Учреждение образования

«Белорусский Государственный Технологический Университет»

**Лабораторная работа №13**

**«**Исследование методов текстовой стеганографии»

Выполнил:

Студент 3 курса 1 группы ФИТ

Дырда Дмитрий Геннадьевич

Минск 2020

**Цель:** изучение стеганографических методов осаждения/ извлечения тайной информации с использованием электронного файла-контейнера текстового формата, приобретение практических навыков программной реализации методов.

**Классификация, сущность и основные особенности базовых методов текстовой стеганографии**

В текстовой стеганографии в качестве контейнера используется файл текстового типа.

Методы текстовой стеганографии:

* Синтаксические (не затрагивают семантику текстового сообщения);
* Лингвистические (трансформация файлов-контейнеров с сохранением смыслового содержания текста).

Базовые синтаксические методы:

* Изменение расстояния между строками текста (Line-Shift coding). Выделяется максимальное и минимальное расстояния между строками, позволяющее кодировать соответственно символы «1» и «0» осаждаемого сообщения. На рис. 1 изображен пример данного метода;

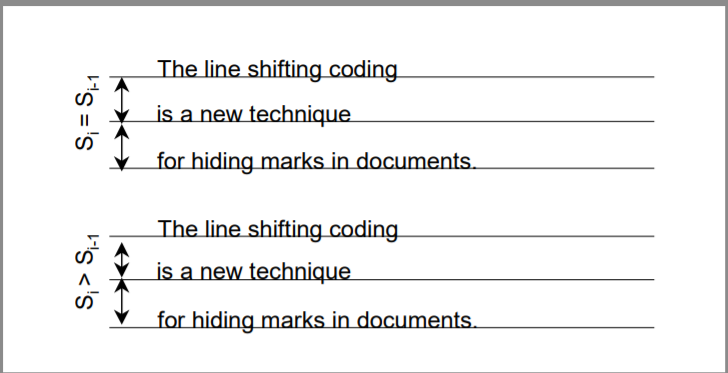


Рисунок 1 – Line-shift coding

* Изменение расстояния между словами (Word-shift coding) в одной строке текста. осаждение информации основано на модификации расстояния между словами текста-контейнера. Пример на рисунке 2.

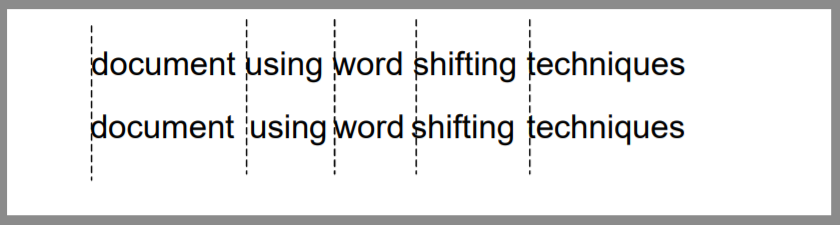


Рисунок 2 – Word-shift coding

* Изменение количества пробелов между словами (частный случай метода Word-Shift Coding). чередование одинарного пробела и двойного (хх­\_хх\_\_хх) кодирует «1», переход же с двойного пробела на одинарный кодирует «0» (хх\_\_хх\_хх). Пример на рисунке 3;

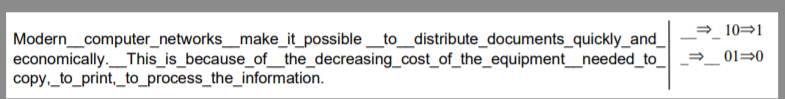


Рисунок 3 – изменение количества пробелов между словами

* Внесение специфических изменений в шрифты, т. е. начертания отдельных букв (Feature Coding). Заключается в изменении написания отдельных букв используемого стандартного шрифта. Пример на рисунке 4;

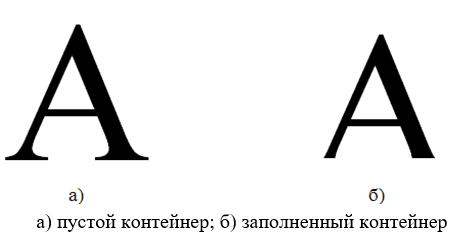


Рисунок 4 – Feature coding

* Изменение интервала табуляции; аналогичен вышеописанному методу изменения количества пробелов, только в этом случае меняется не количество пробелов, а соответственно расстояние между строками и интервал табуляции;
* Null Chipper. Предполагает размещение тайной информации на установленных позициях слов или в определенных словах текста-контейнера, который, как правило, лишен логического смысла (как видно, действительно лепет). Пример на рисунке 5;

