ЛИНГВИСТИЧЕСКИЕ МЕТОДЫ СТЕГАНОГРАФИИ (НА МАТЕРИАЛАХ ПИСЬМЕННОГО РУССКОГО ЯЗЫКА)

А. В. Джунковский
Московский Государственный Лингвистический Университет
Москва

Защита конфиденциальных данных имела особое значение на протяжении всей истории человечества. Эта задача не поменялась и сегодня. На фоне постоянно эволюционирующих средств получения несанкционированного доступа, нахождение надежных методов сокрытия информации не теряет своей актуальности.

Традиционно специалисты выделяют два основных направления, способа сокрытия информации. Криптография осуществляет эту задачу путем использования шифра, с помощью которого осуществляется систематическое искажение текста. Для расшифровки необходимо использование ключа. Таким образом, в рамках криптографического подхода главной задачей является создание наиболее надежного шифра.

В свою очередь, стеганография осуществляет ту же задачу принципиально иными способами. Под стеганографией понимают совокупность методов сокрытия информации с целью достижения конфиденциальности путем маскировки самого факта сокрытия [1].

Термин «стеганография» в переводе с греческого означает «тайнопись». Хронологически он появился и был активно задействован первым, но в ходе исторического развития был вытеснен криптографией. С развитием методов компьютерной дешифровки методы стеганографии вновь становятся приоритетными ввиду их большей надежности [2].

В качестве теоретической базы исследования нами были использованы работы К. Шеннона по теории информации, А.П. Алферова по криптографии, а также широкая выборка работ по стеганографии и стеганализу.

Сегодня развитие компьютерных технологий сделало криптографические способы защиты информации значительно менее надежными, в связи с чем методы стеганографии вновь становятся востребованными [3].

Такая ситуация связана со стремительным развитием компьютерных технологий. В прошлом вскрытие шифров было сопряжено с искусным применением собственного интеллекта [4]. Сейчас это стало вопросом написания необходимых алгоритмов и использования доступных вычислительных мощностей. Компьютерные технологии сделали криптографию методикой замедления, но не предотвращения раскрытия информации [5].

В свою очередь, стеганография оказалась более востребована, поскольку попытка автоматизированного распознавания признаков стеганографического сокрытия информации требовала бы от заинтересованных лиц проведения анализа на объеме, де-факто, генеральной совокупности текстов автора [6]. Грамотное применение методов стеганографии позволяет не просто замедлить, но и предотвратить распознавание контейнера (текста, содержащего скрытую информацию) [7].

Сосредоточимся на исследовании письменных (печатных и рукописных) текстов русского языка. Такой выбор обусловлен тем, что именно в письменной речи риск потери конфиденциальности растянут во времени [8]. Иными словами, контейнер, или носитель скрытой информации, сохраняется чаще, чем в устной форме речи [9]. Повышенная опасность требует улучшенных методов защиты.

В России лингвистические методы стеганографии являются редко затрагиваемой темой. Если исследования стеганографических методов сокрытия в аудио- и видеофайлах, а также изображениях, встречаются с относительной периодичностью, то непосредственно лингвистические методы сокрытия за редкими исключениями не затрагиваются. Существует малое число российских работ, в которых рассматриваются лингвистические методы стеганографии на материалах устного русского языка, однако исследование на материалах письменной речи является научной лакуной.

Если говорить о зарубежных публикациях, то вопросы лингвистических методов стеганографии наиболее активно разрабатываются в Компьютерной лаборатории Кембриджского университета на материалах английского языка. Кроме того, такие исследования пользуются популярностью в Германии на материалах немецкого языка.

Первым отличием нашего исследования является его проведение на материалах русского языка. Кроме того, западные исследователи сосредотачивают свои усилия на противодействие компьютерному стеганализу, редко обращая внимание на возможности человека в данной сфере. Другой особенностью является последовательность проведения нашей работы. Большинство исследований в этой сфере проверяют эффективность заранее разработанных методов стеганографии, в то время как мы стремимся определить эти методы для выполнения практической задачи создания наиболее оптимальной методики лингвистической стеганографии для письменной формы русского языка.

Прикладная задача нашей работы — изучение методов стеганографии, не только невосприимчивых к компьютерному стеганализу, но и наиболее защищенных от такового, проводимого человеком.

В рамках исследования проводился эксперимент, состоящий из двух частей. Выборка — 50 испытуемых для обоих частей эксперимента. Участники эксперимента были осведомлены о том, что тексты могут содержать подобную информацию. Их задача — выявить эти тексты.

Первая часть эксперимента строилась на деликатной манипуляции естественно-языковыми печатными текстами. Ее цель — выявление плана языка, на котором подобная манипуляция наименее заметна для человеческого восприятия. В частности, испытуемым предлагался один случайным образом выбранный текст из заранее подготовленной выборки, в котором стеганографическими методами была внедрена скрытая информация. На морфемном уровне производилась манипуляция посредством внедрения эрративов. На лексическом уровне искажался естественный порядок слов, а также употреблялись наиболее подходящие в данном контексте синонимы. На синтаксическом уровне производится манипуляция средствами пунктуации. Искажение на фонемном, парадигматическом и синтагматическом уровнях не рассматриваются нами по причине особенностей письменной речи [10], а также выбранной формы эксперимента.

