ТЕКСТОВАЯ СТЕГАНОГРАФИЯ В HTML: РЕАЛИЗАЦИЯ СКРЫТЫХ КАНАЛОВ ПЕРЕДАЧИ ДАННЫХ

С.С. Барильник, И.В. Минин, О.В. Минин, Ю.В. Щетинин
Новосибирский государственный технический университет
г. Новосибирск

Информатизация ведет к созданию единого мирового информационного пространства, в рамках которого ведется обмен информацией, в том числе, между различными субъектами. Однако развитие высоких технологий информационного обмена делает их потенциально уязвимой от электронного терроризма и шантажа. Одним из актуальных направлений деятельности является оценка реальных стелсографических угроз, разработка мер по стелсографической защите информации [1].

HTML (Hyper Text Meta Language) – это теговый язык разметки документов. Любой документ на языке HTML представляет собой набор элементов, причём начало и конец каждого элемента обозначается специальными пометками – тегами. Элементы могут быть пустыми, то есть не содержащими никакого текста и других данных (например, тег перевода строки
). В этом случае обычно не указывается закрывающий тег. Кроме того, элементы могут иметь атрибуты, определяющие какие-либо их свойства (например, размер шрифта для элемента font). Атрибуты указываются в открывающем теге. Вот пример фрагмента HTML-документа [2]:

Здесь элемент содержит атрибут href.

Сегодня практически все Web-страницы в internet являются отображением HTML кодов.

В одной из своих статей были рассмотрены методы защиты авторского права на web-страницы при помощи алгоритмов текстовой стеганографии [3]. В этой работе будут предложены методы невидимой передачи информации через web-сайты по средствам текстовой стеганографии в HTML.

Для раскрытия цели данной работы, рассмотрим небольшой пример. Пусть вам необходимо сообщить секретный ключ к вашему банковскому счету родному человеку, но он находится в другой стране или даже на другом континенте. Естественно, по телефону сообщать такую информацию опасно, тем более страшновато передавать ключ через интернет (например, по электронной почте). Можно попробовать зашифровать этот секретный код, но в таком случае, получение вашего ключа посторонним лицам – это лишь вопрос времени. Так как же быть, необходимы такие каналы передачи информации, которые невозможно было бы прослушать классическими методами.

Предположим, что во всемирной паутине вы имеете свою страничку, где отображена некоторая информация, доступная всем пользователям сети. Но вы предпочитаете, чтобы некоторые «избранные» видели немного больше, чем остальные. Для этого можно воспользоваться алгоритмами текстовой стеганографии, адаптированной для HTML [3, 4].

В предложенной статье рассмотрены два алгоритма:

- биты скрываемой информации представляются в виде непечатаемых символов. Такими символами являются «Пробел» и «Горизонтальная табуляция». Таким образом, можно представить биты в виде символов: «1» – «Пробел», «0» – «Горизонтальная табуляция». Каждый байт скрываемой информации преобразуется в последовательность этих символов, где каждому символу соответствует бит скрываемого байта. Например, скрываемый байт – 0x43 = 0100 0011 => « | | | | | | | |», где | | – «Пробел», | | – «Горизонтальная табуляция». Далее полученная последовательность помещается в конец строки и становится «невидима». По такому принципу можно скрыть один байт информации в одной строке [3].
- в Windows для перевода строки используется два символа: 0x0D, 0x0A. В современных Unix операционных системах для этого достаточно одного символа: 0x0A. Большинство текстовых редакторов понимают и правильно отображают оба формата перевода строк. Пользуясь этой особенностью, можно прятать биты скрываемой информации: «0» – 0x0A, «1» – 0x0D 0x0A, т.е. наличие 0x0D является «1». По такому принципу можно скрыть один бит информации в одной строке [3].