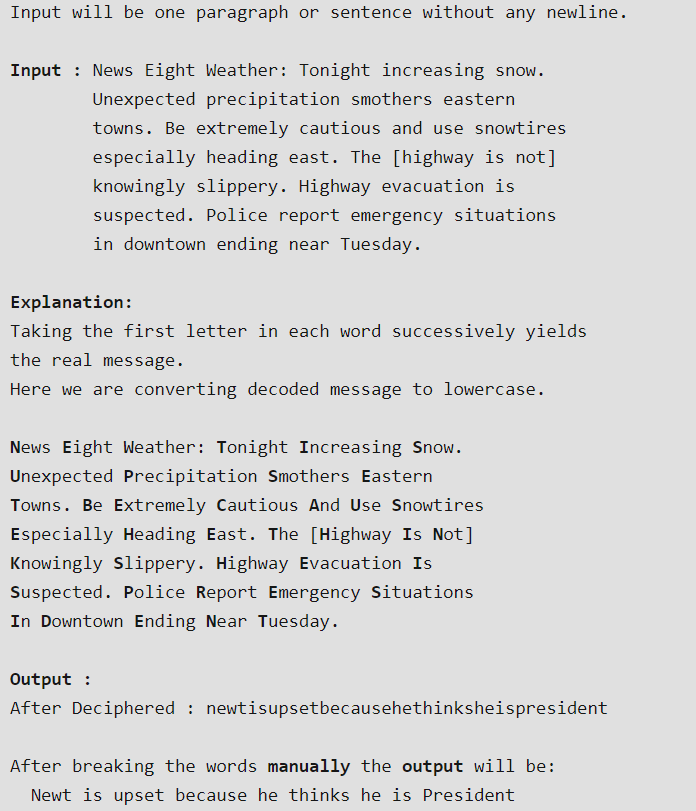


Рисунок 5 – Null cipher

* Увеличение длины строки; предусматривает искусственное увеличение длины каждой строки за счет пробелов: например, нет пробела (определяется положением знака перехода на новую строку) – «0», один пробел – «1». Пример на рисунке 6;



Рисунок 6 – Увеличение длины строки

* Использование регистра букв. Для обозначения бита секретного сообщения, представленного единицей, используется символ нижнего регистра, а нулем — верхнего (или наоборот);
* Использование невидимых символов; знак пробел кодируется символом с кодом 32, но в тексте его можно заменить также символом, имеющим код 255 (или 0), который является «невидимым» и отображается как пробел.

Синтаксические методы неустойчивы к форматированию текста, и поэтому информация может быть потеряна при простом применении иного стиля форматирования текста-контейнера, скрывающего в себе стегосообщение. К тому же с помощью синтаксических методов можно передать незначительное количество информации.

Некоторые лингвистические методы:

* Метод синонимов. Возьмем подмножество синонимов: {«тайный», «секретный», «конфиденциальный», «доверительный»}. В приведенном подмножестве каждое слово имеет единственное одинаковое смысловое значение, что позволяет закодировать каждое слово своим уникальным кодом (т. е. выполнить операцию осаждения), например, «доверительный» – 00, «конфиденциальный» – 01, «секретный» – 10, «тайный» – 11. Подобное кодирование позволяет выбирать одно из четырех слов в зависимости от двух бит секретного сообщения. Отметим, что при этом, независимо какое из четырех слов будет выбрано, семантика сообщения не изменится. Очевидно, что при этом количество символов, соответствующих одному из синонимов используемого подмножества, зависит от общего числа элементов в подмножестве. Кроме того, обеим сторонам стеганосистемы должен быть известен общий алгоритм кодирования, т. е. один из ключей системы. В каждом подмножестве синонимов их упорядочивание должно выполняться по одному и тому же алгоритму и у отправителя сообщения, и у его получателя.
* Метод переменной длины слова. Основан на том, что длина слов в сообщении зависит от содержания секретного сообщения и способа кодирования слов: обычно одно слово текста-контейнера определенной длины кодирует два бита информации из стеганосообщения; например, слова текста длиной в 4 и 8 символов могут означать комбинацию бит «00», длиной в 5 и 9 — «01», 6 и 10 — «10», 7 и 11 букв — «11»; слова короче 4 и длиннее 11 букв можно вставлять где угодно для лексической и грамматической связки слов в предложении – программное приложение, которое декодирует принятое сообщение (извлекает сообщение из стеганоконтейнера), будет просто игнорировать их;
* Метод первой буквы. Программа-помощник в этом методе накладывает ограничение уже не на длину слова, а на первую (можно на вторую) букву; обычно одну и ту же комбинацию могут кодировать несколько букв, например, комбинацию «101» означают слова, начинающиеся с «А», «Г» или «Т;
* 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. Пример на рисунке 7;

