УДК 621.394.147 + 004.056.53

А. В. Еременко[1], канд. техн. наук, доцент каф. "Инфокоммуникационные системы и информационная безопасность", e-mail: nexus-@mail.ru,
А. Е. Сулавко[2], канд. техн. наук, ст. преподаватель каф. "Комплексная защита информации", e-mail: sulavich@mail.ru,
Е. В. Толкачева[1], канд. техн. наук, доцент каф. "Инфокоммуникационные системы и информационная безопасность", e-mail: tolkacheva_ev@mail.ru,
Е. А. Левитская[3], инженер-исследователь, e-mail: laska_kb@mail.ru
[1] Омский государственный университет путей сообщения (ОмГУПС (ОмИИТ), г. Омск,
[2] Омский государственный технический университет (ОмГТУ), г. Омск,
[3] ФГУП "Российский Федеральный Ядерный Центр — Всероссийский научно-исследовательский институт технической физики имени академика Е. И. Забабахина", г. Снежинск

Метод защиты текстовых документов на электронных и бумажных носителях на основе скрытого биометрического идентификатора субъекта, получаемого из подписи[1]

Рассмотрена проблема защиты авторского права (интеллектуальной собственности), возникающего при создании текстового произведения. Объектом исследования в работе выступают методы кодирования информации в текстовых контейнерах. Предложен метод встраивания в текстовый документ цифрового водяного знака, основанного на биометрических признаках автора документа. Определена информационная емкость разработанного метода. Разработан способ проверки документов на электронных и бумажных носителях на предмет неправомерного изменения и их аутентичности.

Ключевые слова: *биометрические признаки, защита интеллектуальной собственности, помехоустойчивое кодирование, цифровой водяной знак, стеганография, ключевая последовательность*

Введение

Развитие информационных сетей постепенно превращает мир в единое информационное и коммуникационное пространство, обостряя проблемы, связанные с защитой авторских прав. Помимо технической стороны вопроса защиты интеллектуальной собственности, актуальным остается вопрос законодательства. В настоящее время действие закона № 187-ФЗ распространяется только на фильмы, но от "пиратства" страдают и другие творческие деятели. По мнению генерального директора ЭКСМО, прозвучавшему на конференции "Право на Download 2013", за минувший год потери отрасли книгоиздания от интернет-пиратства составили 7,5 млрд руб. Согласно прогнозу к 2015 г. ущерб составит уже 30 млрд руб., а к 2018 г. достигнет значения 70 млрд руб. [1]. Основным подходом к техническому решению проблемы нелегального копирования является добавление в каждую распространяемую копию скрытого идентификатора (водяного знака), позволяющего однозначно определить источник утечки. Технологии встраивания цифрового идентификатора разрабатывали для электронных документов. Однако при печати текстовых документов возникает проблема переноса скрытого идентификатора автора на бумагу. Цель работы — определить потенциал подписи субъекта для создания на ее основе скрытого биометрического идентификатора и разработать метод встраивания идентификатора в текстовые документы на электронных и бумажных носителях для подтверждения их целостности и аутентичности.

1. Стеганографические методы защиты от нелегального копирования авторских произведений

На сегодняшний день существует большое число систем внедрения цифровых водяных знаков (ЦВЗ) в мультимедийную информацию. Менее проработанным является вопрос защиты текстовой информации с помощью внедрения ЦВЗ. В лингвистической стеганографии выделяют следующие направления: использование особенностей символов, кодирование смещением строк, кодирование смещением слов, синтаксические и семантические методы. Их достоинства и недостатки рассмотрены в работе [2]. В литературе [3, 4] также можно встретить описание синтаксических и семантических методов внедрения информации, однако отсутствует описание их адаптации для внедрения ЦВЗ.

Настоящая работа посвящена разработке технического метода защиты авторского права (интеллектуальной собственности), возникающего при создании текстового произведения. Исходя из постановки задачи, содержание текста после встраива-

[1] Работа выполнена при финансовой поддержке РФФИ (грант № 15-37-50366).

ния защитной метки не должно изменяться, так как должен быть сохранен авторской стиль произведения, в котором одинаковое значение имеют пунктуационные знаки и слова. Данное условие и требование сохранения возможности проверки защитной метки после вывода документа на печать накладывают ограничения на выбор стеганографического метода встраивания ЦВЗ. Аналитическое исследование проблемы показало, что только кодирование вертикальным смещением строк и горизонтальным смещением слов позволяет решить поставленную задачу, так как при таком подходе изменяется оформление документа, а не его содержательная часть. С учетом опыта [4] необходимо проработать вопрос генерации защитной метки и ее встраивания в различные участки текстового документа для обеспечения связи содержимого документа и идентификационных признаков подписи автора произведения.

