УДК 004.056.5
DOI: 10.25559/SITITO.18.202201.152-166

Научная статья

Метод защиты информации цифровых документов с помощью невидимых цифровых меток и его реализация

К. С. Гуртова
ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова», г. Москва, Российская Федерация
Адрес: 119991, Российская Федерация, г. Москва, ГСП-1, Ленинские горы, д. 1
kristinagurtov@yandex.ru

Аннотация
В настоящее время многие цифровые данные, которые передаются через Интернет, часто подвергаются атакам злоумышленников, что приводит к утечке информации и создает серьезные проблемы в области защиты авторского права, защиты права собственности, аутентификации и т.д. В последние годы большое внимание пользователей и исследователей привлекла технология цифровых водяных меток для применения в задачах защиты контента. Особенно требовательной областью цифрового маркирования являются методы маркирования документов, которые очень чувствительны к любым изменениям текста. В данной статье рассматриваются текущие тенденции в области технологий нанесения и извлечения водяных меток на цифровые документы, чтобы определить самые современные методы и их ограничения. Также разрабатывается общая архитектура алгоритма нанесения и алгоритма извлечения надежных и незаметных водяных меток в документ, основанная на изменении глифов текста, для решения проблемы отслеживания источника утечки информации. Применяя такой алгоритм, мы можем извлечь информацию о водяных метках из скриншотов документа. По сравнению с предыдущими алгоритмами нанесения водяных меток в документы, предлагаемая схема гарантирует независимое от контента встраивание, а также невидимость цифровой метки. Кроме того, предлагаемая схема маркирования показывает высокую точность извлечения.

Ключевые слова: маркирование документов, водяная метка, глиф, невидимость, поиск утечек

Автор заявляет об отсутствии конфликта интересов.

Для цитирования: Гуртова К. С. Метод защиты информации цифровых документов с помощью невидимых цифровых меток и его реализация // Современные информационные технологии и ИТ-образование. 2022. Т. 18, № 1. С. 152-166. doi: https://doi.org/10.25559/SITITO.18.202201.152-166

Original article

Method for Information Protection of Digital Documents Using Invisible Digital Watermarks and Its Implementation

K. S. Gurtova
Lomonosov Moscow State University, Moscow, Russian Federation
Address: 1 Leninskie gory, Moscow 119991, GSP-1, Russian Federation
kristinagurtov@yandex.ru

Abstract

Currently, many digital data transmitted over the Internet are often attacked by intruders. This leads to information leakage and creates serious problems in the field of copyright protection, property rights protection, authentication, etc. In recent years, the technology of digital watermarks for application in content protection problems has attracted great attention from users and researchers. A particularly demanding area of digital marking is the marking of documents, which are very sensitive to any changes of text. This article reviews the current trends in watermarking and watermarking technologies on digital documents to identify state-of-the-art techniques and their limitations. Also, a general architecture of the algorithm for applying and extracting reliable and imperceptible watermarks into a document, based on changing text glyphs, is being developed to solve the problem of tracking the source of information leakage. Using this algorithm, we can extract watermark information from document screenshots. Compared to previous algorithms for watermarking documents, the proposed scheme guarantees content-independent embedding, as well as the invisibility of the digital mark. In addition, the proposed marking scheme shows a high extraction accuracy.

Keywords: document marking, watermark, glyph, invisibility, leak detection

The author declares no conflict of interest.

For citation: Gurtova K.S. Method for Information Protection of Digital Documents Using Invisible Digital Watermarks and Its Implementation. *Sovremennye informacionnye tehnologii i IT-obrazovanie = Modern Information Technologies and IT-Education.* 2022; 18(1):152-166. doi: https://doi.org/10.25559/SITITO.18.202201.152-166

Введение

Развитие цифровых вычислений и сетевых технологий открыло новые возможности для обмена информацией через интернет-сервисы. Сегодня доступ в сеть Интернет означает доступ к огромному количеству научных статей, обучающих курсов, а также возможность общения через сообщения, видео- и аудиосвязь. Однако, передача, хранение и совместное использование данных по незащищенным сетям связи требует высокого уровня безопасности и конфиденциальности. С ростом технологий появились такие проблемы, как нарушение авторского права, открытие врачебной тайны и другие ситуации нелегального копирования и распространения закрытой информации.

