УДК 004.056.55

И.В. Соколенко, Н.Е. Губенко
Донецкий национальный технический университет, г. Донецк
кафедра компьютерных систем мониторинга

АНАЛИЗ ОСОБЕННОСТЕЙ МЕТОДА ЛЕКСИЧЕСКОЙ СТЕГАНОГРАФИИ, ОСНОВАННОГО НА ГЕНЕРИРОВАНИИ ТЕКСТА

Аннотация

Соколенко И. В., Губенко Н. Е. Анализ особенностей метода лексической стеганографии, основанного на генерировании текста. *Рассматривается метод лингвистической стеганографии, основанный на генерирование искусственного текста. Анализируются особенности при реализации данного метода, его недостатки и преимущества.*

Ключевые слова: *лексическая стеганография, генератор текстов, мимикрия.*

Введение

Проблема защиты информации от несанкционированного доступа была актуальна во все времена. Однако, если ранее для решения данной проблемы использовали криптографические методы, то в современном мире все популярнее становится стеганография.

Стеганогра́фия (от греч. στεγανός — скрытый + γράφω — пишу; буквально «тайнопись») — это наука о скрытой передаче информации путём сохранения в тайне самого факта передачи [1]. Для этого с помощью специальных алгоритмов в обычное сообщение, которое называют контейнером, встраивают секретное сообщение. Контейнер подбирается так, чтобы его содержание и факт передачи не вызывали никаких подозрений. Сегодня в сети Интернет передается большое число файлов различных типов: цифровые фотографии, видео, текст музыка и др. Следовательно, все эти файлы могут выступать в качестве контейнера.

В данной статье обсуждается метод стеганографии, который использует в качестве контейнера текстовые файлы.

1 Основные понятия

На конференции Information Hiding: First Information Workshop в 1996 году было предложено использовать единую терминологию и обговорены основные термины стеганографии.

Стеганографическая система или стегосистема - совокупность средств и методов, которые используются для формирования скрытого канала передачи информации. Обобщенная модель стегосистемы представлена на рисунке 1.

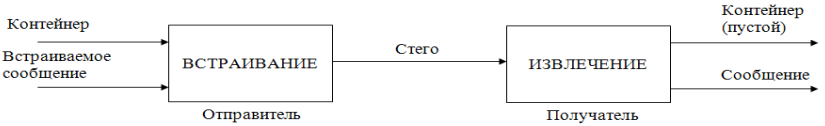

Рисунок 1 – Обобщенная модель стегосистемы

При построении текстовой стегосистемы, как и любой другой, должны учитываться следующие положения:

- противник имеет полное представление о стеганографической системе и деталях ее реализации. Единственной информацией, которая остается неизвестной потенциальному противнику, является ключ, с помощью которого только его держатель может установить факт присутствия и содержание скрытого сообщения;
- если противник каким-то образом узнает о факте существования скрытого сообщения, это не должно позволить ему извлечь подобные сообщения в других данных до тех пор, пока ключ хранится в тайне;
- потенциальный противник должен быть лишен каких-либо технических и иных преимуществ в распознавании или раскрытии содержания тайных сообщений.

В качестве встраиваемого сообщения может быть: текст, изображение и т. п. Контейнер - любая информация, предназначенная для сокрытия тайных сообщений. В случае, если данная информация является текстом, то стегосистему называют текстовой [2].

Текстовая стегосистема, как и любая другая, должна отвечать следующим требованиям:

- Стегосообщение должно быть устойчиво к искажениям, в том числе и злонамеренным. В процессе передачи изображение (звук или другой контейнер) может претерпевать различные трансформации: уменьшаться или увеличиваться, преобразовываться в другой формат и т. д. Кроме того, оно может быть сжато, в том числе и с использованием алгоритмов сжатия с потерей данных.
- Свойства контейнера должны быть так видоизменены, чтобы изменение невозможно было выявить при визуальном контроле. Этим требованием определяется качество сокрытия внедряемого сообщения: заполненный контейнер никоим образом не должен привлечь внимание атакующего.

Зачастую, при преобразовании текста, возникает проблема потери его связности, что, несомненно, приводит к подозрениям со стороны стороннего наблюдателя. Лучше всего с этой проблемой справляется ***лингвистическая стеганография*** – внедрение произвольной информации

в произвольный текст с опорой на свойства языка и лингвистические ресурсы [3]. С развитием информационных технологий появилась возможность генерировать произвольный текст автоматически. Тексты, созданные с помощью генераторов, являются правильными с точки зрения большинства языковых норм.

