МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования «БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ

ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет информационных технологий

**ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №15 НА ТЕМУ:**

**Исследование методов текстовой стеганографии**

Выполнила студентка 3 курса 4 группы

Коржова Валерия Сергеевна

Минск 2023

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

**Теоретические сведения**

К текстовой стеганографии относятся методы, предусматривающие использование в качестве контейнера файла-документа текстового типа.

Текстовая стеганография:

1) Синтаксические методы:

– изменение расстояния между строками электронного документа;

– изменение расстояния между словами;

– изменение количества пробелов между словами;

– на основе внесения специфических изменения в шрифты;

– изменение интервала табуляции;

– Null Chipper;

– увеличение длины строки;

– использование регистра букв;

– использование невидимых символов.

Достоинства:

– Легко применяются к любому тексту (независимо от содержания, назначения, языка);

– Легко реализуются в программном коде (т.к. они полностью автоматические)

Недостатки:

– Невысокая эффективность (объем встраиваемой информации);

– Перечисленные методы работают успешно до тех пор, пока тексты представлены в коде ASCII;

– Неустойчивы к форматированию текса;

– Неустойчивы к изменению масштаба документа.

2) Лингвистические методы:

– Метод синонимов;

– Метод переменной длины слова;

– Метод первой буквы;

– Мимикрия.

**Практическая часть**

**Задание 1.** Разработать приложение, позволяющее использовать стенографический метод на основе изменения расстояния между строками.

В качестве контейнера был выбран word-документ в формате docx. Приложение было реализовано на языке Python с устанавливаемым модулем «docx» для работы с документами этого формата.

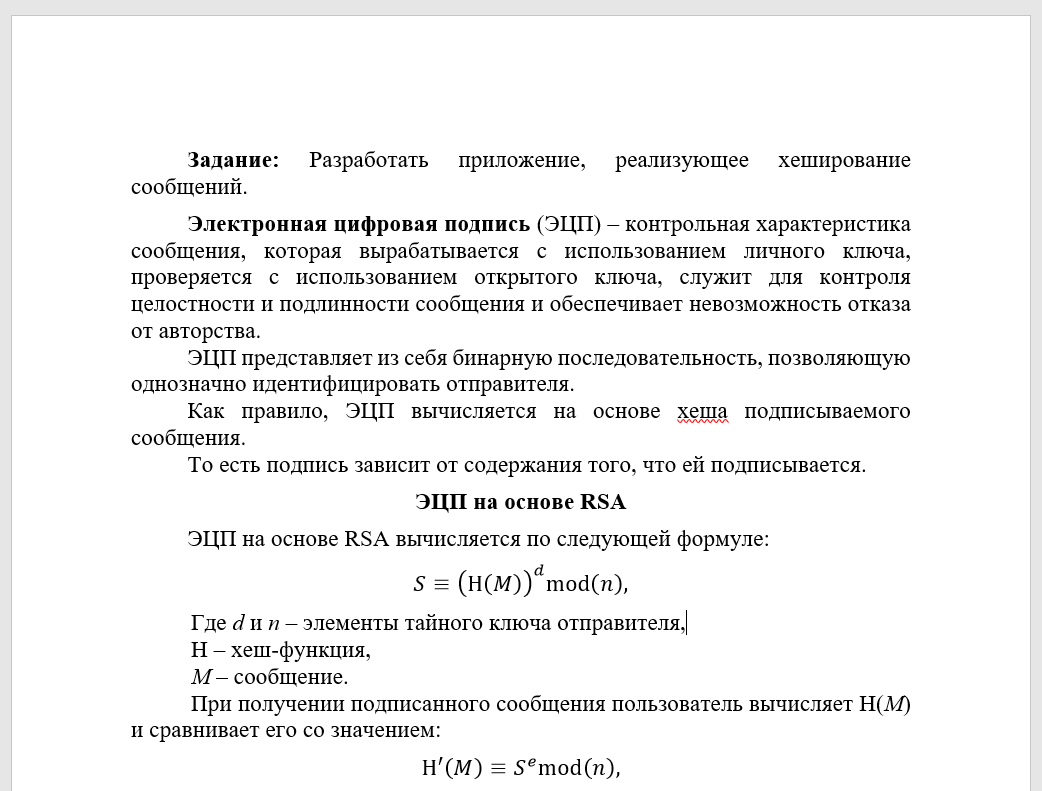


Рисунок 1.1 – Документ-контейнер

Метод предполагает установку для каждого абзаца документа своего межстрочного интервала, значение которого кодирует последовательность из двух битов сообщения.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Интервал** | 1.0 | 1.1 | 1.2 | 1.3 |
| **Биты** | 00 | 01 | 10 | 11 |

При кодировании сначала для каждого абзаца устанавливается значение межстрочного интервала 1.08. Абзацы, имеющие такой интервал, не будут учитываться при вытягивании сообщения из контейнера. Далее сообщение переводится в бинарный вид и разбивается на группы по 2 бита. Абзацам последовательно задаются межстрочные интервалы в соответствии с битовыми группами. В конце контейнер с осаждённым сообщением сохраняется по новому пути.

