УДК 003.26

Тычко А.Ю.
Студент магистратуры
Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А.Бонч-Бруевича
Россия, г. Санкт-Петербург
Герлинг Е.Ю., кандидат технических наук, доцент
доцент кафедры «Защищённые системы связи»
Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А.Бонч-Бруевича
Россия, г. Санкт-Петербург

ЛИНГВИСТИЧЕСКАЯ СТЕГАНОГРАФИЯ, СЛОЖНОСТИ ЦИФРОВОЙ РЕАЛИЗАЦИИ И ВОЗМОЖНЫЕ ОБЛАСТИ ПРИМЕНЕНИЯ

Аннотация: *В статье рассматривается один из методов лингвистической стеганографии. Приводится практический пример реализации. Определяются причины сложности цифровой реализации метода. А также потенциальные сферы применения.*

Ключевые слова: *стеганография, лингвистика, информация, информационная безопасность, лингвистическая стеганография.*

Annotation: *The article considers one of the methods of linguistic steganography. A practical example of implementation is given. The reasons for the complexity of the digital implementation of the method are determined. Also, potential applications are considered.*

Key words: *steganography, linguistics, information, information security, linguistic steganography.*

Сокрытие и передача некой произвольной информации в не вызывающем подозрении покрывающем сообщении называется

стеганографией. Покрывающим сообщением – массивом информации, в котором можно передать скрытое сообщение – в данной статье анализируется осмысленный текст, а методом – замена слов на синонимы. Лингвистическая стеганография с заменой синонимов представляет собой алгоритм вложения скрытого сообщения в текст путём замены некоторых слов на аналогичные, которые не изменяют содержание и основной смысл текста. Выбор заменить синонимичное слово или оставить неизменным определяет, какой символ должен быть вложен – 0 или 1.

Для более наглядного представления этого алгоритма возьмём отрывок из Капитанской дочки А.С. Пушкина и заменим часть слов на синонимы.

«Мне приснился сон, которого никогда не мог я ***забыть*** и в котором до сих пор ***наблюдаю*** нечто пророческое, когда соображаю с ним ***необычные*** обстоятельства моей жизни. Читатель извинит меня: ибо, ***возможно***, знает по опыту, как сродно человеку предаваться суеверию, несмотря на всевозможное презрение к предрассудкам. Я ***пребывал*** в том состоянии чувств и души, когда существенность, уступая мечтаниям, ***сходится*** с ними в неясных ***образах*** первосония. Мне казалось, буран еще свирепствовал и мы еще ***бродили*** по снежной пустыне... Вдруг ***узрел*** я вороты и въехал на барский двор нашей усадьбы. Первою мыслию моею было опасение, чтобы батюшка не прогневался на меня за невольное возвращение под ***крышу*** родительскую и не почел бы его ***намеренным*** ослушанием. С ***волнением*** я выпрыгнул из кибитки и ***наблюдаю***: матушка встречает меня на крыльце с видом глубокого огорчения. «Тише, — говорит она мне, — отец болен при смерти и желает с тобою проститься». Пораженный ***ужасом***, я иду за нею в спальню. Вижу, комната слабо освещена; у ***кровати*** стоят люди с печальными лицами. Я тихонько подхожу к постеле; матушка приподымает полог и говорит: «Андрей Петрович, Петруша приехал; он воротился, узнав о твоей болезни; благослови его». Я стал на колени и устремил глаза мои на больного. Что ж?.. Вместо отца моего вижу в постеле лежит мужик с черной

бородою, весело на меня ***посматривая***. Я в ***непонимании*** оборотился к матушке, говоря ей: «Что это значит? Это не батюшка. И к какой мне стати просить благословения у мужика?» — «Все равно, Петруша, — отвечала мне матушка, — это твой посажёный отец; поцелуй у него ручку, и пусть он тебя благословит...» Я не соглашался. Тогда мужик вскочил с ***кровати***, выхватил топор из-за спины и стал махать во все стороны.»

