Министерство образования Республики Беларусь

Учреждение образования «Белорусский государственный технологический университет»

Факультет информационных технологий

Кафедра информационных систем и технологий

**Отчет к лабораторной работе №13**

**«Исследование методов текстовой стеганографии»**

Выполнил:

студентка 3 курса 2 группы

Черноок Ю. С.

Проверил:

ассистент

Копыток Д. В.

Минск 2021

Цель: изучение стеганографического метода встраивания/извлечения тайной информации с использованием электронного файла-контейнера текстового формата, приобретение практических навыков программной реализации методов.

**Теоретическая часть**

Стеганографическая система (stegosystem, стегосистема или стеганосистема – в русскоязычной тематической литературе используются оба сокращения) – совокупность средств и методов, которые используются для формирования скрытого канала передачи (или хранения) информации.

При этом скрытый канал организуется на базе и внутри открытого канала с использованием особенностей восприятия информации. «Скрытость» канала передачи тайной информации отличает стеганографии от криптографии: в первом случае тайной является сам факт наличия канала (передачи информации).

Абстрактно стеганографическая система обычно определяется, как некоторое множество отображений одного пространства (множества возможных сообщений, М) в другое пространство (множество возможных стеганосообщений, S, и наоборот. Основные компоненты стеганосистемы:

* контейнер, С (файл-контейнер или электронный документ произвольного формата), в котором размещается (осаждается, скрывается) тайное сообщение, М; именно контейнер является упомянутым скрытым каналом;
* тайное сообщение, М, осаждаемое в контейнер для передачи или хранения (например, с целью доказательства или защиты авторских прав на документ-контейнер; здесь речь может идти о невидимых цифровых водяных знаках, ЦВЗ);
* ключи или ключевая информация, K системы, выполняющие ту же функцию, что и криптографические ключи; ключей может быть несколько, в соответствии с этим современные стеганосистемы характеризуют как многоключевые: один ключ отождествляется с методом осаждения/извлечения тайной информации, другой – с выбором элементов (например, битов) контейнера для его модификации при осаждении тайной информации, третий (или третьи) – для предварительного (перед осаждением) преобразования тайной информации (например, на основе помехоустойчивого кодирования, сжатия или зашифрования) и т. д.;
* контейнер с осажденным сообщением или стеганоконтейнер, S, который передается по открытому каналу, также являющемуся важным компонентом анализируемой системы; стеганоконтейнер будем именовать также стеганосообщением;
* для полноты упомянем также субъектов системы: отправителя и получателя.
* В зависимости от формата документа-контейнера цифровую (или компьютерную) стеганографию подразделяют на классы:
* аудиостеганография,
* видеостеганография,
* графическая стеганография,
* текстовая стеганография и др.

Стеганографической системой ∑ будем называть совокупность сообщений M, контейнеров C, ключей K, стеганосообщений (заполненных контейнеров) S и преобразований (прямого F и обратного F -1), которые их связывают:

∑ = (M, C, K, S, F, F -1).

Как видим, сущностью рассматриваемой системы является тайное хранение или передача одной информации в другой информации, которая является открытой.

При построении стеганосистемы должны, таким образом, учитываться следующие основные положения:

* свойства контейнера должны быть модифицированы так, чтобы изменение невозможно было выявить при визуальном контроле; это требование определяет качество сокрытия внедряемого сообщения: для обеспечения беспрепятственного прохождения стеганосообщения по каналу связи оно никоим образом не должно привлечь внимание атакующего;
* противник (интруз) имеет полное представление о стеганографической системе и деталях ее реализации; единственной информацией, которая остается ему неизвестной, 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 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.

Многообразие методов текстовой стеганографии подразделяется на синтаксические методы, которые не затрагивают семантикутекстового сообщения, и лингвистические, которые основанына эквивалентной трансформации текстовых файлов-контейнеров, сохраняющей смысловое содержание текста, его семантику.

К синтаксическим методам компьютерной стеганографии, которые характеризуются сравнительно невысокой эффективностью(с точки зрения объема встраиваемой информации), относятся следующие (такие методы мы отнесем к числу базовых синтаксических методов):

* метод изменения межстрочныхинтервалов
* изменение расстояния между словами
* изменение количества пробелов между словами
* на основе внесения специфических изменений в шрифты
* изменение интервала табуляции
* Null Chipper
* увеличение длины строки
* использование регистра букв
* использование невидимых символов

Существуют также стеганографические методы, которые интерпретируют текст как двоичное изображение. Необходимо отметить,что данные методы нечувствительны к изменению масштаба документа, что обеспечивает им хорошую устойчивость к большинствуискажений, которые могут иметь место при активных атаках.