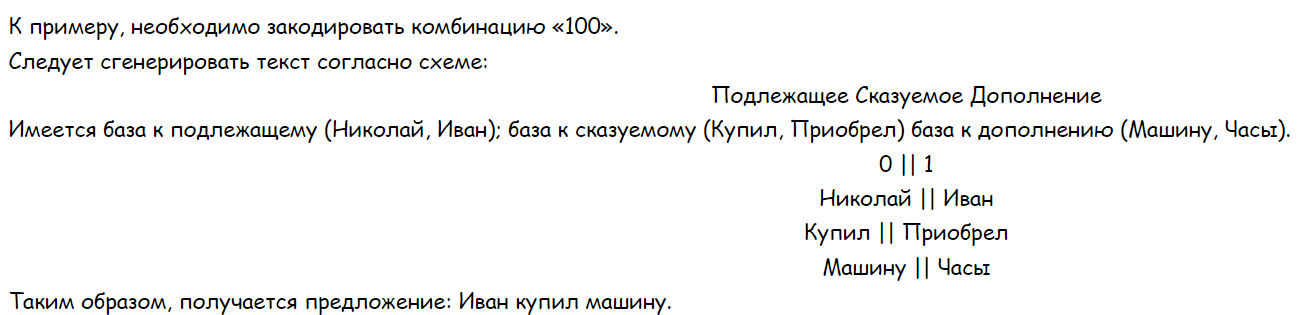


Рисунок 7 – Мимикрия

**Метод текстовой стеганографии на основе модификации цветовых параметров символов текста-контейнера**

По сути, данный метод схож с классическим метод наименее значащих бит. Он опирается на использовании цветовой модели RGB. Данный метод был рассмотрен в лабораторной работе 12. Здесь же мы еще раз кратко опишем его: суть метода замена наименее значащего бита (Least Significant Bits - LSB) заключается в сокрытии информации путем изменения последних битов изображения, кодирующих цвет на биты скрываемого сообщения. Разница между пустым и заполненным контейнерами должна быть не ощутима для органов восприятия человека.

**Метод на основе апроша**

Апрош (рис. 8) определяет расстояние между соседними символами текста. Фактически апрош состоит из двух таких расстояний – полуапрошей, являющихся как бы пространством, прилегающим к каждому из символов-соседей.



Рисунок 8 – Апрош

Идея метода заключается в следующем. Встраивание сообщения в контейнер может быть основано на модификации базового (устанавливаемого текстовым процессором по умолчанию) значения апроша, *ао*, его изменением от базового до некоторого максимального *аmax* (или минимального), *аmin*, которое зрительно не должно отличаться от стандартного. Такое изменение производится с определенным шагом (дискретно) Δ*аi*, каждому значению которого присваивается определенный бит или определенная комбинация бит.

Особенностью рассматриваемого метода является возможность одноразового размещения (в апроше одного символа) числа бит, определяемого дискретной разницей между минимальным и максимальным значениями Δ*а*. Например, если отсчет вести от Δ*амин* до установленного интервала Δ*аt* в виде параметра 0,1\**n*д (пт), то количество условных дискретных единиц *n*д, представленное в бинарном виде, определяет число бит, которые можно таким образом разместить; например, Δ*амин* = –0,5 пт, а Δ*аt* = 0,3 пт. Разница между этими величинами составляет 0,8 пт: 8\*0,1 или *n*д*t* = 8 ( в двоичном виде – 1000; в первом приближении именного такую бинарную комбинацию можно разместить (осадить) путем модификации конкретного апроша). На этой основе могут быть разработаны различные варианты кодировки осаждаемых комбинаций.

**Метод на основе кернинга**

В текстовых документах встречаются такие сочетания знаков, которые образовывают визуальные «дыры» либо «сгущения». Например, в текстах на основе кириллицы – это такие сочетания: «ГА», «TA», «ATA», «ЬТ» и т. п., на основе латиницы – «AY», «AV», «T;», «ff», а на основе греческого алфавита – «ΘΑ», «ΔΟ», «λκ» и др. Такие сочетания называются кернинговыми парами.

