВО «Иркутский государственный университет путей сообщения»
msibgu@rambler.ru

Аннотация. Большинство современных исследований в области стеганографии либо сосредоточены на концепции скрытой информации, встроенной в мультимедийные контейнеры (изображения, аудио- и видеофайлы) разных форматов, либо направлены на использование телекоммуникационных сетей (сетевая стеганография). Предлагаемая работа является начальной стадией исследования, посвященного текстовой стеганографии, и рассматривает модификацию цветовых параметров текста и извлечение информации из контейнера без использования специального программного обеспечения.

Ключевые слова: текстовая стеганография, модификация цветовых параметров текста.

Для решения задачи защиты информации от несанкционированного доступа используется множество способов и методов. Два основных способа — криптография и стеганография — имеют свои преимущества и недостатки.

Стеганография применяется тогда, когда необходимо не только скрыть смысл сообщения, но и спрятать сам факт передачи. Основной принцип стеганографии — замаскировать конфиденциальное сообщение, используя открытую, всем доступную информацию. Стеганография используется очень давно, но с появлением и развитием компьютерной техники вышла на новый, более качественный уровень. Различают несколько способов стеганографии, применяющих вычислительную технику и сети: компьютерная стеганография, использующая некоторые особенности файловых систем, неиспользуемых областей дисков, форматов файлов и т.д.; цифровая стеганография, использующая мультимедиаконтент. В последнем случае используется мультимедийный контейнер, в который вкладывается скрываемое сообщение. Существует также сетевая стеганография, использующая особенности протоколов передачи данных в различных сетях [7].

Разновидность компьютерной стеганографии, которая использует текстовый контейнер, называется текстовой стеганографией (text steganography) [5, 10]. Для сокрытия информации в текстовых контейнерах используются разнообразные методы: цветовое кодирование на основе перестановки и систем счисления [4], использование цвета шрифта невидимых символов в документах Microsoft Word [2], модель нормализации интервалов (CSNTSteg) [6], OCR-модели для встраивания и извлечения информации [1], метод на основе изменения междустрочного расстояния неотображаемых символов строк [9], метод модификации контуров символов текста [14] и др.

Наше внимание привлек метод текстовой стеганографии на основе модификации цветовых координат символов. Суть метода заключается в том, что цвет символа в текстовом процессоре MS Word представлен в цветовой модели RGB, то есть цвет каждого символа

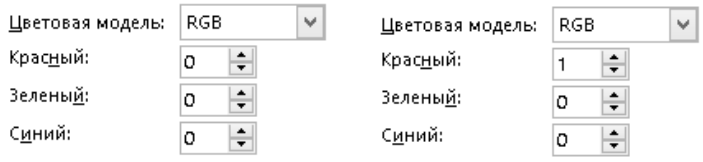

Рис. 1. Параметры шрифта до и после изменения цветового спектра в окне *Цвета* MS Word 2016

формируется из палитры трех цветов. Незначительное изменение одного, двух или трех цветов символа не воспринимается человеческим глазом. Используя данную физиологическую особенность, можно незаметно производить встраивание информации. Для использования данного метода И.М. Ажмухамедовым и Н.А. Колесовой было создано программное обеспечение (далее — ПО) [8]. Для извлечения информации также используется созданное этими разработчиками ПО. Таким образом, для реализации данного метода нужно две специализированные программы. Это накладывает определенные ограничения на использование перспективного способа. И если внедрение сообщения в контейнер еще допустимо, то извлечение при помощи сторонней программы крайне нежелательно. Весьма вероятно, что получатель сообщения не сможет использовать специализированую программу из-за тех или иных ограничений (чужой компьютер, групповые политики и т.д.).

Нами предлагается метод текстовой стеганографии, особенностью которого является отсутствие использования специального программного продукта. Он заключается в том, что в параметры шрифта текущего текста вручную вносится изменение цвета букв, находящееся ниже порога чувствительности глаза нормального человека. В результате в RGB-модели буквы спектр незначительно меняется (рис. 1). Таким образом, из «искаженных» букв составляется сообщение, которое необходимо спрятать, и исходный текст превращается в стегоконтейнер.