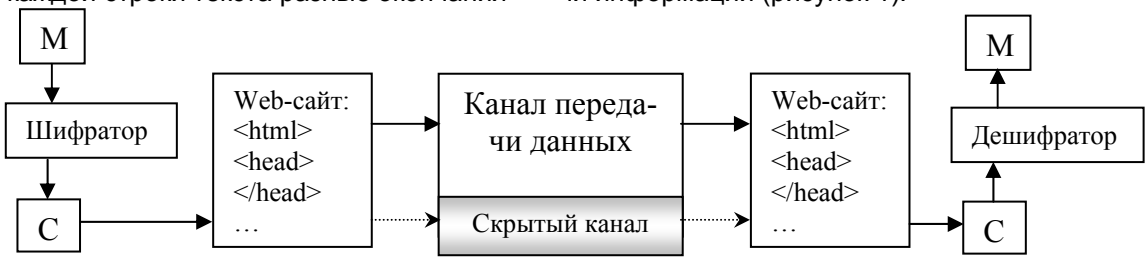
В нашем случае контейнером является HTML файл. Для увеличения скрытности ин-

формации будем вставлять полученные последовательности не в конец каждой строки, а только в конце строк, заканчивающихся на тэг (<html>, </titled>, </body>,
, <td> и т.д.). Это позволяет не отображать скрытых пробелов и горизонтальных табуляторов на странице при работе с первым алгоритмом.

Второй алгоритм тоже можно использовать для скрытной передачи данных. Если злоумышленник будет знать, что данные передаются, и догадается, что они скрываются таким образом (наличие 0x0D в конце строки, так как в конце разных строк будут разные окончания), то, с целью маскировки спрятанных бит информации от их визуального анализа в Hex – редакторе, можно записать в конец каждой строки текста разные окончания (случайным образом), а считывать только необходимые окончания (в данном случаи, только строки, заканчивающиеся на тэг). Поэтому прочитать такие биты, не зная маски, невозможно.

В статье [3] также приведен сравнительный анализ предложенных алгоритмов, который показал, что второй алгоритм обладает большей степенью защищенности и, следовательно, создает более скрытый канал, но его пропускная способность гораздо меньше первого.

Защищенность передаваемой информации также можно повысить путем шифрации передаваемого сообщения. Таким образом, получаем следующую схему скрытой передачи информации (рисунок 1):



M – передаваемое сообщение
C – зашифрованное передаваемое сообщение

Рисунок 1 – Схема скрытой передачи информации

Таким образом, в статье были рассмотрены способы реализации скрытых каналов передачи информации при помощи алгоритмов текстовой стеганографии адаптированных под HTML коды.

Следует отметить, что описанный алгоритм и его практическая реализация внедрены на официальном сайте фирмы, занимающейся разработкой программного обеспечения ООО «Графические программные системы», в качестве защиты авторского права на проект «МирКибер» в июне текущего года.

Предложенные в данной работе алгоритмы хорошо адаптированы под современные языки высокого уровня, такие, как Java, C#, php и т.д. В перспективе планируется дальнейшее исследование алгоритмов текстовой стеганографии и применение их в HTML, а также в других стандартах хранения текстовой информации (XHTML, XML, PDF и т.д.).

Для демонстрации работы алгоритма предложена программа, реализованная на Java, которая позволяет увидеть в полном объеме возможности алгоритма. Java является объектно – ориентированным кросс – платформенным языком программирования, что позволяет сделать программу модульной и работоспособной в любой операционной системе или даже устройстве, имеющем виртуальную Java машину.

Список литературы

1. Голубев Е.А. Стелсографические угрозы / Е.А. Голубев // Материалы 2 межведомственной конференции «Научно-техническое и информационное обеспечение деятельности спецслужб», Москва, 1998. – С. 58-60.
2. Соколов С.А. HTML и CSS в примерах, типовых решениях и задачах. Профессиональная работа / С.А. Соколов – М.: Вильямс, 2007. – С. 416.
3. Барильник С.С. Адаптация алгоритмов текстовой стеганографии для HTML / С.С. Барильник, И.В. Минин, О.В. Минин// 8 международная сибирская школа-семинар по электронным приборам и материалам EDM'2007, 2007, Novosibirsk, NSTU, p. 225 – 228.
4. Минин И.В. Стелсографическая защита интеллектуальной собственности на документы в WWW. И.В. Минин, О.В. Минин, Н.Е. Герасимов // Восьмой международной симпозиум ТЕХНОМАТ 2007 **«***Материалы, Методы и Технологии», Болгария,* 28 май – 1 июнь.