Традиционно для обеспечения безопасности цифровой информации использовались различные методы, такие как криптография, стеганография и их комбинационные подходы[1], но все они имеют различные ограничения, зависящие от типа приложения, в котором используются и модифицируются цифровые данные. Чтобы решить проблему традиционных методов, исследователи разработали концепцию цифровых подписей и цифрового маркирования[2], которые повышают безопасность, обеспечивая целостность и конфиденциальность цифровых данных, и защищают содержимое от несанкционированного доступа. Методы цифровой подписи и нанесения водяных меток очень похожи друг на друга. Цифровая подпись используется для проверки подлинности содержимого цифровых данных. Однако цифровая подпись имеет существенное ограничение – она может идентифицировать изменения, внесенные в цифровые данные, но не может найти область, в которой данные были изменены. Поэтому в некоторых ситуациях используется технология цифровых водяных меток для предоставления некоторых дополнительных функций, которые преодолевают ограничения и проблемы метода цифровой подписи.

Области применения цифрового маркирования

Цифровые метки могут широко применяться во многих сферах нашей жизни. В таблице 1 приведены несколько основных областей, которые используют или могут использовать цифровое маркирование для защиты данных.

Таблица 1. **Области применения цифрового маркирования**
Table 1. **Areas of application of digital marking**

Область применения	**Описание**
Защита авторских прав и идентификация владельца	Водяные метки чаще всего используются для защиты авторских прав. Информация, скрытая в метке, содержит информацию о владельце данных
Поиск утечки данных	Внедрение информации о владельце копии данных также позволяет выявить источник незаконного копирования и распространения информации
Защищенная передача данных	Цифровые метки могут быть использованы для передачи сообщений таким образом, чтобы оно не могло быть обнаружено и прочитано посторонним лицом
Целостность данных	Ненадежные цифровые метки позволяют обнаруживать изменения в данных
Врачебная тайна	Метки позволяют повысить безопасность имени пациента и данных о его болезнях, записанные в электронных медицинских картах

Методы цифрового маркирования

A. Цифровое маркирование

Бумажные водяные метки были изобретены еще в 13 веке в Италии для идентификации работ мастеров, которые создавали бумагу для картин знаменитых художников. Далее, водяные метки также стали наносить для проверки подлинности банкнот, а также различных банковских бумаг. Концепция продолжала развиваться в цифровую эпоху, чтобы обеспечить аутентификацию и защиту цифровых носителей.

Сегодня цифровое маркирование – это мощный инструмент для внедрения различного типа данных в другие данные [1]. Все методы цифрового маркирования обязательно определяются двумя основными алгоритмами – это алгоритм внедрения метки в данные и алгоритм извлечения метки из помеченных данных. Методы внедрения и извлечения меток широко изучены и существует большое множество классификаций этих методов, которые позволяют определить нужный метод для поставленной задачи.

B. Требования к методам цифрового маркирования

Несмотря на различия между методами нанесения и извлечения водяных меток, которые могут применяться в совершенно разных сценариях, существуют требования, которым должна удовлетворять любая система цифрового маркирования.

1) Надежность

Под требованием надежности метода цифрового маркирования имеется в виду возможность восстановления водяной метки из поврежденной версии данных. Обычно это означает, что метка может выдерживать, наиболее распространенные изменения и искажения, которым могут подвергаться данные в течение их жизненного цикла.

[1] Mohanty S. P. Digital Watermarking: A Tutorial Review. Report on Dept. of Electrical Engineering, Indian Institute of Science. Bangalore, India, 1999. p. 1-24.

[2] Lin C.-Y. Watermarking and Digital Signature Techniques for Multimedia Authentication and Copyright Protection: Thesis of Doctor of Philosophy. Columbia University, 2000. 238 p.