Обнаружить данную маскировку сложно, т.к. проверить каждую букву вручную представляется трудоемкой задачей. Используя такую технологию, можно замаскировать свое сообщение, заранее обговорив с адресатом необходимый цвет. При этом размер контейнера меняется, но не критично. Для эксперимента начальный фактический размер контейнера составил 13837 байт, после внедрения сообщения (одно слово) его объем увеличился до 13993 байт, или на 1,12%. Объем внедрения составил 0,63% (6 букв) от объема текста в 788 знаков. При этом размер файла на диске остался прежним — 16384 байта.

Таким образом, мы скрыли наше сообщение в нейтральном тексте. Теперь перед нами встает не менее важная задача: извлечь необходимую информацию из контейнера, используя при этом только стандартные и повсеместно используемые программы, желательно бесплатные. Предлагается следующий метод:

1. Запаковать файл условно-бесплатной программой WinRAR.
2. Открыть архив.
3. Просмотреть файл средствами архиватора (Alt+V) или выбрав в контекстном меню команду *Просмотреть файл*.
4. В открывшихся окнах найти файл *document.xml* (данное имя файлу присваивает программа-архиватор).
5. Открыть его.
6. В открывшемся в режиме HTML файле найти буквы с измененной цветовой моделью (на рис. 2 область выделена черным прямоугольником): они обрамлены тэгами *<w: t>* и одиночны.
7. Собрать сообщение из разрозненных букв.

Оценим показатели визуального искажения, основанные на анализе пиксельной структуры контейнера нашей стеганографической системы. Количественное и качественное оценивание стойкости системы защиты к внешним воздействиям представляет собой достаточно сложную задачу, которая обычно на практике реализуется методами системного анализа, математического моделирования или экспериментального исследования [11].

Используемые в исследовании тексты были набраны шрифтом Times New Roman, размер 12 и 14 пт (2,8 и 3,3 мм). Количество пикселей для 12 пт — 121 пиксель на букву, для 14 пт — 169 пикселей на букву. Данные для других шрифтов и размеров вычисляются аналогично.

Так как наша система относится к компьютерной и цифровой стеганографии, используем метод объективного оценивания качества. Большое распространение получили корреляционные показатели качества изображения, например коэффициент корреляции Пирсона (1):

$$k = \frac{\sum_c \sum_r (A(c,r) - A_m) \cdot (B(c,r) - B_m)}{\sqrt{\sum_c \sum_r ((A(c,r) - A_m)^2) \cdot (\sum_c \sum_r (B(c,r) - B_m)^2)}} \quad (1)$$

Рис. 2. Просмотр файла-контейнера document.xml в режиме HTML

где c, r — координаты пикселя изображения; $A(c, r)$, $B(c, r)$ — исходное и искаженное изображение; A_m, B_m — среднее арифметическое пикселей исходного и искаженного изображения. Значения коэффициента корреляции определены в диапазоне от –1 до 1. В случае полной идентичности изображений он принимает значение, равное 1. Для вычисления корреляции Пирсона в Matlab предусмотрена функция *corr2*. При вычислении получили значение, равное 1. Значения c и r меняются от 1 до 11 (для 12 пт) или от 1 до 13 (14 пт) пикселей.

В нашем исследовании использовалась операционная система (далее — ОС) Windows 10, браузер Internet Edge, текстовый процессор MS Word 2016 и условно-бесплатная программа — архиватор Win RAR6.11. Для апробации был взят текстовый документ, содержащий 127 слов, 914 (788) знаков с пробелами (без пробелов). За рамками исследования остались механизмы взаимодействия архиватора и файлов архива. В частности, детально не рассматривалась команда *Просмотреть файл* (вызов Alt+V) архиватора.

Данный метод находится на границе компьютерной и цифровой стеганографии. К компьютерной стеганографии его относит контейнер — текстовый файл, к цифровой — замена цвета объекта. Основным достоинством нашего метода можно назвать возможность обнаружить сообщение, не прибегая к специальным программам. Использование широко распространенного и/или бесплатного ПО допускает применение нашего метода в тех случаях, когда приходится пользоваться «чужим» оборудованием.