Вторая часть эксперимента отличалась еще более узкой специализацией. Так, в ней рассматривалась эффективность манипуляций с графемами. Было принято решение об исследовании таких искажений на материалах рукописного текста. Сравнивалась эффективность манипуляции с расположением отдельных графем и их совокупностей в тексте, с техникой переходов между отдельными графемами, а также искажением формы отдельных графем.

Результаты обеих частей эксперимента обрабатывались математически и статистически. Был сделан вывод о том, что искажения на уровне графем и морфем обладает наибольшим потенциалом, т.к. интерпретировались участниками эксперимента как случайные искажения.

В частности, мы выделяем 2 группы случаев, представляющих интерес. Для понимания важности предварительных результатов нашего исследования необходимо помнить, что материалы первой части эксперимента представляли собой печатные тексты. Первая группа представляет собой замену буквы на созвучную («непр**и**клонными» вместо «непреклонными») Вторая группа состоит из трех случаев — искажение слов на уровне графем посредством опущения букв слова («постоветского» вместо «пост**с**оветского»), замены буквы на соседнюю с точки зрения расположения на клавиатуре при стандартной раскладке QWERTY («простра**м**ство» вместо «пространство»), а также обратного порядка букв внутри слова («прадз**ник**» вместо «праздник»). Мы склонны интерпретировать результативность данных методов искажения следующим образом: первая группа строится на фонетическом созвучии и воспринимается читателем как признак субстандартного уровня грамотности автора. Вторая группа не воспринимается испытуемыми в качестве стеганографического контейнера по причине презумпции совершения автором ошибки печати.

Искажения на лексическом уровне привлекли внимание большинства участников эксперимента. Искажения на синтаксическом уровне дали смешанные результаты и требуют дальнейших исследований.

Проведение второй части эксперимента на материалах рукописного текста выявило необходимость доработки методов его проведения. Большинство испытуемых столкнулись с значительными трудностями при чтении рукописного текста, в следствие чего требуется его повторение после корректировки содержания материалов эксперимента.

В своем исследовании мы выдвинули две гипотезы. Первая — о том, что манипуляции с различными уровнями языка отличаются по своей заметности для читателя текста. Вторая гипотеза идентична первой по форме, но рассматривает заметность различных типов искажения графем. Первая гипотеза по предварительным результатам эксперимента подтвердилась.

В ходе исследования экспериментально было выявлена реальная эффективность различных лингвистических методов стеганографии, что может быть использовано для оптимизации надежности и безопасности связи [11]. На данный момент исследование продолжается, и мы намерены использовать результаты данного и последующих исследований в данной области для решения задачи оптимизации выбора методов стеганографии применительно к письменной форме русского языка

Мы надеемся использовать полученные статистические данные для разработки рекомендаций по наиболее эффектным лингвистическим методам стеганографии. Актуальность проводимого исследования предвосхищает свой пик [12]. С каждым годом развитие технологий автоматизированного криптоанализа делает методы стеганографии все более востребованным [13]. Парадигмальный сдвиг в сфере защиты информации становится в этих условиях вопросом времени.

ЛИТЕРАТУРА

1. Алферов А. П., Зубов А. Ю. и др. Основы криптографии. М.: Гелиос АРВ, 2012. 480 с.

2. Конахович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография. Теория и практика. Киев: МК-Пресс, 2006. 420 с.
3. Потапова Р. К. Речь: коммуникация, информация, кибернетика. М.: Изд-во УРСС, 2010. 600 с.
4. Шеннон К. Работы по теории информации и кибернетике. М.: Изд-во иностранной литературы, 1963. 830 с.
5. Baliga A., Kilian J. On Covert Collaboration. New York: Gate Publishing, 2007. 182 p.
6. Bennet K.Linguistic Steganography: Survey, analysis and robustness concerns for hiding information in text. Lafayette: Purdue University, 2004. 120 p.
7. Beutelspacher A. Kryptologie: Eine Einführung in die Wissenschaft vom Verschlüsseln, Verbergen und Verheimlichen. Berlin: Springer-Verlag, 2009. 155 S.
8. Johnson N.F. Steganalysis. Boston: Artech House, 2000. 314 p.
9. Katzenbeisser S. Principles of Steganography. Boston: Artech House, 2000. 186 p.
10. Kerckhoffs A. La Cryptographie Militaire. Paris, 1883. 163 p.
11. Provos N. Defending Against Statistical Steganalysis. Michigan: University of Michigan Press, 2001. 343 p.
12. Wayner P. Disappearing Cryptography: Information Hiding. San Francisco: Morgan Kaufmann, 2002. 221 p.
13. Wayner P. Strong Theoretical Steganography. – Berlin: Cryptologia, 2005. 410 p.