В соответствии с предлагаемой технологией при завершении работы над текстом автор произведения (текстового документа) его подписывает. В качестве подписи может быть использован не только автограф, но и текстовый пароль. Из образа подписи автора формируется криптографический ключ. Далее на основе хеша текста и ключа генерируется биометрический идентификатор (защитная метка), который с помощью методов стеганографии связывается непосредственно с содержимым документа с возможностью его восстановления для проверки авторства и целостности текста после печати.

2. Информативные признаки подписи авторов текстовых произведений

Число признаков, необходимое для успешного решения некоторой задачи генерации ключа, зависит от их информативности. В рамках работы по гранту РФФИ 15-07-09053 "Исследование методов получения криптографических ключей шифрования из динамических биометрических характеристик пользователей компьютерных систем" авторами настоящей работы были апробированы различные виды ортогональных базисных функций для описания динамики подписи и оценена информативность признаков, получаемых на их основе.

Для создания эталона подписи пользователь несколько раз воспроизводит ее на графическом планшете. Из каждой реализации подписи вычисляются параметры, характеризующие внешний вид подписи — расстояния между координатами контурного образа, отношение длины к ширине, центр подписи, угол наклона, и ее динамику — коэффициенты корреляции между функциями динамики подписи, коэффициенты вейвлет-преобразования. После этого выполняется расчет средних значений указанных параметров по всем реализациям. Полученный вектор значений является эталоном подписи. Затем эталонные значения признаков округляют и представляют в виде последовательности m бит A_m. Описание процедуры округления будет представлено ниже.

3. Встраивание защитной метки в текст документа

Метод встраивания защитной метки в документ с помощью вертикального сдвига строк и горизонтального смещения слов описывается в работе [4]. В работе рассматриваются различные виды искажений в изображении документа (шум, перекос строк, растяжение и сжатие текста, размытость, случайный сдвиг и равномерное изменение интенсивности), возникающие при его ксерокопировании, сканировании и выводе на печать, и предложены способы компенсации данных искажений.

Изображение страницы может быть представлено в виде функции $f(x, y) = 0$ или 1, $x \in [0, W]$, $y \in [0, L]$, где W и L — ширина и длина документа, зависящие от разрешения сканирующего устройства. Для простоты примем, что $x \in R$, $y \in R$ и функция $f(x, y)$ принимает непрерывные значения. Изображение текстовой строки описывается функцией $f(x, y) = 0$ или 1, $x \in [0, W]$, $y \in [t, b]$, где t и b — верхняя и нижняя границы строки.

Горизонтальное представление строки на изображении можно описать интегральной величиной $h(y)$, т. е. суммой всех интервалов, где функция $f(x, y)$ принимает ненулевые значения $h(y) = \int_0^W f(x, y)dx$, $y \in [t, b]$. Вертикальная область текстовой строки может быть представлена суммой интервалов в вертикальном направлении, где функция $f(x, y)$ принимает ненулевые значения $v(x) = \int_t^b f(x, y)dy$, $x \in [0, W]$.

Для снижения ошибок декодирования при возникновении искажений в изображении документа предлагается кодировать строку только в том случае, если она и ближайшие соседние строки полностью заполнены. Один бит сообщения может быть закодирован в тексте через смещение строки вверх или вниз по вертикали от стандартной позиции, определяемой межстрочным интервалом, заданным оформителем документа. При этом соседние строки предлагается использовать как контрольные и их положение не должно изменяться. Также для кодирования битов можно использовать сдвиг слова по горизонтали влево или вправо. Для этого в строке выделяют три группы слов: два слова по краям строки и оставшиеся слова между ними, объединенные в одну среднюю группу. Слова по краям строки рассматривают как контрольные, и их положение не должно изменяться по завершении операции кодирования. Таким образом, для описания процедуры кодирования можно оперировать единой моделью представления текста в виде блоков (рис. 1): средний блок используется для ко-

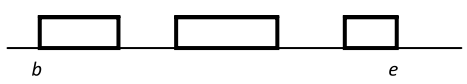

Рис. 1. Модель представления текста в виде блоков, использующихся для кодирования и декодирования битов сообщения

дирования бита сообщения, крайние блоки являются контрольными и их положение не изменяется по завершении операции кодирования.