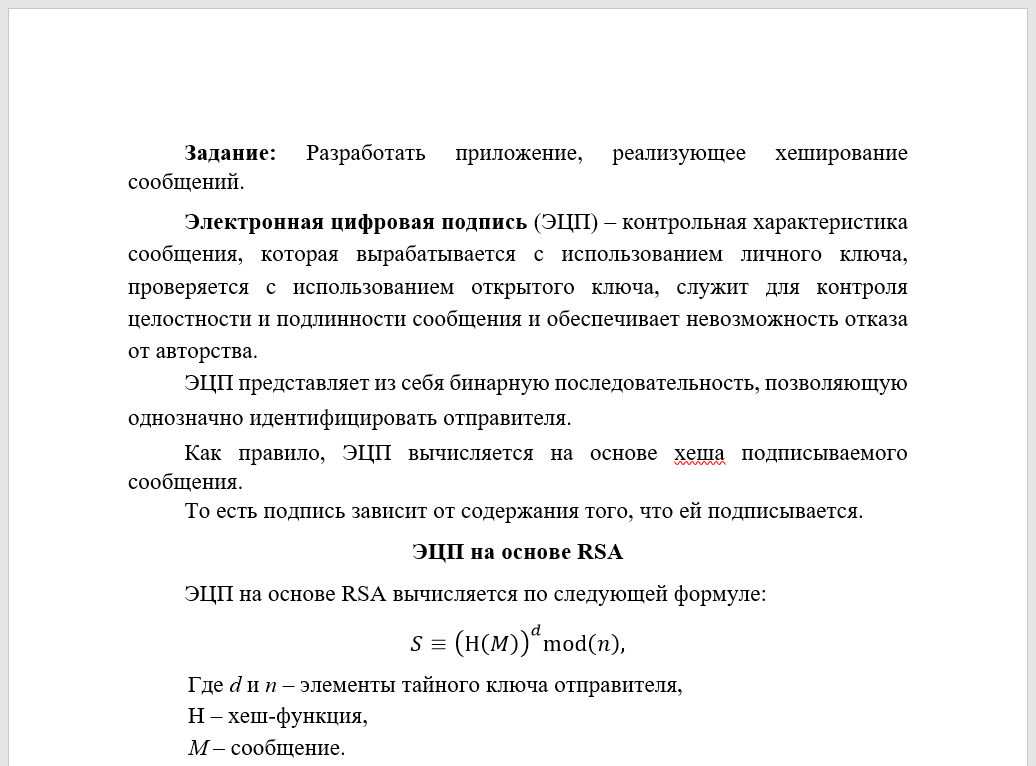


Рисунок 1.2 – Контейнер с осаждённым сообщением «mikhail»

При вытягивании осаждённого сообщения анализируются межстрочные интервалы каждого абзаца. Если интервал равен одному из значений, кодирующих битовую группу, в буфер декодированного сообщения записываются соответствующие биты.

**Задание 2.** Разработать приложение, позволяющее использовать стенографический метод на основе кернинга.

Некоторые пары символов (кернинговые пары) образуют между собой заметные интервалы, если они стоят рядом в тексте. Поэтому для таких пар применяется кернинг – сдвиг между символами. Он может быть как положительным, так и отрицательным. Не стоит путать кернинг с апрошем. Апрош – расстояние между двумя символами текста. Кернинг – сдвиг относительно апроша.



Рисунок 2.1 – Пример кернинга

Приложение для использования данного стенографического метода было разработано на языке JavaScript для платформы NodeJS с использованием устанавливаемого модуля «jsdom». 2 разработанных скрипта позволяют осаждать сообщения в html-файлах и вытягивать их.