2) Точность

Требование точности системы цифрового маркирования подразумевает, что наличие водяной метки не должно ухудшать качество данных с внедренной меткой.

3) Емкость

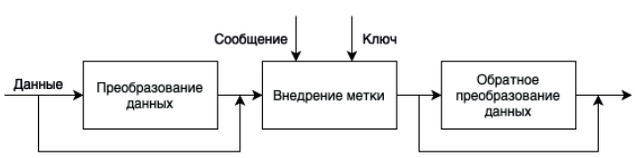
Емкость или полезная нагрузка определяется как количество бит, из которых состоит сообщение, внедряемого с водяной меткой.

Большая полезная нагрузка является желательной особенностью алгоритма водяных меток, но увеличение полезной нагрузки напрямую противоречит требованию надежности метки. Самый простой способ повысить надежность, соблюдая при этом критерий точности, состоит в том, чтобы распространить водяную метку на большое количество выборок данных (например, на несколько областей изображения), что быстро исчерпывает пространство, доступное для встраивания водяных меток. В качестве альтернативы, кодирование с исправлением ошибок может быть использовано поверх алгоритма водяных меток, что позволит восстановить некоторые биты, возможно, потерянные из-за манипуляций с данными. Однако даже в этом случае, пространство, доступное для размещения сообщения, уменьшается.

Исходя из этих рассуждений, требование к емкости цифровых меток вступает в противоречие с двумя другими требованиями. Поэтому любая система цифрового маркирования способна удовлетворять только двум требованиям и при проектировании такой системы нужно учитывать особенности разрабатываемого приложения, а также сценарии его использования.

C. Классические методы цифрового маркирования

Первая большая категория методов для внедрения и извлечения водяных меток – это классические методы, которые имеют богатую историю создания. Алгоритм внедрения водяных меток в классической модели цифрового маркирования, изображен на рисунке 1. Первым шагом, данные, в которые требуется внести водяную метку, могут быть преобразованы в необходимую для внедрения форму. После того, как данные находятся в нужном виде, водяная метка вводится в документ с возможным использованием ключа *K*. Классические методы часто используют ключ, как дополнительный параметр алгоритма, чтобы избежать получения информации из цифровой метки неавторизованными пользователям, то есть пользователям, которые не знают значения ключа. Затем содержимое с водяной меткой возвращается в исходный вид, чтобы получить данные с цифровой меткой и не потерять изначальное содержимое.



Р и с. 1. Основные шаги классических методов внедрения цифровых меток
F i g. 1. The main steps of the classical methods of digital labels implementing

1) Однобитовые и мультибитовые метки

С точки зрения количества встраиваемой информации в метку, классические методы цифрового маркирования могут быть разделены на мультибитовые и однобитовые [2], [3]. В однобитовых схемах результатом алгоритма извлечения метки является простое заключение, была ли в данных цифровая метка. Тем не менее, чаще всего требуется внедрение большего количества информации и в таких случаях используется мультибитовая цифровая метка, при извлечении которой можно получить дополнительную информацию, например, о табельном номере сотрудника или IP-адресе компьютера, на котором были сохранены данные.

2) Надежные и ненадежные метки

Еще одна классификация делит водяные метки на ненадежные, надежные и гибридные [4], [5]. Ненадежные водяные метки очень чувствительны к изменениям в данных и могут быть легко удалены, поэтому используются в основном для обнаружения несанкционированного доступа, в то время как надежные водяные метки сохраняют встроенную информацию при атаках, изменяющих данные. Водяная метка со смешанными свойствами называется гибридной водяной меткой.

3) Метки, зависимые и независимые от исходных данных

Методы получения цифровых меток можно также разделить на три основные категории, а именно: независимые, слабо зависимые и зависимые от исходных данных [6], [7], [8]. Для зависимых водяных меток во время их встраивания и извлечения требуются как исходные данные, так и данные, записанные внутри метки. При этом при независимом нанесении метки ни исходные данные, ни информация о метке не требуются для ее извлечения. Слабо зависимая метка требует только знание об исходных данных для корректного получения метки из данных. Таким образом, зависимые метки легче внедрять и извлекать, но в большинстве случаев данные, использованные для внедрения метки, недоступны или слишком разнообразны.

