Метод Текстовой Стеганографии на Основе Модификации Цветовых Координат Символов

Надежда Урбанович, Владимир Пласковицкий
Факультет издательского дела и полиграфии
Белорусский государственный технологический университет, Минск, Беларусь
nadya_ur@rambler.ru

Аннотация

В статье рассматриваются особенности предлагаемого авторами метода компьютерной текстовой стеганографии на основе изменения цветовых параметров символов текста. Метод может применяться, в том числе, для защиты прав интеллектуальной собственности. Описывается программное средство для изучения и анализа эффективности метода.

Ключевые слова: *текстовая стеганография, цвет символа, защита авторского права.*

1. ВВЕДЕНИЕ

Известны два основных направления решения задачи защиты информации от несанкционированного использования: криптография и стеганография.

Целью криптографии является скрытие содержимого сообщений за счет их шифрования. В отличие от этого, в стеганографии скрывается сам факт существования тайного сообщения. Дальнейшее изложение касается компьютерной стеганографии, предусматривающей размещение информации в текстовых документах.

Слово «стеганография» имеет греческие корни и буквально означает «тайнопись». С помощью различных методов можно тайно (с различным уровнем тайности) передавать сообщение, «осажденное» в документ (файл).

Кратко охарактеризуем основные понятия, относящиеся к предметной области: сообщение, контейнер и ключ. Термин «контейнер» употребляется в отечественной литературе большинством авторов, поскольку является дословным переводом устоявшегося английского термина «container», обозначающего несекретную информацию, которую используют для сокрытия сообщений. По сути же, контейнер в стеганографической системе является ни чем иным как носителем сокрытой информации, поэтому вполне возможно использование и такого термина. В некоторых источниках термин контейнер также заменяют названием «стего», который также является производным от английского сокращения «stego» (полное название «stegano»).

Контейнером (носителем) называют несекретные данные, которые используют для сокрытия сообщений.

В компьютерной стеганографии в качестве контейнеров могут быть использованы различные оцифрованные данные: не только текстовые электронные документы, но и растровые графические изображения, звук, видео и др.

Сообщением (стегосообщением) называют секретную информацию, наличие которой в контейнере необходимо скрыть.

Ключом называют секретную информацию, известную только законному пользователю, которая и определяет конкретный вид алгоритма сокрытия [1].

Кроме скрытой передачи сообщений, стеганографические методы являются одним из самых перспективных инструментов для аутентификации и маркировки авторской продукции с целью защиты авторских прав на цифровые объекты от пиратского копирования. В качестве стегосообщения можно использовать данные об авторе, дату и место создания произведения, номера документов, подтверждающих авторство, дату приоритета и т.п. Такие специальные сведения могут рассматриваться в качестве доказательств при рассмотрении споров об авторстве или для доказательства нелегального копирования.

Многие существующие методы текстовой стеганографии [2] недостаточно эффективно скрывают сообщения, факт наличия секретной информации (ключ) является очевидным либо почти очевидным. Этот вывод сделан авторами данной статьи на основе сравнительного анализа известных методов текстовой стеганографии с помощью специально разработанного программного средства. Таким образом, актуальной является задача разработки новых методов, повышающих устойчивость к атакам, т. е. снижающим вероятность извлечения сообщения из контейнера.

2. СУЩНОСТЬ МЕТОДА МОДИФИКАЦИИ ЦВЕТОВЫХ ПАРАМЕТРОВ СИМВОЛА

Цвет символа в текстовом процессоре Microsoft Word представлен в цветовой модели RGB. Незначительное изменение цвета символа не воспринимается человеческим глазом. Используя данную физиологическую особенность, можно незаметно производить встраивание информации.

Подобный подход используется в случаях, когда контейнером является изображение (графический файл). Известный метод имеет название Least Significant Bit (LSB). При его реализации встраивание производится в последние 1–2 бита цвета пикселя изображения. При адаптации вышеупомянутого алгоритма к тексту встраивание можно производить в последние 3–5 битов цвета символа. Увеличение числа используемых бит цвета в тексте, по сравнению с графикой, происходит из-за того, что изображение, как правило, содержит градации и переходы от одного цвета к другому. Текст — монотонен и выполняется в большинстве случаев одним цветом, поэтому становится возможным увеличение используемого для встраивания цветового диапазона.