Для сравнения ниже представлен неизменённый оригинал:

«Мне приснился сон, которого никогда не мог я позабыть и в котором до сих пор вижу нечто пророческое, когда соображаю с ним странные обстоятельства моей жизни. Читатель извинит меня: ибо, вероятно, знает по опыту, как сродно человеку предаваться суеверию, несмотря на всевозможное презрение к предрассудкам. Я находился в том состоянии чувств и души, когда существенность, уступая мечтаниям, сливается с ними в неясных видениях первосония. Мне казалось, буран еще свирепствовал и мы еще блуждали по снежной пустыне... Вдруг увидел я вороты и въехал на барский двор нашей усадьбы. Первою мыслию моею было опасение, чтобы батюшка не прогневался на меня за невольное возвращение под кровлю родительскую и не почел бы его умышленным ослушанием. С беспокойством я выпрыгнул из кибитки и вижу: матушка встречает меня на крыльце с видом глубокого огорчения. «Тише, — говорит она мне, — отец болен при смерти и желает с тобою проститься». Пораженный страхом, я иду за нею в спальню. Вижу, комната слабо освещена; у постели стоят люди с печальными лицами. Я тихонько подхожу к постеле; матушка приподымает полог и говорит: «Андрей Петрович, Петруша приехал; он воротился, узнав о твоей болезни; благослови его». Я стал на колени и устремил глаза мои на больного. Что ж?.. Вместо отца моего вижу в постеле лежит мужик с черной бородою, весело на меня поглядывая. Я в недоумении оборотился к матушке, говоря ей: «Что это значит? Это не батюшка. И к какой мне стати просить благословения у

мужика?» — «Все равно, Петруша, — отвечала мне матушка, — это твой посажёный отец; поцелуй у него ручку, и пусть он тебя благословит...» Я не соглашался. Тогда мужик вскочил с постели, выхватил топор из-за спины и стал махать во все стороны.»

Изменённые слова демонстрируют лишь возможность использования синонимов. Сам алгоритм вложения информации предполагает изменение слова на синоним лишь при некоторых условиях. В каждой паре синонимов один будет использоваться при вложении «1» другой при вложении «0», и если слово, встречающееся в тексте, соответствует вкладываемому биту – то оно остаётся неизменным.

Применение лингвистической стеганографии в стилистическом художественном тексте легко обнаруживается при наличии оригинала, а потому для надёжности следует исключить использование в качестве покрывающих сообщений ранее опубликованные открытые тексты.

В данной статье рассматриваются сложности реализации автоматической системы вложения.

Наречия – самая удобная для вложения часть речи, но при этом не так часто используемая в каналах передачи текстовых сообщений. Нет сложностей преобразования – слова заменяются без изменений.

Существительные – самая распространённая часть речи, но имеет множество свойств, которые от автоматической системы замены будут требовать значительных ресурсов в случае, если выбранные синонимы будут отличаться друг от друга родом или склонением. Более того – для определения существительного в тексте будет требоваться или значительно расширить словарь, или сделать сложную систему анализа слов. В обоих случаях это увеличивает время обработки каждого слова – в первом случае значительно расширяется словарь, во втором – увеличивается структурная сложность. Каждый падеж имеет собственные окончания, которые разнятся от склонения и самих слов. Создание зависимости окончаний от слов вызовет

значительные проблемы в расширяемости словаря синонимов – каждый новый элемент необходимо подписывать множеством идентификаторов. Обработку множества идентификаторов в итоге разумнее заменить массивом слов в разных падежах. Но тут возникает необходимость дополнительных проверок – одно и то же слово в разных падежах может иметь одинаковые окончания и замена на другое может произойти с грамматической ошибкой. Систему, учитывающую все эти переменные создать возможно, но её целесообразность вызывает вопросы – стоит ли создавать столь сложную систему если эффективнее использовать оператора?

Прилагательные обладают свойствами, близкими к существительным, а потому также могут использоваться в системе замены синонимов. Большинство прилагательных можно заменять, учитывая только их окончания. Самое важное учитывать контекст текста и не использовать разговорные, редко используемые или контекстные замены в текстах, где они будут не к месту.