4) Заметные и незаметные метки

Заметные водяные метки используются для встраивания данных, которые должны быть видимыми или слышимыми [9]. Например, такими метками могут быть QR-коды, встраиваемые в документы или водяные знаки, содержащие логотип владельца, помещаемые на изображения. Существует два важных условия, которым должна удовлетворять хорошая заметная цифровая метка. Во-первых, такую метку должно быть трудно удалить, а также она должна быть устойчивой к фальсификации. Поскольку встраивать шаблон или логотип в данные относительно легко, необходима проверка, что заметная метка действительно была вставлена корректным пользователем.

Незаметные цифровые метки встраивают данные, которые должны быть невидимыми или неслышимыми, но которые могут быть извлечены компьютером [10].

D. Методы цифрового маркирования документов

Еще одной широкой классификацией методов цифрового маркирования является классификация по типу данных, в которые внедряется метка. Самыми распространенными данными для маркировки являются текстовые документы, изображения и видео [11].

Текст, будучи самым простым способом коммуникации и обмена информацией, создает различные проблемы, когда речь заходит о защите авторских прав и аутентификации. Любые изменения в тексте должны сохранять ценность, полезность, смысл и грамматику текста. Короткие документы сложнее защитить и аутентифицировать, так как простой анализ легко выявит водяную метку, что сделает текст небезопасным.

Помимо требований надежности и невидимости, водяные метки в документах должны быть независимы от содержимого его текста. Поскольку в реальном сценарии водяные метки различны, но документы одинаковы, мы не можем полагаться на содержимое документа для встраивания водяной метки. Кроме того, алгоритмы, независимые от исходных данных, требуют меньшего времени встраивания, чем алгоритмы, связанные с контентом.

Традиционные алгоритмы внедрения водяных меток в документы, как правило, можно разделить на три категории: схемы, использующие семантику текста, схемы, использующие структуру текста и схемы, использующие изображение текста.

1) Схемы, основанные на семантике текста

Схемы, основанные на семантике текста, фокусируются на использовании семантики текста для выполнения процесса встраивания [12], [13]. Глаголы, существительные, прилагательные, местоимения, синонимы и другие грамматические особенности текстового содержимого используются для того, чтобы скрыть сообщение водяной метки. Эти грамматические изменения делаются без ущерба для первоначального смысла текста. Порядок слов в предложениях также можно изменить, чтобы скрыть биты. Это можно сделать, изменив структуру текста и вставив водяную метку, например, переместив деепричастную фразу, добавив тему или изменив предложение с активного на пассивное предложение. Как можно увидеть, такая схема требует модификации исходного файла, которая не может быть применена в большинстве сценариев, так как слова документа не могут быть изменены в официальных случаях.

2) Схемы, основанные на структуре текста

Аналогичным образом, метод, основанный на структуре текста, также не может быть использован в большинстве случаев. Структурный подход изменяет структуру текста, чтобы встроить необходимые биты [14], [15]. Он включает в себя общее форматирование свойств текста, в котором текстовое содержимое изменяется с использованием его слов или предложений, чтобы скрыть информацию о водяных метках. Расположение слов и букв или стиль письма также могут быть изменены, чтобы скрыть биты водяных меток. Это включает в себя повторение некоторых букв или изменение особенностей текста.

Данная схема изучает общие свойства текста, и некоторые физические свойства в макете текста используются для того, чтобы скрыть биты водяных меток. Одним из примеров метода, основанного на структуре текста, является метод сдвига слов и предложений вверх или вниз для встраивания битов водяной метки. Он может состоять из алгоритма сдвига строк, в котором предложение перемещается вверх или вниз, алгоритма сдвига слов, который перемещает слова по горизонтали, или кодирования объектов определенного текста, которые изменяются для встраивания битов водяных меток. Анализ среднего расстояния между словами в каждой строке также представляет собой один из структурных методов, применяемых в водяных метках, где расстояние для встроенной водяной метки основано на некоторых формулах. Поскольку метод на основе структуры изменяет структуру текста, такую как интервалы между словами или знаки препинания, чтобы встроить водяную метку, он также не подходит для встраивания водяных меток в реальные документы.