Принципы кодирования защитной метки. Процедура кодирования битов защитной метки в документе выглядит следующим образом.

1. В отличие от работы [4], где для кодирования битов сообщения предложено использовать каждую вторую строку, соответствующую указанным выше условиям, в настоящей работе в целях увеличения емкости метода для кодирования выбирают строки, составляющие абзац, который является структурной единицей при форматировании документа (изменение межстрочного интервала применяют ко всем строкам, входящим в абзац). Положение первой и последней строк абзаца остается неизменным и используется для вычисления оригинального межстрочного интервала. При этом смещение выбранных для кодирования строк необходимо проводить в одном направлении применительно ко всему документу (либо вверх, либо вниз).

2. В отличие от работы [4], где кодирование бита сообщения предложено выполнять смещением центральной группы слов по горизонтали, в целях увеличения емкости метода в настоящей работе предлагается использовать все слова в строке за исключением первого и последнего, которые являются вспомогательными при определении оригинального расстояния между словами (до изменения их положения в строке). Смещение выбранных для кодирования слов также необходимо выполнять в одном направлении (вправо или влево). Смещение кодирует 1, отсутствие смещения — 0.

3. Защитную метку встраивают в каждую страницу документа. В зависимости от наполненности страницы текстом (емкости контейнера) проводят формирование защитной метки с числом бит 2^n с соблюдением условия $2^n \leqslant E$, где E — доступная для встраивания сообщения емкость контейнера в битах. При этом заполняется весь доступный объем контейнера. Зависимость надежности проверки защитной метки от числа n будет оценена в другом разделе настоящей работы.

4. Операция смещения строки является приоритетной перед операцией сдвига слова при кодировании очередного бита сообщения.

Декодирование (считывание) защитной метки из цифровой копии документа. После получения изображения документа в электронном виде проводим декодирование защитной метки. Рассмотрим обобщенную модель, которая описывает оба способа кодирования бит сообщения. Модель представляет собой область $h(x)$, $x \in [b, e]$ (см. рис. 1), состоящую из трех блоков. В случае если кодирование проводим смещением слов, область $h(x)$ состоит из трех блоков слов, первый и последний из которых являются контрольными и их положение не изменяется. Если речь идет о модели кодирования сдвигом строк, то $h(x)$ обозначает область, состоящую из трех блоков строк, первый и последний из которых являются контрольными и не изменяют своего положения по завершении операции кодирования.

Декодирование строк. Для определения смещения строк выполняют расчет центра тяжести каждого из трех блоков, входящих в область $h(x)$, определяемых соответственно тремя интервалами $[b_1, e_1]$, $[b_2, e_2]$ и $[b_3, e_3]$.

Центр тяжести блока определяем по формуле

$$c_i = \frac{\int_{b_i}^{e_i} x h(x)\,dx}{\int_{b_i}^{e_i} h(x)\,dx},$$

где i — номер блока, $i = 1, 2, 3$.

Достаточно определить ординату центра тяжести каждого блока для проверки факта смещения строки. На следующем шаге по ординатам центров тяжести y_1, y_2 и y_3 определяем расстояние между блоками. Признаком того, что строка сдвинута вверх, является выполнение неравенства $|y_2 - y_1| < |(y_3 - y_1)/2|$. Выполнение неравенства $|y_2 - y_1| > |(y_3 - y_1)/2|$ означает, что строка смещена вниз.

Декодирование слов. Пусть мы имеем область $h(x)$ и три интервала $[b_1, e_1]$, $[b_2, e_2]$ и $[b_3, e_3]$, которые определяют три группы слов. Полагаем, что $h(x) = 0$ в интервалах между группами слов. Среднюю группу используем для кодирования сообщения в тексте, в то время как положения крайних слов остаются неизменными.

Определим область, описывающую строку со сдвигом центрального блока влево, как

$$h^l(x) = \begin{cases} h(x),\ x < b_2 - \varepsilon \text{ или } x > e_2; \\ h(x + s),\ b_2 - \varepsilon \leqslant x \leqslant e_2 - \varepsilon; \\ 0,\ e_2 - \varepsilon \leqslant x < e_2. \end{cases}$$

После сдвига центрального блока вправо $h^r(x)$ описывает новую область, получившуюся из $h(x)$:

$$h^r(x) = \begin{cases} h(x),\ x < b_2 \text{ или } x > e_2 + \varepsilon; \\ 0,\ b_2 \leqslant x < b_2 + \varepsilon; \\ h(x - s),\ b_2 + \varepsilon \leqslant x \leqslant e_2 + \varepsilon, \end{cases}$$

где ε — значение сдвига, не превышающее половины расстояния между словами.