В ходе нашего исследования остались не освещены вопросы общего использования данного метода:

1. Использование данного метода для различных ОС и других версий программы WinRAR.
2. Использование большого объема контейнера.
3. Использование большого объема информации.
4. Подбор размера исходного файла и контейнера с целью сокрытия использования стеганографии.
5. Создание программ для автоматизированного ввода и вывода текста в контейнер. Нашей целью было использовать только самые распространенные программные средства, поэтому в дальнейшем мы намерены использовать язык программирования VBA, так как он встроен в пакет MS Office.

Предлагаемая технология является начальной стадией нашего исследования, и обозначенные вопросы станут частью дальнейших разработок.

ЛИТЕРАТУРА

1. Ding K., Hu T., Niu W., Liu X., He J., Yin M., Zhang X.A. Novel Steganography Method for Character–Level Text Image Based on Adversarial Attacks. Sensors 2022, 22, 6497.
2. Khairullah M.A Novel Text Steganography System Using Font Color of the Invisible Characters in Microsoft Word. URL: www.researchgate.net/publication/232615004_A_Novel_Text_Steganography_System_Using_Font_Color_of_the_Invisible_Characters_in_Microsoft_Word_Documents.
3. Liu Q., Sung A.H. Feature Mining and Neuro-Fuzzy Inference System for Steganalysis of LSB Matching Stegangoraphy in Grayscale Images https://www.researchgate.net/publication/220814327_Feature_Mining_and_Neuro-Fuzzy_Inference_System_for_Steganalysis_of_LSB_Matching_Stegangoraphy_in_Grayscale_Images.
4. Sadié J.K., Metcheka L.M., Ndoundam R. Two high capacity text steganography schemes based on color coding. URL: https://arxiv.org/abs/2004.00948.
5. Shniperov A.N. A text steganography method based on Markov chains // Automatic Control and Computer Sciences.— 2016.— T. 50 (№ 8).— С. 802–808.
6. Thabit R., Udzir N.I., Yasin S.M., Asmawi A., Gutub A.A.— A. CSNTSteg: Color Spacing Normalization Text Steganography Model to Improve Capacity and Invisibility of Hidden Data. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9795032.
7. Абазина Е.С., Ерунов А.А. Сравнительный анализ и классификация методов цифровой и компьютерной стеганографии, и перспективные направления ее развития // Труды Военно–космической академии имени А.Ф. Можайского.— 2016.— № 655.— С. 5–16.
8. Ажмухамедов И.М., Колесова Н.А. Защита от передачи стегосообщений в графических файлах // Инфокоммуникационные технологии.— Том 10.— № 3.— 2012. С. 93–97.
9. Блинова Е.А. Стеганографический метод на основе изменения междустрочного расстояния неотображаемых символов строк электронного текстового документа // Труды БГТУ. № 6. Физико-математические науки и информатика.— 2016.— № 6(188).— С. 166–169.
10. Герлинг Е.Ю. Обзор современного программного обеспечения, использующего методы стеганографии // Экономика и качество систем связи.— 2019.— № 3(13).— С. 51–58
11. Конахович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография. Теория и практика. «МК-Пресс», Киев, 2006.— 288 с.
12. Красов А.В. Метод обнаружения сетевой стеганографии на основе машинного обучения // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки.— 2022.— № 3.— С. 100–108.— DOI 10.37882/2223–2966.2022.03.17
13. Михаэлис В.В. Защита беспроводных сетей // Информационные технологии и проблемы математического моделирования сложных систем.— 2015.— № 14.— С. 4–10.
14. Нистюк О.А., Урбанович П.П. Метод и математическая модель стеганографического преобразования информации на основе модификации контура символов текста-контейнера // Труды БГТУ. Серия 3: Физико-математические науки и информатика.— 2022.— № 2(260).— С. 92–98.

Журнал «Современная наука: актуальные проблемы теории и практики»