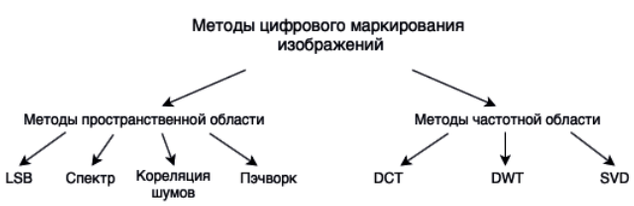
3) Схема, основанная на изображениях

Подход, основанный на изображениях, относится как к категориям цифрового маркирования текста, так и к категориям цифрового маркирования изображений. В схеме цифрового маркирования на основе изображения текстовое содержимое понимается как серия текстовых изображений, в которых изображение водяной метки или сообщение встроены в текст. Текст с водяными метками рассматривается как изображение, и текст больше не может быть скопирован и вставлен и должен быть перепечатан для воспроизведения.

Исходя из этих соображений, далее будут описаны методы цифрового маркирования изображений.

E. Методы цифрового маркирования изображений

Алгоритмы, используемые в классических методах цифрового маркирования изображений делятся по типу области, в которую переводятся исходные данные [16], [17]. Выделяют две основные области, в которые может быть внедрена метка: пространственная область и частотная область. Водяные метки пространственной области изменяют значения пикселей в одной или нескольких областях изображения. Цифровые метки частотной области основаны на изменении некого коэффициента преобразования. Частотное преобразование переводит изображение в его частотное представление, и изображение сегментируется на несколько частотных диапазонов. В то время как методы пространственной области имеют небольшую сложность и могут быть легко применимы, они не могут противостоять обычным атакам на обработку изображений. Встроенная цифровая метка в частотной области изображения может обеспечить бóльшую надежность, чем в пространственной области. На рисунке 2 изображены основные методы цифрового маркирования изображений в различных областях.



Р и с. 2. Наиболее распространенные методы цифрового маркирования изображений

F i g. 2. The most common digital image labeling methods

1) Метод, основанный на LSB изображения

Одним из самых простых методов, использующих пространственную область, является метод, использующий LSB (Least Significant Bit) [18], [19], при котором пиксель внутри изображения изменяется на значение бита секретного сообщения. Встроенные биты зашифровываются с помощью ключа. Что-

бы извлечь водяную метку, этот ключ должен быть известен. Основным недостатком метода, основанного на LSB, является низкая надежность по отношению к шуму, который легко деформирует метку. Кроме того, если используется достаточно простой алгоритм встраивания, закодированное сообщение может быть легко восстановлено и даже изменено третьей стороной. Тем не менее, LSB является одним из самых популярных алгоритмов цифрового маркирования из-за его огромной информационной емкости.

2) Метод, основанный на спектре изображения

Еще один метод, использующий преобразование в пространственную область, рассматривает изображение как канал связи и цифровую метку как сигнал, проходящий через него. Данный метод называется методом, основанном на спектре изображения [20]. Таким образом, атаки и непреднамеренные искажения сигнала рассматриваются как шум, к которому погруженный сигнал должен быть невосприимчив. В самом простом виде данного метода биты кодируемой информации модулируются зашифрованным псевдошумовым сигналом, масштабируются для уменьшения видимости и добавляются к пикселям изображения [21].

3) Метод, основанный на корреляции шумов

Данный метод использует аддитивный белый гауссовский шум для внедрения метки в изображения [22]. Для получения результирующего изображения с водяной меткой к изображению добавляется белый гауссовский шум, умноженный на коэффициент зашумленности. Данный коэффициент зависит от кодируемого бита сообщения водяной метки. При извлечении метки из изображения считается корреляции между исходным шумом и шумом на одной из области картинки и делается вывод о закодированном бите. Тем самым, данный метод цифрового маркирования имеет дополнительный параметр – коэффициент зашумленности, который нужно подбирать для каждого изображения. Но правильный выбор степени добавленного шума позволит защититься от атак, применяющих различные фильтры на изображения [23].