Например, необходимо внедрить секретное сообщение «101» в текст-контейнер «А», используя текстовый процессор MS Office Word 2007. В данном процессоре цвет символов представлен в системе RGB (цвет представлен тремя стандартными цветами: red, green, blue) с 8 битами на канал (каждый из 3 стандартных цветов представлен 8 битами — числом от 0 до 255). Цвет текста-контейнера «А» — черный: данный цвет представлен в MS Office Word 2007 как (00000000, 00000000, 00000000). Встраивание

секретного сообщения будем производить в младшие биты цвета символов. В итоге будет получен цвет: (00000001, 00000000, 00000001).

Результат внедрения сообщения «101» в текст-контейнер (символ «А»), при использовании текстового процессора MS Office Word 2007, показан на рисунке 1 (показано увеличенное графическое изображение). Как видно, визуально оба символа одинаковы.

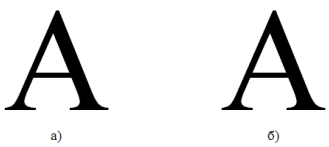

Рис. 1: Применение метода изменения цвета символов

а) стандартное графическое изображение символа; б) модифицированный символ (со встроенным секретным сообщением «101»)

3. ОПИСАНИЕ ПРОГРАММНОГО СРЕДСТВА НА ОСНОВЕ ИЗМЕНЕНИЯ ЦВЕТОВЫХ КООРДИНАТ СИМВОЛОВ ТЕКСТА

Описанный выше метод реализован в программном продукте, характеризующемся следующими техническими параметрами.

Язык разработки: C# (Framework 2.0 + библиотеки Microsoft.Vbe.Interop.dll, Microsoft.Office.Interop.Word.dll, office.dll).

Требования к ОС: Windows + Framework 2.0 (XP, Vista, Seven).

Поддерживаемые форматы документов для скрытия: любые, который можно открыть с помощью Office Word и способные хранить цвет символов: Word 93-2010 (*.doc, *.docx), *.rtf (межплатформенный формат хранения размеченных текстовых документов) , *.odt *(открытый формат документов для офисных приложений*). Поддержка данных форматов должна быть доступна из установленного на компьютере пользователя приложения MS Word.

Общий принцип работы приложения проиллюстрирован на диаграмме деятельности (рис. 2).

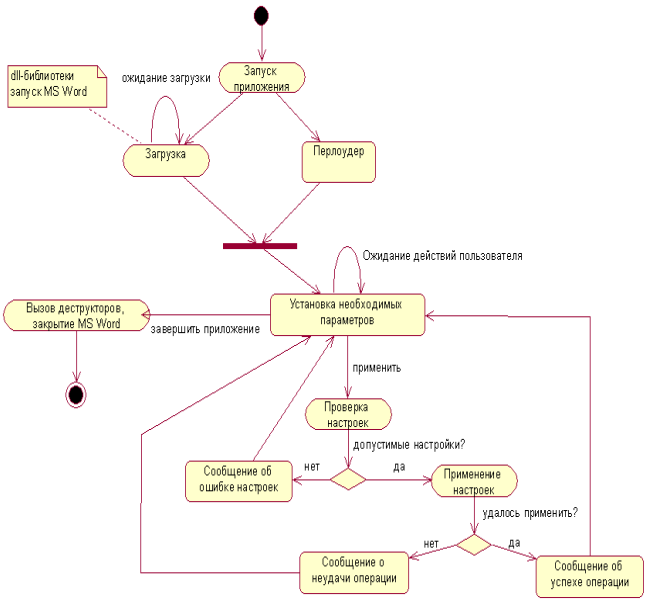

Рис. 2: Диаграмма деятельности программного средства.

Главное окно программного средства имеет вид, представленный на рис. 3.