2 Метод генерации искусственного текста

Одним из распространенных методов передачи скрытой информации, является мимикрия. Мимикрия генерирует осмысленный текст, встраивая информацию, выбирая при этом из базы определенные фразы и слова. Причем база состоит из статических слов, фраз, узлов, мест, где может быть принято решение, какое слово или фразу дальше вставлять в текст. Мимикрия создает бинарное дерево и составляет текст, выбирая те из листьев дерева, которые кодируют нужный бит.

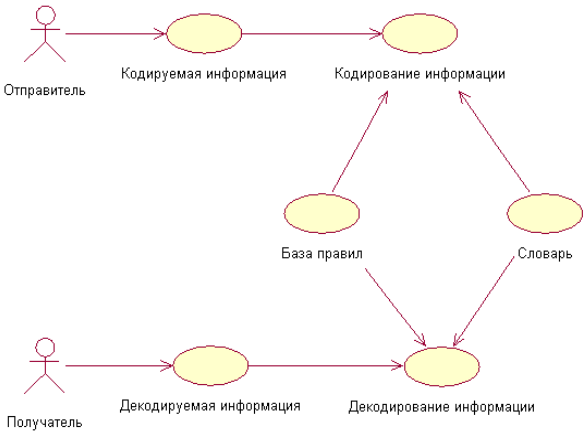

Рисунок 2 – Use Case диаграмма метода

К примеру, необходимо закодировать комбинацию «100». Следует сгенерировать текст согласно схеме:

Подлежащее Сказуемое Дополнение

Имеется база к подлежащему (Николай, Иван); база к сказуемому (Купил, Приобрел) база к дополнению (Машину, Часы).

0	**1**
Николай	Иван
Купил	Приобрел
Машину	Часы

Таким образом, получается предложение: ***Иван купил машину.***

Безусловно, для реализации данного метода необходимо грамотно и четко составить базу слов. В идеале базу необходимо дополнить фразами и т. н.

узлами переходов.

Так, примером реализации данного метода на английском языке является программа spammimic, генерирующая спам. В качестве встраиваемого сообщения было выбрано слово hello (рисунок 3).

Рисунок 3 – Кодирование слова hello в программе spammimic

В результате было выведено сообщение, состоящее из 199 слов (рисунок 4).

Рисунок 4 – Вывод закодированного сообщения

Данный метод обладает как преимуществами, так и недостатками. К последним можно отнести слабую производительность метода, передачу небольших объемов информации, сложность в составлении базы. Устойчивость методов, генерирующих искусственный стеготекст, обеспечивается заданными правилами грамматики. 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. Наиболее эффективный метод анализа использует прогнозирование для выявления искусственной природы текста, порожденного программами *Texto* и *Markov-Chain-Based,* где используются методы, учитывающие корреляцию слов между предложениями. Так, считается, что предложения, содержащие слова, встречающиеся только в технических текстах, не могут стоять рядом с предложениями, содержащими слова, встречающиеся только в текстах художественной литературы [4].

Выводы

В данном докладе представлен метод лексической стеганографии, основанный на генерировании текста, подобного естественному. Настоящий метод является развитием метода замены синонимов, описанного в работе [5]. Однако и он требует усовершенствования. В дальнейшем планируется решение задач плотного заполнения контейнера стегоинформацией, повышения «естественности» текстов, генерируемых стегосистемой для того, чтобы они не выделись из обычной массы файлов такого же формата и наполнения.

Список литературы

1. Стеганография. Материал из Википедии – свободной энциклопедии. Электронный ресурс. Режим доступа: http://ru.wikipedia.org/wiki/Стеганография

2. Основные положения стеганографии.Опубликовано: журнал "Защита информации. Конфидент", №3, 2000

3. Два метода синонимического перефразирования в лингвистической стеганографии. И. А. Большаков. Центр Компьютерных Исследований [Электронный ресурс]. Режим доступа: http://www.dialog-21.ru/Archive/2004/Bolshakov.htm

4. Разработка методов обеспечения безопасности использования информационных технологий, базирующихся на идеях стеганографии. Нечта И. В. Автореферат. [Электронный ресурс]. Режим доступа: www.sibsutis.ru/images/259_avtoreferat_nechta.docx

5. Метод кодирования произвольной двоичной информации на основе лингвистических ресурсов. Ларионова К. Е., Губенко Н. Е., [Электронный ресурс]. Режим доступа: http://masters.donntu.edu.ua/2009/fvti/larionova/library/article11.htm