К числу основных лингвистических методов относятся:

* метод синонимов
* метод переменной длины слова
* метод первой буквы
* мимикрия

**Метод на основе апроша**

Апрош определяет расстояние между соседними символамитекста. Фактически апрош состоит из двух таких расстояний – полуапрошей, являющихся как бы пространством, прилегающимк каждому из символов-соседей.Согласно существующим техническим правилам набора нормальный апрош должен быть равен половине кегля (размера) шрифта.

Идея метода заключается в следующем. Встраивание сообщения в контейнер может быть основано на модификации базового (устанавливаемого текстовым процессором по умолчанию)значения апроша, его изменением от базового до некоторого максимального, которое зрительно недолжно отличаться от стандартного. Такое изменение производится с определенным шагом (дискретно), каждому значениюкоторого присваивается определенный бит или определенная комбинация битов.

**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. Особенности «кернингования» приведены на рис. 13.6.Под кернингом обычно понимается процесс изменения межсимвольного расстояние между отдельными парами символов иликернинговыми парами (именно фактор парности отличает кернингот апроша).

**Практическая часть**

На рисунке 1 показана реализация алгоритма сохранения информации методом модификации апроша.

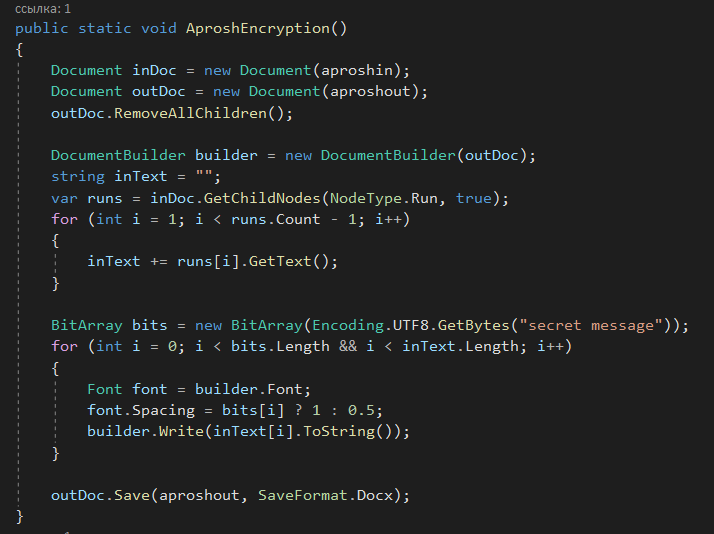


Рисунок 1 – Модификация апроша

Сперва получается текст-контейнер, в данном случае читается данный текстовый файл. Сохраняемое сообщение преобразуется в набор битов. При переборе всех символов контейнера каждому символу задается значение апроша в зависимости от значения соответствующего бита, при этом символ с заданным апрошем сохраняется в выходной файл. Обход происходит, пока не закончится текст контейнера либо не закончится секретное сообщение. Таким образом на каждый бит секретного сообщения приходится один символ контейнера, однако это отношение можно увеличить, ставя не 2 варианта размера апроша, а больше.

На рисунке 2 показывается реализация получения секретного сообщения из контейнера.

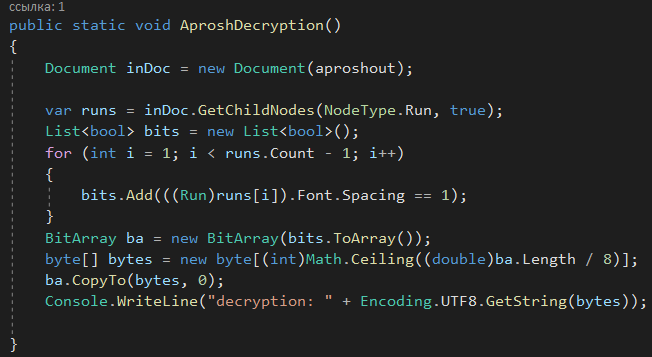


Рисунок 2 – Получения секретного сообщения

Для расшифрования обходится каждый символ исходного текста и берется значение его апроша. В данном случае максимальное значение апроша соответствует значению бита равным единице, при минимальном апроше бит равняется нулю. Таким образом, из каждого символа получается бит сообщения, а из них получают текст сообщения.

На рисунке 3 показывается реализация модификации длины строки путем добавления пробела в конец.