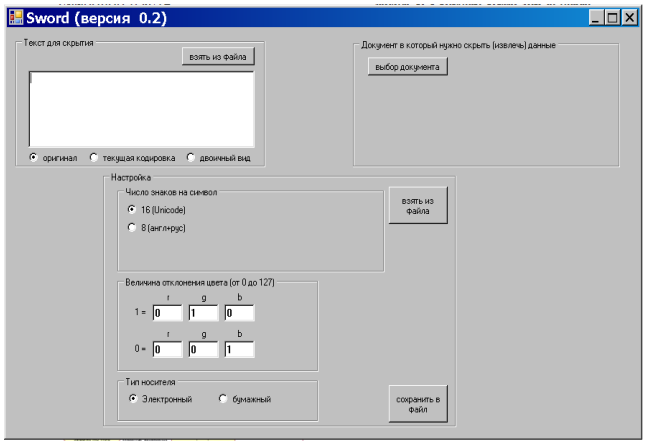

Рис. 3: Общий вид программы

Установка необходимых параметров осуществляется с помощью трех блоков, в которых задается: стегосообщение (в случае извлечения заполнять не надо), контейнер и настройки, на основе которых происходит скрытие/извлечение сообщения (ключ).

Последовательность скрытия текста:

- перевод текста в двоичный вид (на основе выбранной кодировки);
- проверка того, достаточно ли в документе-контейнере символов для скрытия стегосообщения. Если последнее состоит из K символов и при этом выбрана кодировка, в которой один символ представлен Q знаками, то в документе должно быть не меньше K*Q*2 знаков. Умножение на 2 предусматривает то, что каждому скрываемому символу нужна пара, по которой будет определяться отклонение его цвета;
- если проверка прошла успешно, составляется список символов, в которых будет производиться скрытие. Эта процедура необходима для того, чтобы модифицированные символы располагались через случайный промежуток. Поэтому нужно убедиться, что алгоритм, задающий этот шаг, укладывается в границы документа. Правило расстановки следующее: анализируется максимальное расстояние, на котором могут располагаться скрываемые символы. Например, это расстояние равно S. Для каждого символа генерируется шаг в диапазоне от 2 до (2*S-2), т. е. иногда шаг будет превышать это значение, иногда наоборот - будет меньше. Затем происходит составление списка. Если в момент составления происходит отклонение в большую от S сторону, то значение S уменьшается на единицу и процедура повторяется. В худшем случае S будет равняться 2, т. е. скрытые символы будут идти с плотностью через один;
- если список составлен успешно, происходит осаждение символов стегосообщения. Время встраивания пропорционально числу скрываемых

символов и практически не зависит от объема документа-контейнера;

- чтобы просмотреть позиции, в которых скрыты символы, необходимо воспользоваться кнопкой «Показать», а затем отметить эти символы (кнопка «Отметить»). После этого соответствующие символы будут помечены отдельным (например, синим маркером; рис. 4);

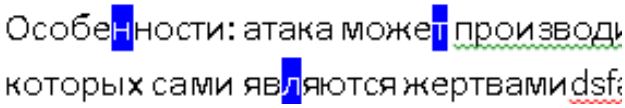

Рис. 4: Вид документа с выделением модифицированных символов

- перед сохранением либо изменением документа необходимо очистить эти метки (кнопка «Очистить»);
- параметры, которые используются для скрытия текста, можно сохранять в файлы для последующего считывания. Эти файлы являются обычными текстовыми документами, в которых хранятся настройки, структурированные специальным способом. Для того чтобы отличить файлы с такими настройками от других текстовых документов, первые сохраняются с расширением *.sword.

Последовательность извлечения текста:

- выбор документа-контейнера, содержащего стегосообщение;
- настройка параметров (ключа), на основе которых он был скрыт (если были сохранены, можно загрузить из файла);
- запуск операции извлечения (кнопка «Извлечь»);
- после извлечения текст будет помещен в соответствующее поле. Время, затраченное на извлечение, пропорционально числу символов в документе и практически не зависит о числа скрытых символов. Скорость извлечения составляет приблизительно 50 знаков в секунду. Во время дешифровки происходит уведомление пользователя о текущем состоянии выполнения с помощью специального текстового поля.