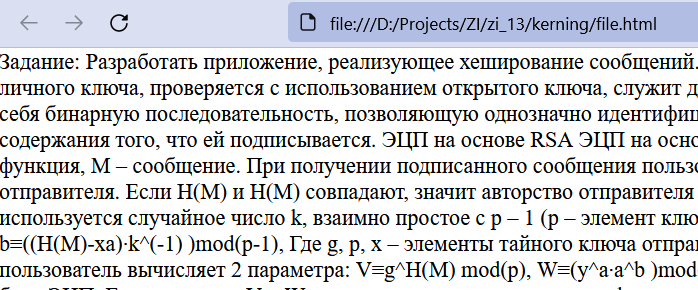


Рисунок 2.2 – Пустой HTML-контейнер

При осаждении сначала сообщение переводится в бинарный вид. Затем для каждого символа в тексте контейнера задаётся значение кернинга, равное половине пикселя. Символы с таким значением сдвига при вытягивании не учитываются. Затем последовательно для каждого символа текста контейнера задаются значения кернинга в соответствии с битами сообщения: 0 пикселей для нулевых битов и 1 пиксель для единичных битов.

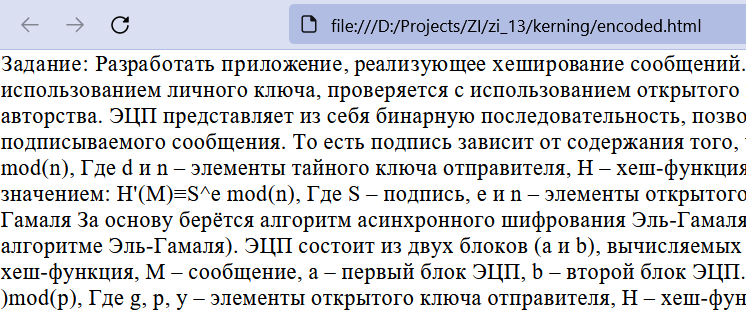


Рисунок 2.3 – Контейнер с осаждённым сообщением

При вытягивании сообщения анализируются все символы текста контейнера. Если значение кернинга равно 0 или 1 пиксель, соответствующий бит записывает в буфер для декодированного сообщения.

**Вывод:** В ходе лабораторной работы были изучены методы текстовой стеганографии и разработаны приложения для использования методов на основе расстояний между строками и кернинга.

**Контрольные вопросы**

**1. В чем состоит сущность методов текстовой стеганографии?**

Сущность методов текстовой стеганографии заключается в скрытом передаче информации в текстовых данных, таких как текстовые документы, электронные письма или веб-страницы, без вызывания подозрений у наблюдателей.

**2. Охарактеризовать методы синтаксической текстовой стеганографии**

Синтетические методы:

– изменение расстояния между строками электронного документа;

– изменение расстояния между словами;

– изменение количества пробелов между словами;

– на основе внесения специфических изменения в шрифты;

– изменение интервала табуляции;

– Null Chipper;

– увеличение длины строки;

– использование регистра букв;

– использование невидимых символов.

Достоинства:

– Легко применяются к любому тексту (независимо от содержания, назначения, языка);

– Легко реализуются в программном коде (т.к. они полностью автоматические)

Недостатки:

– Невысокая эффективность (объем встраиваемой информации);

– Перечисленные методы работают успешно до тех пор, пока тексты представлены в коде ASCII;

– Неустойчивы к форматированию текса;

– Неустойчивы к изменению масштаба документа.

**3. Охарактеризовать методы лингвистической текстовой стеганографии.**

Лингвистические методы:

– Метод синонимов;

– Метод переменной длины слова;

– Метод первой буквы;

– Мимикрия.

**4.** Д**ать оценку стеганографической стойкости методов текстовой стеганографии при конвертации текста-контейнера в иной текстовый формат.**

Оценка стеганографической стойкости методов текстовой стеганографии при конвертации текста-контейнера в другой текстовый формат зависит от различных факторов, включая выбранный метод и специфические особенности конвертации.

Замена символов: этот метод может быть относительно простым, но стойкость зависит от способа замены и сложности обнаружения изменений. Если замена символов проводится случайным образом и сложно обнаружить паттерны, это может быть стойким методом.

Внедрение в пробелы и форматирование: стойкость этого метода зависит от того, насколько естественно выглядят изменения в пробелах и форматировании. Если изменения явно выделяются или вызывают подозрения, метод может быть менее стойким.

Внедрение в форматирование документов: этот метод может быть стойким, если изменения в форматировании выглядят естественными и не вызывают подозрений. Однако, если специфическое форматирование может быть обнаружено или удалено при конвертации в другой формат, стойкость может быть низкой.