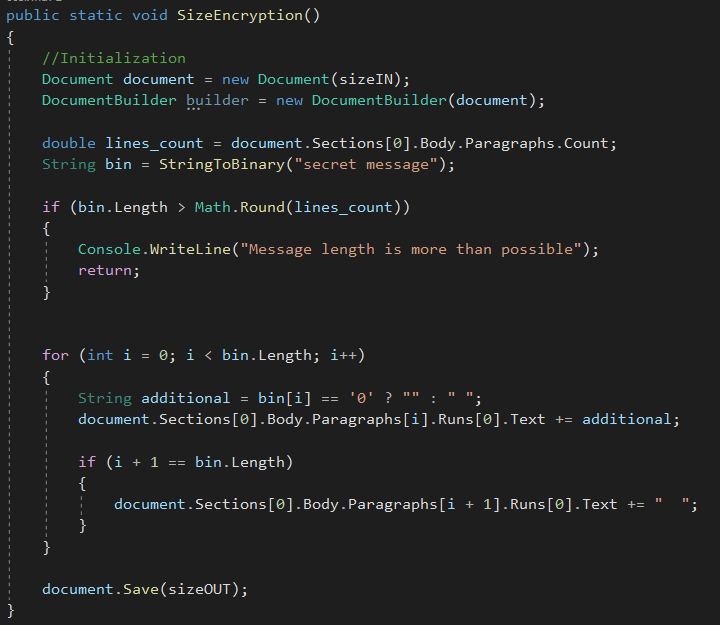


Рисунок 3 – Реализация модификации расстояния между строками

### Символы контейнера перебираются, после чего, при нахождении символа пробела в конце, берет его согласно текущему шифруемому биту сообщения. Таким образом контейнер должен быть в разы больше, чем при модификации без добавлении пробела.

Алгоритм снова повторяет соответствующий алгоритм на основе модификаций апроша, однако в данном случае учитываются только пробелы, где опять же минимальный апрош означает нулевой бит а максимальный – единичный.

На рисунке 4 показан исходный контейнер.

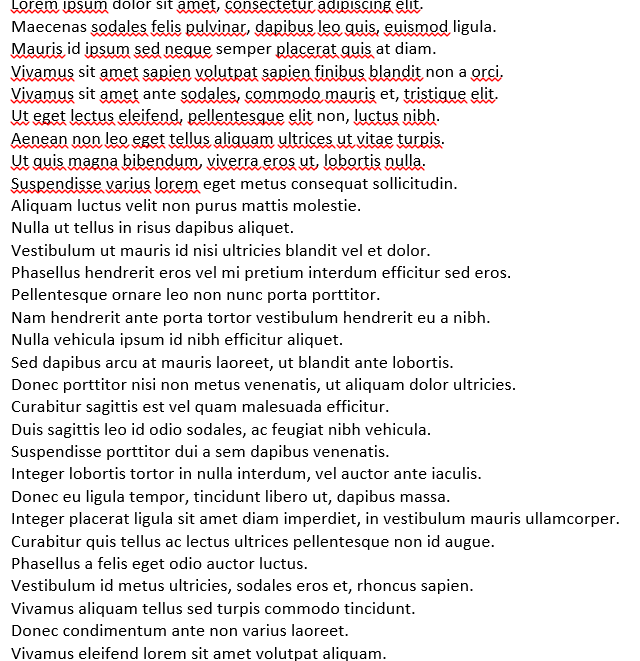


Рисунок 4 – Исходный контейнер

Контейнер должен быть большим, чтобы вместить все символы.

На рисунке 5 показывается заполненный контейнер.

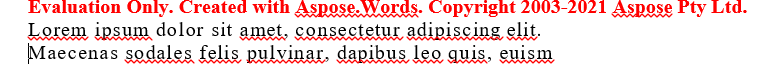


Рисунок 5 – Заполненный контейнер апроша

На рисунке 6 показан заполненный контейнер с разными длинами строк.

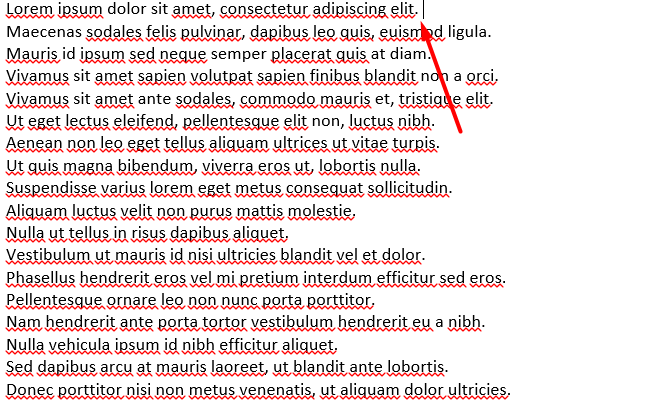


Рисунок 6 – Заполненный контейнер с различными длинами строк

В этом случае длины слов не различаются, однако длины строк различны из-за пробела в конце строки.

Вывод: я изучила стеганографич2еский метод встраивания/извлечения тайной информации с использованием электронного файла-контейнера текстового формата, приобрела практические навыки программной реализации методов.