Обозначим $h'(x)$ область строки, полученную после оцифровывания документа. Для того чтобы определить направление смещения средней группы слов вычислим сумму:

$$\sum_{b_1}^{e_3} h'(x)(h^l(x) - h^r(x)) \geqslant 0.$$

Положительный либо равный нулю результат означает, что средняя группа слов была сдвинута влево, т.е. область $h'(x)$ в большей степени совпала с описанием $h^l(x)$, и сдвинута вправо в противном

случае. Независимо от примененного типа выравнивания абзаца можно определить положение каждого слова в строке относительно первого и последнего слов. На следующем шаге определяется фактическое положение слов в строке, после чего происходит декодирование содержимого.

Оценка информационной емкости метода и контейнера. Одним из вопросов, который необходимо исследовать в настоящей работе, является определение зависимости сложности криптографического ключа (его надежности) и емкости текста, в который данный ключ может быть встроен.

При разработке системы защиты текстовых документов на электронных и бумажных носителях необходимо учитывать емкость контейнера. Поэтому первым этапом в процессе стеганографии является выбор файла, который необходимо скрыть. Его еще называют информационным файлом. В сопроводительной документации к большинству известных программ по стеганографии говорится, что для сокрытия информации объем памяти файла-контейнера должен примерно в 8 раз превышать объем памяти информационного файла. Следовательно, чтобы спрятать файл размером 710 Кбайт, понадобится графический файл объемом 5680 Кбайт. Но если рассмотреть методы сокрытия информации в тексте, то становится очевидным, что объем памяти файла-контейнера должен в 50—200 раз превышать объем памяти информационного файла [5].

В то же время эффективность методов встраивания дополнительной информации в скрывающие данные (контейнеры) в первую очередь определяется информационной емкостью метода. Приведем несколько интересных оценок лингвистов. Средняя длина слова на корпусе текстов частотного словаря О. Н. Ляшевской и С. А. Шарова составляет 5,28 символа, а на корпусе частотного словаря Л. Н. Засориной — 5,4 символа. Средняя длина предложения в русском языке составляет 10,38 слов [6]. Размер условно-стандартной (учетной) страницы формата А4 равен 1800 знакам с пробелами [7]. В таблице представлен расчет информационной емкости контейнера для страницы формата А4. Нижняя граница оценки получена на основании данных из открытой литературы, верхняя граница — по результатам собственных измерений типовой страницы, плотно заполненной текстом. Число слов в первой строке таблицы получено следующим образом. Средняя длина слова в русском языке 5,28 символа была округлена в большую сторону до 6. Длина слова с учетом пробелов с обеих сторон составляет 8 символов, тогда число слов в документе равняется 225 (частное от деления 1800 на 8). За вычетом крайних слов в строках ($2 \times 35 = 70$), которые нельзя использовать для кодирования битов сообщения, остается 155 слов, пригодных для внедрения сообщения.

Оценка информационной емкости контейнера

Источник данных	Число знаков с пробелами	Число строк	Число слов на странице А4, пригодных для кодирования	Информационная емкость, %
Открытый источник	1800	35	155	1
Собственные измерения	2690	35	267	1,5

С учетом того, что на каждую строку абзаца и на каждое слово, встреченное в текстовой строке, приходится один бит скрываемого сообщения, средняя информационная емкость предложенного метода невысока. В соотношении количества скрываемой информации к объему контейнера (скрывающего текста) она составляет 1...1,5 %. Применительно к текстовой стеганографии информационная емкость зачастую оценивается не прямым соотношением количества скрываемой информации к исходному объему контейнера, а отношением бит/символ. Другими словами, оценивается среднее число битов, приходящихся на один символ текстового сообщения. Такая оценка продиктована прежде всего тем, что минимальной единицей текстового сообщения является символ, а не отдельный бит. Информационная емкость метода в отношении бит/символ составляет 0,02, что соответствует эффективности использования контейнера на 2 % (1 бит на 48 бит 6-символьного слова). Для сравнения — эффективность использования контейнера методом, основанным на использовании синонимов, при искусственной генерации текстов не превышает 11 % в отношении бит/символ. Примерно ту же эффективность имеют и методы, основанные на использовании пробельных символов и знаков пунктуации.