Стегосообщение можно вводить с клавиатуры, вставлять из буфера, открывать из файла. При выборе необходимого документа предоставляется три фильтра, а также возможность выбора любого файла (см. рис. 5).

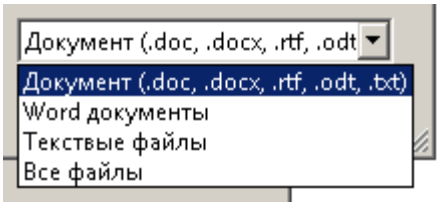

Рис. 5: Возможные типы документов

Во время работы с текстом можно просматривать его в двоичной системе (рис.6). Редактировать текст в этом режиме нельзя.

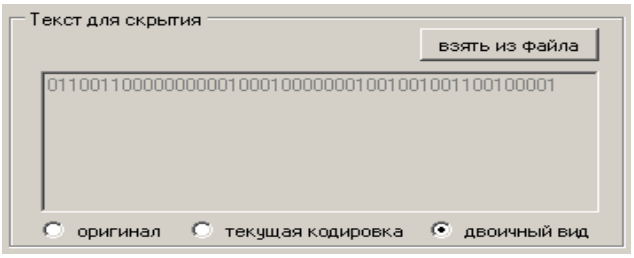

Рис. 6: Стегосообщение, представленной в двоичном виде

По умолчанию выбранный документ не отображается на экране, чтобы просмотреть его, необходимо нажать на кнопку «Показать». После открытия документа с его содержимым можно производить любые действия. При внедрении исследуется именно текущее содержание документа, а не то, которое было на момент открытия.

Если документ будет закрыт пользователем не из приложения MS Word, то он автоматически закроется и в программе. Он также будет закрыт, если в программе выбрать другой документ, поэтому при необходимости его надо предварительно сохранить.

Алгоритм модификации цветовых координат символов текста состоит в следующем. В настройках задается отклонение от основного цвета для символов «0» и «1» в системе RGB. В процессе скрытия берутся два идущих подряд символа: первый — образец, второй — тот, в котором будет скрыт текст. Если в настройках запрещено прятать текст в пробельных участках (тип носителя — «бумажный»), тогда сдвиг осуществляется до тех пор, пока оба символа будут не пробельными.

Цвет символа, в котором будет производиться скрытие, формируется исходя из цвета символа-образца и заданного в настройках смещения. По умолчанию это смещение добавляется к основному цвету. Если при этом значение выходит за допустимый диапазон — отнимается. Например, если цвет образца задан как (0,200,100), а смещение задано как (100,60,50), то результирующим будет цветовая координата (100,140,150). Чтобы избежать ситуации, когда значение может выйти за оба диапазона, в полях для отклонения не рекомендуется задавать значения больше 128.

4. ЗАКЛЮЧЕНИЕ

Описанный метод и программное средство используются в настоящее время для изучения возможности размещения в текстовых документах информации, предусматривающей защиту прав интеллектуальной собственности. В ходе достаточно объемных экспериментов установлено, что модификация до 4-х младших символов цветовой координата каждого канала (RGB) в 100% случаев остается незамеченной пользователем, которому не известен факт осаждения в документе невидимой информации.

5. БЛАГОДАРНОСТИ

Работа была выполнена при частичной поддержке гранта ГБ 11-025.

6. ССЫЛКИ

[1] Конахович Г. Ф., Пузыренко А.Ю. *Компьютерная*

стеганография. Теория и практика. – К.: «МК-Пресс», 2006. – 288 с.

[2] Ярмолик В. Н., Портянко С. С., Ярмолик С.В. *Криптография, стеганография и охрана авторского права – Минск: Изд.центр БГУ, 2007. – 240с.*

Об авторах

Надежда Урбанович — студентка БГТУ. Ее адрес: nadya_ur@rambler.ru.
Владимир Пласковицкий — студент БГТУ. Его адрес: vaplas20@gmail.com.