4) Метод, основанный на пэчворке

Алгоритм пэчворка является популярным методом цифрового маркирования документов через пространственную область. Пэчворк – это алгоритм, который использует статистические свойства небольших областей изображения во время встраивания метки и ее обнаружения. В своем самом простом виде он выбирает случайные пары наборов пикселей, а затем изменяет значения пикселей каждого набора различными способами. Затем выборочные средние значения каждого набора изменяются соответствующим образом или местоположение среднего значения сдвигается в соответствии со значением и его знаком. Аддитивный алгоритм пэчворка обнаруживает водяные метки путем вычисления различий между средними значениями двух наборов и применения проверки сформулированной гипотезы [24].

Независимость алгоритма пэчворка является его существенной особенностью. Этот значит, что из двух наборов мы можем извлекать информацию о водяных метках даже без знания оригинального изображения. Также, выбранные наборы пикселей скрывают информацию на небольшом участке изображения, чтобы скрыть один бит. Таким образом, артефакты встраивания метки менее заметны.

5) Метод, использующий DCT

Дискретно-косинусное преобразование (DCT) – это популярный метод цифрового маркирования в частотной области [25], [26]. DCT позволяет разбить изображение на области высоких, средних и низких частот. Для внедрения метки в основном выбираются области средних частот, потому что встраивание водяной метки в полосу низких частот может рассеивать информацию по наиболее визуально важным частям изображения, а выбор высоких частот чувствительно к чрезмерному удалению информации с помощью сжатия и атак шумом. Небольшая часть изображения переводится в частотную область и из области частот выбираются два значения, в которые встраиваются один бит информации о водяной метке.

6) Метод, использующий DWT

Дискретное вейвлет-преобразование (DWT) – это еще один метод, оперирующий в частотной области. Он включает в себя разложение изображения на частотные каналы. Изображение разделяется на четыре поддиапазона, а именно низкие и высокие частоты для строк и столбцов. Каждый поддиапазон тоже может быть разделен аналогичным образом и данное преобразование повторяется до тех пор, пока не будет достигнута желаемая глубина [27]. Глубина зависит от размера метки, а также от количества меток, которые будут внесены в изображение.

Изменения поддиапазона низких частот строк более заметны зрительной системой человека, поэтому цифровые метки обычно встраиваются в один или несколько из трех других поддиапазонов [28].

7) Метод, использующий SVD

Алгоритм сингулярного разложения (SVD) является инструментом для анализа матриц. При использовании преобразования SVD матрица разлагается на три матрицы U, D, V, где U и V – унитарные матрицы, а D – диагональная матрица.

При нанесении цифровой метки с использованием SVD, изображение рассматривается как матрица и над ней выполняется преобразование (1) [29].

$$SVD(image)=U*D*V \quad (1)$$

Затем к полученной диагональной матрице D прибавляется метка (также представленная в виде матрицы), умноженная на коэффициент видимости метки (2).

$$D_w=D+a*W \quad (2)$$

Далее, над полученной матрицей также выполняется SVD преобразование и полученная диагональная матрица используется в качестве диагональной для изначального преобразования SVD. То есть, маркированное изображение описывается формулой (3).

$$image_w=U*D_w*V \quad (3)$$

F. Новые методы цифрового маркирования

В последние годы все большее внимание привлекает использование машинного обучения и глубоких нейронных сетей для внедрения и извлечения цифровых меток из данных [30-32]. Нейронные сети – это мощный метод для извлечения и использования информации, содержащейся в данных. Они распознают закономерности в наборе данных путем обучения, а не путем программирования. Благодаря этому, нейронные сети строят модели, которые в большей степени отражают структуру данных за значительно меньшее время. Также, нейронные сети хорошо адаптируются в меняющихся условиях.