4. Генерация защитной метки и проверка авторства документа

Чтобы определиться со структурой внедряемой маркировки, необходимо понять, какие требования к ней предъявляются. Для этого следует определить, что включает в себя задача защиты электронного документа. Были выделены следующие части данной задачи:

- защита документа и его частей от подделки;
- проверка авторства документа.

Для решения этих задач было решено внедрять в документы следующую информацию:

- идентификатор автора текстового документа, связанный с биометрическими характеристиками его подписи;
- значение хеш-функции от содержимого документа.

Из этих требований и из описанного выше способа сокрытия информации было принято решение о целесообразности внедрения маркировки в каждую страницу текстового документа. Кроме того, ввиду необходимости проверки наличия изменений в документе, следует привязывать маркировку

не только к ключу, получаемому из биометрических характеристик подписи автора документа, но и к самому контейнеру.

После подготовки произведения автор подписывает документ. Затем необходимо вычислить статические и динамические признаки подписи автора документа, провести оценку информативности каждого признака. Выработку защитной метки проводят постранично, для этого текст с каждой страницы загружают в буфер. С учетом требования распределения данных скрытой маркировки по всему документу, а также необходимости обеспечения возможности проверки маркировки по каждой малой части документа отдельно, алгоритм встраивания маркировки должен включать в себя следующие основные шаги.

1. Провести оценку информационной емкости текстового контейнера описанным ранее способом.

2. С помощью псевдослучайного генератора выработать ключевую последовательность *Key*, равную *m* бит. Биометрический идентификатор автора не может быть непосредственно использован для шифрования текстовой строки, так как значения биометрических признаков даже при последовательном вводе нескольких реализаций будут иметь определенный разброс. Ключевую последовательность необходимо кодировать помехоустойчивым кодом, после чего с помощью операции побитового сложения по модулю 2 с биометрическими данными автора получают открытую строку. Ключ после защиты документа уничтожить. Открытую строку, описание произведения и его автора сохранить на специализированном сервере. Связь криптостойкости ключа длиной *m* бит с надежностью его восстановления из открытой строки, хранящейся на сервере (чем более длинный ключ используется, тем больше малоинформативных признаков применяется для его защиты, следовательно, тем больше ошибки 1-го и 2-го рода при его восстановлении) будет показана ниже.

3. Для получения биометрического идентификатора (защитной метки) используют *n*-битовую хеш-функцию двух аргументов. Логическая структура рекомендуемой хеш-функции представлена на рис. 2. На вход хеш-функции подать выбранную для маркировки текстовую строку и сгенерированную ключевую последовательность. На первом шаге текстовую строку и ключевую последовательность по-

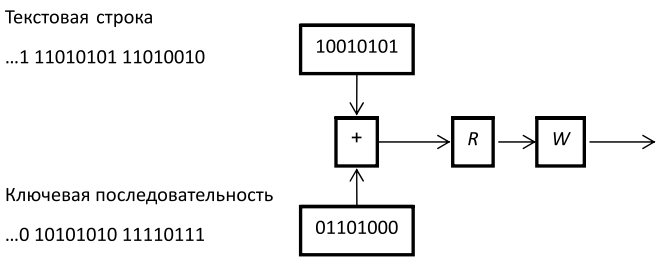

Рис. 2. Вычисление однобитовой хеш-функции от заданных текстовой строки и ключевой последовательности

дают на вход байтового сумматора хеш-функции в байтовом представлении. Байтовый сумматор может быть представлен функцией битового сложения по модулю два (операция XOR) при достаточной длине ключевой последовательности или функцией поточного шифрования текстовой строки на ключе *Key*. В реализации байтового сумматора с использованием функции битового сложения по модулю два каждый бит байта текстовой последовательности складывается по модулю два с соответствующим битом ключевой последовательности. С выхода байтового сумматора биты последовательно подают на вход регистра сдвига $R \rightarrow$ с линейной обратной связью. Выбор конкретного примитивного многочлена для регистра сдвига может быть произвольным, в нашем случае выбран многочлен $x^8 + x^4 + x^3 + x^2 + 1$.

