УДК 004.056

Н.В. Ржеутская, асп. (БГТУ, г. Минск)

СОВРЕМЕННЫЕ МЕТОДЫ ТЕКСТОВОЙ СТЕГАНОГРАФИИ И ОЦЕНКА ИХ СТЕГАНОСТОЙКОСТИ

Стеганография – это область знаний, которая занимается вопросами скрытой передачи информации. В отличие от криптографии, скрытым является сам факт передачи информации. Самым эффективным является использование стеганографических методов совместно с криптографическими. Основой стеганографических методов является то, что скрываемое сообщение помещается в любой объект, передаваемый адресату в виде открытого текста, в котором наличие скрытой информации остается незаметным для внимания злоумышленника.

Работа с текстовыми документами подразумевает, что содержимое документа должно быть абсолютно точно передано при обратном преобразовании. Поэтому в отличие от методов преобразования изображений и звуков текстовая информация должна основываться на других методах.

Методы текстовой стеганографии принято подразделять на синтаксические, которые не затрагивают семантику текстового сообщения, и лингвистические методы, которые основаны на эквивалентной трансформации текстовых файлов, сохраняющей смысловое содержание текста и его семантику.

Рассмотрим более подробно синтаксические методы встраивания скрытой информации в текстовые контейнеры. В текстовых контейнерах скрытая информация чаще всего кодируется путем изменения количества пробелов, использования невидимых символов, путем изменения межстрочных интервалов, регистра букв, табуляций и т.д.

Стеганографический метод «Line-shift coding» строится на изменении расстояния между строками электронного текста документа-контейнера.

Стеганографический метод «Word-shift coding» основывается на изменении расстояния между словами строки электронного текста. Частным случаем данного метода является метод изменения количества пробелов.

Стеганографический метод «feature coding» основывается на внесении специфических изменений в шрифты отдельных букв.

Достоинства синтаксических методов: данные методы легко встраиваются в любой текст независимо от его содержания, назначения и языка. Данные системы легко разрабатываются и выполняются автоматически.

Недостатки синтаксических методов: вышеперечисленные методы легко взламываются, и секретная информация может легко устраняться путем простейших атак. Еще один большой недостаток – этими методами нельзя передавать большое количество скрытой информации.

Теперь рассмотрим лингвистические методы встраивания скрытой информации в текстовые контейнеры.

Лингвистические методы – это системы, которые основываются на лексической структуре самого текста. При использовании лингвистической стеганографии передача скрытого сообщения производится посредством внедрения этого сообщения в некоторый текст, который сам по себе не содержит полезной для получателя информации и для посторонних глаз выглядит безобидным.

Стеганографический метод переменной длины основывается на том, что в одно слово обычно кодируется два бита информации из стегосообщения.

Достоинством данного метода является его простота.

В качестве недостатка можно выделить следующее – слова, набираемые пользователем, должны иметь длину, которую укажет программа.

Стеганографический метод первой буквы подразумевает то, что в первую букву каждого слова кодируется шифр. Одну и ту же комбинацию символов можно закодировать несколькими буквами.

Достоинством данного метода является то, что данный метод может передавать еще больше скрытой информации в одном слове. Такой метод дает автору сообщения больше свободы действия при создании стегосообщения и текст не будет нелепым.

Недостатком данного метода является его слабая производительность, а также передача небольших объемов информации и низкая степень скрытности.

Метод, основывающийся на системе синонимов языка. Принцип работы данного метода прост. Часто в тексте одно слово может быть заменено другим, которое является синонимом исходного слова.

Достоинство стеганографического метода, основанного на замене синонимов, позволяет сохранить синтаксическую структуру предложения и его смысловую нагрузку. Такую замену слов достаточно легко проделать человеку. Этот метод нельзя реализовать простым машинным алгоритмом.

В качестве недостатка можно выделить следующее – в русском языке существует достаточно большое количество пар, состоящих из слова и его синонима. Использование всех таких пар для целей стега-

нографического сокрытия информации, когда слову ставится в соответствие двоичный «0», а его синониму «1» очередного бита скрываемого сообщения, часто приводит к значительным искажениям смысловой нагрузки скрывающего текста.

Как следствие, из-за неправильного употребления синонимов текст, содержащий скрытую информацию, становится легко идентифицируемым, и, в свою очередь, позволяет злоумышленнику установить наличие скрытого сообщения.

Мимикрия. Мимикрия генерирует осмысленный текст, используя синтаксис, описанный в Context Free Grammar (CFG), и встраивает информацию, выбирая из CFG определенные фразы и слова. CFG – это один из способов описания языка, который состоит из статических слов, фраз, узлов, мест, где может быть принято решение, какое слово или фразу дальше вставлять в текст. Мимикрия создает бинарное дерево, которое основано на возможностях CFG, и составляет текст, выбирая те из листьев дерева, которые кодируют нужный бит.

Устойчивость методов, генерирующих искусственный стего-текст, обеспечивается заданными правилами грамматики.

В качестве недостатка можно выделить слабую производительность метода, передачу небольших объемов информации. Отсутствие грамматических и орфографических ошибок в предложениях делает затруднительным поиск отличий искусственного текста от естественного. Анализ осмысленности текста можно производить только с участием человека, что не всегда возможно из-за огромного объема анализируемой информации.

Основной проблемой при использовании всех выше описанных методов текстовой стеганографии является потеря либо частичная деформация скрытого сообщения при изменении или редактировании основного текстового контейнера. Данную проблему можно решить, объединив использование различных методов стеганографии, а также криптографии и помехоустойчивого кодирования для обеспечения эффективного решения сокрытия информации.