Запрограммированные системы ограничены ситуацией, для которой они были разработаны; когда условия меняются, такие системы могут оказаться не способны воспроизводить хорошие результаты. Хотя нейронным сетям может потребоваться некоторое время, чтобы научиться внезапным резким изменениям, они превосходно адаптируются к постоянно меняющейся информации. Нейронные сети могут обрабатывать очень сложные взаимодействия и, таким образом, могут легко моделировать данные, которые слишком сложно моделировать с помощью традиционных подходов.

В отличие от значительных достижений в области классического цифрового маркирования, маркирование с помощью нейронных сетей менее развито и изучено.

Основная проблема развития методов цифрового маркирования с помощью нейронных сетей – это большое количество требований к ним. В большинстве исследований используются простые модели нейронных сетей, в комбинации с классическими методами цифрового маркирования [33], [34]. Целью этих методов является невидимость метки, а не ее надежность. Но такие методы менее практичны, так как являются чувствительными к изменениям данных. Чтобы добиться надежности цифровых меток, используются более сложные модели нейронных сетей, которые показывают более высокую производительность в различных задачах, но их разработка и внедрение занимает больше времени [35], [36].

Основные атаки на методы цифрового маркирования

В большинстве приложений, которые применяют методы цифрового маркирования, маркированные данные, вероятно, будут каким-то образом обработаны до того, как они попадут к получателю. Например, такой обработкой может быть сжатие с потерями, добавление шумов или преобразование в другой формат. Некоторые типы обработки могут быть применены с явной целью удаления или искажения метки. Таким образом, встроенная водяная метка может быть преднамеренно или непреднамеренно быть повреждена.

В терминологии цифрового маркирования атака – это любая обработка, которая может помешать обнаружению метки или передаче информации, передаваемой водяной меткой. Важным аспектом любого метода цифрового маркирования является его устойчивость к атакам. Одна из распространенных классификаций существующих атак на цифровые метки содержит четыре класса атак: удаляющие атаки, геометрические атаки, криптографические атаки и протокольные атаки. В данном разделе мы в общих чертах опишем эти четыре типа атак и приведем несколько примеров.

A. Удаляющие атаки

Удаляющие атаки направлены на полное удаление водяной метки из маркированных данных без взлома алгоритма нанесения водяной метки, то есть, например, без ключа, используемого для встраивания. В результате такой атаки никакая обработка, даже чрезмерно сложная, не сможет восстановить информацию, содержащуюся в метке, из атакованных данных. Примерами данной категории являются удаление шума, размытие, сжатие с потерями и атака по сговору. Атаки по сговору [37] могут быть применимы, когда злоумышленник или группа злоумышленников могут получить множество копий данных, промаркированных разными метками. В таком случае успешная атака может быть достигнута путем усреднения всех копий или взятия только небольших частей из каждой отдельной копии.

Не все из этих методов всегда приближаются к своей цели полного удаления водяных меток, но, тем не менее, они могут значительно повредить метку.

B. Геометрические атаки

В отличие от удаляющих атак, геометрические атаки не удаляют саму встроенную водяную метку, а пытаются ее исказить. Такими атаками, например, являются кадрирование и масштабирование.

Устойчивость к глобальным геометрическим искажениям часто зависит от использования либо области, инвариантной к преобразованию, либо метода, сравнивающего исходные данные с полученными, либо специально разработанных периодических водяных меток, функция автоковариации которых позволяет оценивать геометрические искажения. Однако, злоумышленник может разрабатывать специальные атаки, используя знания о схеме синхронизации.

Тем не менее, большинство современных методов цифрового маркирования решают проблему устойчивости к глобальным аффинным преобразованиям [38]. Однако устойчивость к локальным случайным изменениям, по-прежнему остается открытой проблемой для большинства коммерческих инструментов для нанесения водяных меток.