Биты с выхода регистра сдвига подают во временный буфер *W*, реализующий хранение последних 2^n поступивших в него битов (исходя из условия $2^n \leqslant E$, где 2^n — длина ключа; E — информационная емкость контейнера). По завершении циклической обработки всех битов последовательности, полученной от исходной текстовой строки с выхода байтового сумматора, последние 2^n битов с выхода регистра сдвига $R \rightarrow$, оставшиеся в буфере *W*, и формируют результат хеш-функции. Альтернативно можно использовать любую из широко распространенных, проверенных хеш-функций одного аргумента, например функцию SHA-256, подавая ей на вход результат гаммирования двух аргументов.

Для проверки авторства текста субъект, заявляющий данное право, расписывается на графическом планшете. Полученные на основании его подписи биометрические признаки используют для извлечения криптографического ключа из открытой строки (с помощью операции сложения по модулю 2 и последующего применения кода, исправляющего ошибки), хранящейся на сервере. Вычисляют значение хеш-функции от текстовой строки и полученного криптографического ключа способом, описанным выше. Выполняют декодирование защитной метки, содержащейся в тексте и ее сравнение с вычисленной меткой на основании предъявленных данных. При совпадении защитных меток авторство считается подтвержденным, а содержание документа неизменным.

Для определения связи криптостойкости ключа длиной *m* бит и надежности его восстановления из открытой строки, хранящейся на сервере, был проведен вычислительный эксперимент с использованием предложенного способа восстановления ключа, который заключается в следующем. Предварительно исходные значения признаков кодируют с помощью преобразования вида: $y \in f(x)$, где $x \in X$, X — множество возможных значений признака, $y \in \{0, 1, 3, 7, 15, 31, 63, 127, 255, 254, 252, 248, 240, 224, 192, 128\}$. Выходное значение y представлено в двоичном виде для того, чтобы исклю-

чить неинформативные биты данных. Далее проводят "склейку" битовых последовательностей в одну результирующую, которую "объединяют" со случайной строкой. В работе [8] показана связь эффективности коррекции ошибок с методами группирования битов с разной вероятностью единичной ошибки. Несмотря на предпринятые в данном направлении усилия, единого подхода для решения этого вопроса до сих пор выработано не было. В настоящей работе предпринята попытка развития данного направления по модернизации нечетких экстракторов. Последовательность признаков задается случайным образом для каждого субъекта индивидуально во время формирования открытой строки. Осуществляется конкатенация битовых представлений только наиболее информативных (стабильных) признаков для субъекта. Для каждого признака по всем отобранным для создания эталона преобразованным реализациям вычисляют относительную частоту появления единичных (или нулевых) бит. Далее определяют интегральную вероятность появления единичного бита во всех разрядах, при этом относительную частоту берут как вероятность. Чем больше разрядов будут иметь частоты, близкие к 0 или 1, тем меньше получится итоговое произведение и тем выше интегральная оценка стабильности (информативности) признака для субъекта. Несложно заметить, что при подсчете частот появления нулевых бит (вместо единичных), значение производящей функции будет тем же. Далее все признаки ранжируют по информативности (изменяется их порядок — от самого информативного к самому малоинформативному) и отбирают определенное число признаков, проводят конкатенацию битовых представлений этих признаков, остальные отбрасывают. Чем больше использовано признаков, тем выше длина ключа, но и выше сумма ошибок 1-го и 2-го рода. Итоговую битовую последовательность используют для формирования открытой строки.

Для кодирования битовой последовательности *Key* в предложенном методе использовали коды, исправляющие ошибки Боуза — Чоудхури — Хоквингема (БЧХ), для декодирования — алгоритм Питерсона — Горенстейна — Цирлера (ПГЦ), по аналогии с [9].

Результат эксперимента показан на рис. 3 и должен быть учтен для определения компромисса между надежностью восстановления ключа и его криптостойкостью. Он составляет: вероятности ошибок выработки идентификатора 1-го рода 0,008...0,064 и 2-го рода 0,03...0,01 при длине криптографического ключа от 64 до 328 бит соответственно. Полученный результат указан для оптимальной исправляющей способности кода (при которой сумма ошибок 1-го и 2-го рода восстановления ключа была наименьшей при заданном числе признаков), вычисляемой в процессе эксперимента. Достоверность указанных на рис. 3 результатов со-