Нынешние цифровые системы способны проводить множественные операции в секунду, скорость обработки будет достаточно велика, но генерация самого текста и дальнейшее расширение словарей и алгоритмов будет зависеть от человека – что в значительной мере влияет на скорость вложения и передачи скрытой информации. Изменения всегда будут направлены на расширение системы и словарей, что может привести к тому что сама цифровая обработка по объёму может превосходить сами обрабатываемые покрывающие сообщения.

Из этого исходит дополнительная задача лингвистической стеганографии – передача самих словарей, с помощью которых будет производиться вложение и извлечение информации. Также для должной надёжности необходим алгоритм выбора слова, которое будет нести в себе бит сообщения – в таком случае даже наличие словаря не даст шанса извлечь сообщение, так как любое из двух слов может обозначать 0 или 1. При этом

при наличии псевдослучайной последовательности вкладываемых бит атакующий не сможет определить, видит он перед собой факт вложения или нет. В теории при использовании блочного вложения с контрольной суммой можно сделать сообщение устойчивое к атаке замены/удаления одного или нескольких слов. Но в таком случае информация передастся не полностью и поднимется вопрос касательно целостности полученного сообщения.

Словарь состоящий из слов различных частей речи, учитывающий изменение формы слов по объёму может превосходить текст, который будет использоваться для сокрытия сообщения. При этом важно учитывать, что описанные особенности касаются только русского языка. Другие языки имеют совершенно иные правила написания осмысленного текста. В английском языке одно и то же слово при различных предлогах имеет совершенно разные значения, порой совершенно не похожие между собой. А потому перед использованием лингвистической стеганографии в том или ином языке важно подробно изучить вариативность и возможность замены слов.

В качестве потенциальных каналов передачи подобных сообщений могут служить:

- Новостные колонки;
- Сообщения на форумах;
- Обмен сообщениями в мессенджерах и в прочих средствах обмена информацией.

Возможно применение методов лингвистической стеганографии для создания водяных знаков в копиях различных, особенно переведённых с других языков, литературных и научных произведений. Большие объёмы текста позволят сделать множество копий стегосообщения, которое затем может быть обнаружено даже после частичной редакции исходного текста.

Создать программу для считывания сообщения не сложно – достаточно считать текст, выявить из него нужные синонимичные слова, перевести их в цифровую последовательность и уже исходя из методов цифрового кодирования извлечь сообщение. Гораздо сложнее реализовать автоматическую систему вложения сообщения, особенно без участия оператора – после вложения необходим анализ итогового текста, который компьютер провести не в состоянии, либо затрачиваемые ресурсы будут слишком велики и найдут куда более эффективное применение в прочих средствах стеганографии.

А потому программа должна быть ориентирована именно на облегчение работы оператора.

Для вложения информации в готовый текст необходимо обеспечить отсутствие его копий без стегосообщения – иначе атакующий сможет выявить наличие вложения без сторонних средств и в случае возможности – повредить его целостность или выявить отправителя сообщения. Впрочем, эта проблема поддаётся решению если в качестве канала передачи используется обмен сообщениями в мессенджере. Но в таком случае длительность передачи сообщения может быть значительной, а в случае низкой информативности и малом использовании слов – достаточно медленной. В этом случае носителем информации, в случае разумеющейся предварительной легенды – устоявшейся манеры общения, можно использовать неологизмы, междометия и даже разновидности смайликов – всё это может быть использовано как для передачи сообщения, так и для подачи сигналов определения есть ли в исходном сообщении вложение или нет.

Использованные источники:

1. Пушкин А.С. «Капитанская дочка» // «Современник» – 1836 – т. IV – С. 42-215.

2. Бабина, О.И. Лингвистическая стеганография: современные подходы. Часть 1 // Вестник ЮУрГУ. Серия «Лингвистика». – 2015. – Т. 12, № 3. – С. 27-33.

3. Словарь русских синонимов и сходных по смыслу выражений / Абрамов Н. // М. Русские словари – 1999.

4. Большаков И.А. Использование синонимов, ограниченных контекстными словосочетаниями, для целей лингвистической стеганографии // Информационные процессы и системы – 2004, №5 – С. 23-30.