Под кернингом обычно понимается процесс изменения межсимвольного расстояние между отдельными парами символов или кернинговыми парами. Пример кернинга приведен на рисунке 9.



Рисунок 9 – Пример кернинга

Таким образом, технология кернинга, появившаяся в полиграфии после внедрения фотонабора (а затем и компьютерного набора), включает подбор межбуквенных интервалов для конкретных пар букв с целью улучшения внешнего вида и удобочитаемости текста. Такой избирательный подбор позволяет компенсировать неравномерности визуальной плотности текста, получаемой при использовании стандартных апрошей для каждой буквы.

Следуя общепринятой методике, кернинг, как параметр, будем измерять в тысячных долях круглой шпации (*Еm*) — единицы измерения, которая определяется относительно текущего размера шрифта и равна ширине символа «М». Например, для шрифта размером 6 пунктов 1 круглая шпация равна 6 пунктам, а для шрифта размером 10 пунктов — 10 пунктам. Таким образом, размер кернинга, σ, строго пропорционален текущему размеру шрифта. Сдвиги букв относительно автоматически установленного межсимвольного расстояния (измеряемого, например, апрошем) можно производить с различным шагом: от 0,01 до 0,04 величины *Еm*, в зависимости от нужной точности.

**Особенности программной реализации методов**

Общий принцип работы некоторого условного приложения проиллюстрирован на диаграмме деятельности (рис. 10).

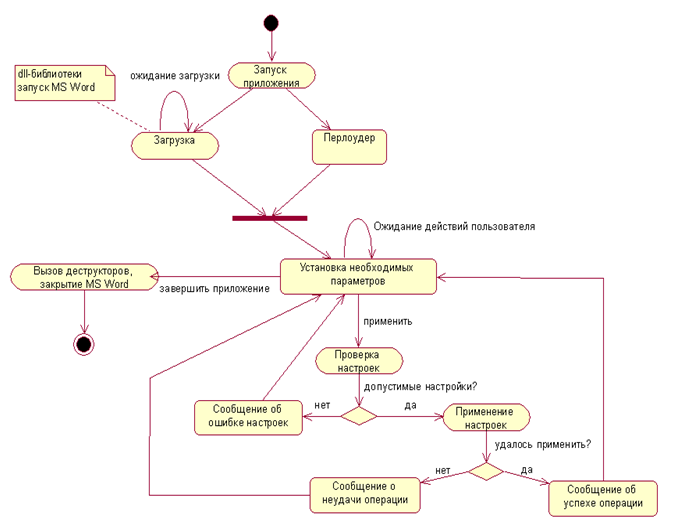


Рисунок 10 - Диаграмма деятельности программного средства,

реализующего методы текстовой стеганографии на основе MS Word

**Краткое описание специализированных программных средств**

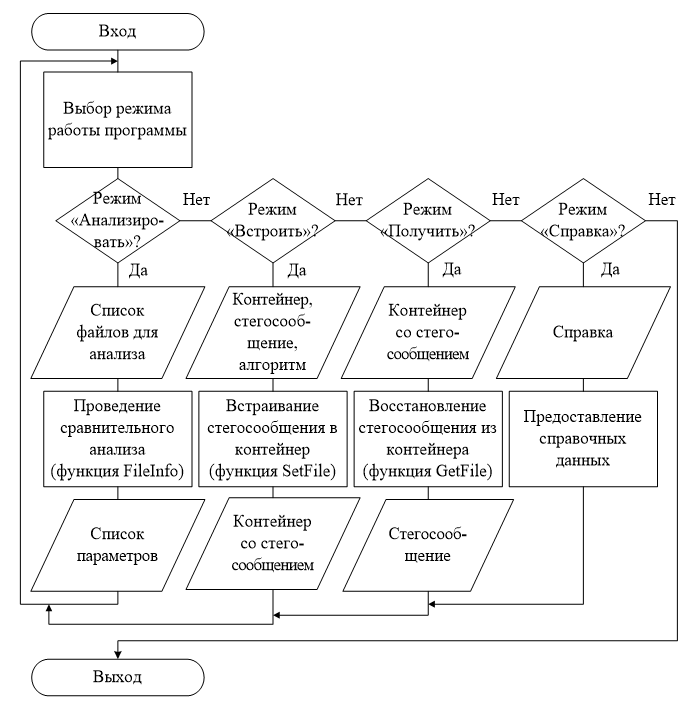


Рисунок 11 - Пример общей блок-схема функционирования программного средства, реализующего методы текстовой стеганографии