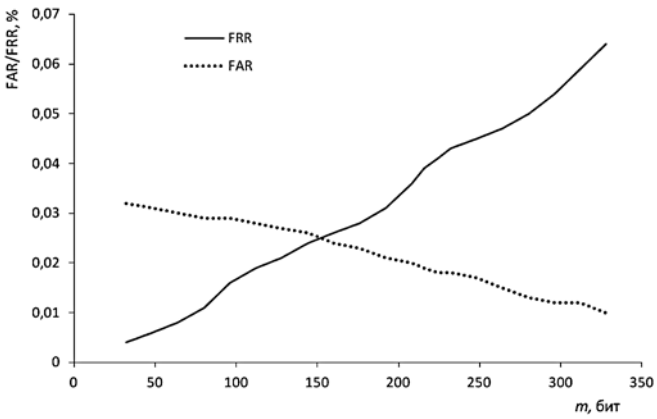

Рис. 3. Связь криптостойкости ключа длиной *m* бит и надежности его восстановления из открытой строки, хранящейся на сервере

ставляет более 0,98 при доверительном интервале вероятности ошибок 1-го (FRR) и 2-го рода (FAR) 0,01 и 0,002 соответственно, что определялось исходя из экспериментально полученных вероятностей и числа проведенных опытов (8000).

Перед встраиванием защитной метки в текст документа для компенсации возможных ошибок на этапе ее извлечения, связанных с искажением изображения, целесообразно выполнять помехоустойчивое кодирование защитной метки. Следует учитывать, что данная операция увеличит размер защитной метки в несколько раз, что повлечет за собой необходимость использования хеш-функции, возвращающей значение меньшей разрядности ввиду ограничения информационной емкости контейнера. Если объема текста на странице недостаточно для внедрения значения хеш-функции, то он может быть "присоединен" к содержанию предыдущей или последующей страницы.

Заключение

Разработан метод генерации идентификатора автора на основе содержимого документа и биометрических характеристик подписи его владельца с вероятностью ошибок выработки идентификатора 1-го рода от 0,008 до 0,064 и 2-го рода от 0,03 до 0,01 при длине криптографического ключа от 64 до 328 бит соответственно. В качестве способа встраивания защитной метки в документ с учетом поставленной задачи выбрано кодирование битов сообщения смещением слов и строк. Способ обеспечивает скрытность и устойчивость встраиваемого сообщения для определения целостности и аутентичности документа как на электронном, так и на бумажном носителях с соотношением количества скрываемой информации к объему контейнера 1...1,5 %. Информационная емкость предложенного метода достаточна для решения поставленной в работе задачи — защиты документа от незаконного копирования и изменения содержимого. Защитную метку можно внедрять непосредственно в сам до-

кумент, поддерживающий форматирование текста, либо изменения в формате выполнять непосредственно при выводе документа на печать.

Для подавления шумов, возникающих при получении цифровой копии документа, перед выполнением процедуры декодирования сообщения следует применять известные методики, описанные в открытой литературе, например эффективным способом подавления шума типа "*salt and pepper*" является медианный фильтр [10].

Список литературы

1. **Горовцова М.** Новая "антипиратская" инициатива: окажутся ли под защитой все объекты авторских прав ГАРАНТ.РУ. URL: http://www.garant.ru/article/501059/#ixzz3PHwgpwoV. (дата обращения: 15.01.2015).
2. **Балакин А. В., Елисеев А. С., Гуфан А. Ю.** Использование стеганографических методов для защиты текстовой информации // T-Comm. Спецвыпуск. 2009, апрель. С. 42—50.
3. **Аграновский А. В., Балакин А. В., Хади Р. А.** Запатентованные методы стеганографии в технологиях цифровых водяных знаков // Информационные технологии. 2002. № 9. С. 2—7.
4. **Brassil J., Low S., Maxemchuk N., O'Gorman L.** Document marking and identification using both line and word shifting // Technical report, AT & T Bell Laboratories, 1994. P. 853—860.
5. **Компьютерная** стеганография — защита информации или инструмент преступления? [Электронный ресурс]. URL: http://www.crime-research.org/library/Steganos.htm, свободный (дата обращения: 20.12.2011).
6. **Куянов Ю. В., Тришин В. Н.** Количественный анализ Большого русского словаря-справочника синонимов // Научное обозрение: гуманитарные исследования. 2015. № 9. С. 105—111.
7. **Дупленский Н.** Письменный перевод. Рекомендации переводчику и заказчику / под ред. Е. Масловского (СПР). 3 ред. М.: Союз переводчиков России, 2015.
8. **Santos M. F., Aguilar J. F., Garcia J. O.** Cryptographic key generation using handwritten signature // Proc. of SPIE, Orlando, Fla, USA, Apr. 2006. Vol. 6202. P. 225—231.
9. **Еременко А. В., Сулавко А. Е.** Исследование алгоритма генерации криптографических ключей из биометрической информации пользователей компьютерных систем // Информационные технологии. 2013. № 11. С. 47—51.
10. **Сато Ю.** Без паники! Цифровая обработка сигналов. М.: Додэка XXI, 2010. 176 с.