Лингвистические стеганографические методы: стойкость этих методов зависит от сложности анализа текста и способа внедрения скрытой информации. Если методы лингвистического анализа сложны и обнаружение скрытой информации требует специализированных инструментов, это может быть стойким методом.

**5. Дать оценку стеганографической стойкости методов текстовой стеганографии при визуальном стеганоанализе текста-контейнера.**

Замена символов: этот метод может быть уязвимым при визуальном стеганоанализе, особенно если замена символов приводит к отклонениям от ожидаемого распределения символов или вызывает подозрения. Простые методы замены символов могут быть относительно легко обнаружены.

Внедрение в пробелы и форматирование: если изменения в пробелах и форматировании визуально не выделяются и выглядят естественно, метод может быть стойким при визуальном стеганоанализе. Однако, если есть специфические паттерны или аномалии, которые могут быть обнаружены визуально, стойкость может быть низкой.

Внедрение в форматирование документов: стойкость этого метода при визуальном стеганоанализе зависит от того, насколько изменения в форматировании могут быть обнаружены. Если изменения не вызывают подозрений и не выделяются визуально, метод может быть стойким.

Лингвистические стеганографические методы: эти методы могут быть уязвимыми при визуальном стеганоанализе, особенно если изменения в тексте вызывают подозрения или нарушают естественные языковые паттерны. Если визуальный анализ текста выявляет аномалии, стойкость методов лингвистической стеганографии может быть низкой.

**6. Дать общую характеристику стеганоанализу в области текстовой стеганографии на основе метода «χ-квадрат».**

Метод "χ-квадрат" является одним из основных методов стеганоанализа в области текстовой стеганографии. Он используется для статистического анализа текста с целью обнаружения скрытой информации. Общая характеристика стеганоанализа на основе метода "χ-квадрат" включает следующие аспекты.

*Статистический анализ*

Метод основан на сравнении наблюдаемых частот встречаемости символов или других статистических параметров в тексте с ожидаемыми частотами, которые можно вычислить на основе модели статистики текста. При стеганографической внедрении скрытой информации частоты могут измениться, и метод "χ-квадрат" позволяет обнаружить эти аномалии.

*Проверка гипотезы*

Метод используется для проверки статистических гипотез о наличии скрытой информации в тексте. Стеганоанализ заключается в сравнении наблюдаемого значения "χ-квадрат" статистики с критическим значением, определенным на основе выбранного уровня значимости. Если наблюдаемое значение превышает критическое значение, то гипотеза о наличии скрытой информации отвергается.

*Использование моделей статистики текста*

Метод требует знания или создания моделей статистики текста, которые представляют ожидаемые частоты символов или других статистических параметров. Модели могут быть созданы на основе обучающих данных или предварительного анализа текста. Они служат базой для сравнения и вычисления "χ-квадрат" статистики.

*Чувствительность к изменениям*

Метод может быть чувствителен к изменениям в тексте, вызванным внедрением скрытой информации. Если изменения вызывают значительные отклонения от ожидаемых частот или статистических параметров, "χ-квадрат" статистика может показать высокие значения, что указывает на возможное наличие скрытой информации.

**7. Что такое апрош? В чем состоит сущность стеганометода на основе модификации апроша?**

Апрош - расстояние между соседними символами текста. Встраивание сообщения в контейнер может быть основано на модификации базового (устанавливаемого текстовым процессором по умолчанию) значения апроша, его изменением от базового до некоторого максимального значения, которое зрительно не должно отличаться от стандартного. Такое изменение производится с определенным шагом, каждому значению которого присваивается определенный бит или определенная комбинация битов.

**8. Что такое кернинг? В чем состоит сущность стеганометода на основе модификации кернинга?**

Кернинг – это сдвиг между символами, используемый, как правило, для лучшей читаемости текста. Сущность метода на основе модификации кернинга заключается в использовании изменений промежутков между символами для внедрения скрытой информации. Небольшие изменения в кернинге между символами в тексте-контейнере используются для кодирования скрытого сообщения.

**9. Дать сравнительную оценку методов на основе модификации пространственно-геометрических и цветовых параметров символов текста-контейнера.**

Методы на основе изменения цветовых параметров символов схожи с методом замены наименьших битов. Их эффективности при визуальном анализе напрямую зависит от числа модифицируемых битов цветов. Чем больше используется бит, тем меньше стойкость метода. При этом файловый вес контейнера в таких методах не изменяется.

Изменение же пространственно-геометрических параметров имеет меньшую стойкость, так как легко обнаруживается при визуальном анализе (легко заметить отсутствие засечек у символов).