A. V. Eremenko[1], Ph. D., Associate Professor, nexus-@mail.ru,
A. E. Sulavko[2], Ph. D., Assistant, sulavich@mail.ru,
E. V. Tolkacheva[1], Ph. D., Associate Professor, tolkacheva_ev@mail.ru,
E. A. Levitskaya[3], Research Engineer, laska_kb@mail.ru,
[1]Omsk State Transport University (OSTU), [2]Omsk State Technical University (OmSTU),
[3]Federal State Unitary Enterprise "Russian Federal Nuclear Center — Academician E. I. Zababakhin All-Russian Research Institute of Technical Physics" (FSUE "RFNC-VNIITF")

The Method of Protection of Electronic and Paper-Based Text Documents with Biometric Identifier of the Subject Obtained from his Signature

The problem of the protection of copyright (intellectual property) that occurs when creating a text works is considered. The object of study in the article are the methods of coding information in the text containers. A method of embedding into a text document a digital watermark based on the biometric features of the author of the document is offered. The informational capacity of this method is determined. A method for inspection of documents in electronic and paper form for unauthorized changes and their authenticity is designed.

Keywords: *biometric features, intellectual property protection, error-correcting codes, digital watermark, steganography, the key sequence*

References

1. **Gorovcova M.** *Novaja "antipiratskaja" iniciativa: okazhutsja li pod zashhitoj vse ob'ekty avtorskih prav GARANT.RU* [The new "antipiracy" Initiative: will be all objects under the protection of copyright GARANT.RU], Access: http://www.garant.ru/article/501059/#ixzz3PHwgpwoV. (Date of circulation: 01.15.2015).
2. **Balakin A. V., Eliseev A. S., Gufan A. Ju.** Ispol'zovanie steganograficheskih metodov dlja zashhity tekstovoj informacii [Using steganography techniques to protect the textual information], *T-Comm*. april 2009, pp. 42—50.
3. **Agranovskiy A. V., Balakin A. V., Hadi R. A.** Zapatentovannye metody steganografii v tehnologijah cifrovyh vodjanyh znakov [The patented techniques in the steganographic digital watermark technology], *Information technologies*, 2002, no. 9, pp. 2—7.
4. **Brassil J., Low S., Maxemchuk N., O'Gorman L.** Document marking and identification using both line and word shifting, *Technical report, AT & T Bell Laboratories*, 1994, pp. 853—860.
5. **Komp'juternaja** *steganografija — zashhita informacii ili instrument prestuplenija?* [Computer steganography — information security or instrument of the crime?]. Access: http://www.crime-research.org/library/Steganos.htm, free (date of circulation: 20.12.2011).
6. **Kuyan J. V., Trishin V. N.** Kolichestvennyj analiz Bol'shogo russkogo slovarja-spravochnika sinonimov [Quantitative analysis of large-Russian dictionary of synonyms directory], *Scientific Review: humanities research*, 2015, no. 9, pp. 105—111.
7. **Duplensky N.** *Pis'mennyj perevod. Rekomendacii perevodchiku i zakazchiku* [Translation. Recommendations to the translator and the customer], editor E. Maslowski, 3 red. Moscow, Union of Translators of Russia, 2015.
8. **Santos M. F., Aguilar J. F., Garcia J. O.** Cryptographic key generation using handwritten signature, *Proceedings of SPIE, Orlando, Fla, USA, Apr. 2006,* 2006, vol. 6202, pp. 225—231.
9. **Eremenko A. V., Sulavko A. E.** Issledovanie algoritma generacii kriptograficheskih kljuchej iz biometricheskoj informacii pol'zovatelej komp'juternyh sistem [Investigation of algorithm for generating cryptographic keys from biometric information of users of computer systems], *Informacionnye tehnologii*, 2013, no. 11, pp. 47—51.
10. **Sato Yu.** Bez paniki! Cifrovaja obrabotka signalov [Do not panic! Digital signal processing], Moscow, Dodeka XXI, 2010, 176 p.