C. Криптографические атаки

Криптографические атаки направлены на взлом методов защиты в схемах нанесения и извлечения водяных меток и, таким образом, поиск способа удаления встроенной информации или внедрения вводящих в заблуждение водяных меток. Одним из таких методов является поиск внедренной секретной информации методом перебора. Другой атакой в этой категории является так называемая атака с оракулом [39], которая использует устройство обнаружения цифровой метки для создания данных без нее. Практическое применение этих атак ограничено из-за их высокой вычислительной сложности.

D. Протокольные атаки

Протокольные атаки направлены на атаку всей концепции приложения для нанесения водяных меток. Один из типов протокольных атак основан на обратимости водяных меток [40]. Предположим, что владелец авторских прав A вносит свою водяную метку W_A в данные I с помощью алгоритма кодирования E_A. Таким образом, создаются данные I_A с водяной меткой, доступный как законному пользователю, так и злоумышленнику по имени B. Предположим, что B обладает инструментом для нанесения цифровых меток, состоящим из алгоритма кодирования E_B и алгоритмом декодирования D_B. Если B способен создать водяную метку W_B, а также поддельные данные I' из I_A таким образом, что:

(1) $E_B(I', W_B) = I_A$;

(2) $D_B(I_A, W_B) = I$;

(3) I' схож с I_A

То тогда система цифрового маркирования (E_B, D_B) называется обратимой. В соответствии с описанной выше операцией как подлинная водяная метка W_A, так и поддельная водяная метка W_B могут быть обнаружены из I_A с использованием D_A и D_B со-

ответственно. Таким образом, законный владелец I_A не может быть идентифицирован на основе извлеченной водяной метки. Решением этой проблемы может использование односторонних функций.

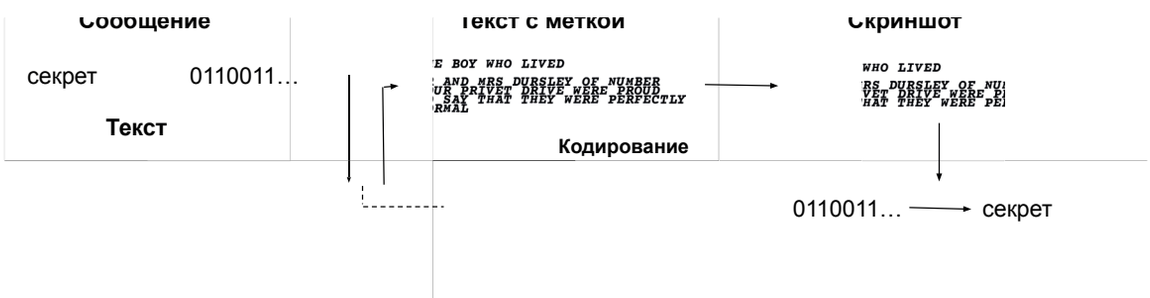
Другой протокольной атакой является атака копирования [41]. В этом случае цель состоит не в том, чтобы уничтожить метку или затруднить ее обнаружение, а в том, чтобы получить водяную метку из маркированных данных и скопировать ее в некоторые другие данные, называемые целевыми данными. Атака копирования применима, когда цифровая метка в целевых данных может быть создана без знания алгоритма технологии нанесения водяной метки или знания ключа для нанесения водяной метки.

Предлагаемая архитектура метода цифрового маркирования документов

Во всех рассмотренных областях методы встраивания информации удовлетворяют двум требованиям: маркирующиеся данные минимально возмущены, то есть алгоритм пытается оставить данные максимально неизмененными; и встроенное сообщение может быть надежно восстановлено даже при наличии некоторых ошибок декодирования.

Тем не менее, методы встраивания информации в текстовые документы очень требовательны. Текстовое цифровое маркирование считается более сложным для удовлетворения вышеупомянутых требований, чем его аналоги для изображений, видео и аудио. Это связано с тем, что “пиксель” текстового документа представляет собой отдельные буквы, которые, в отличие от пикселя изображения, не могут быть изменены на другие буквы, не вызывая заметных различий.

Поэтому в предлагаемой архитектуре метода цифрового маркирования документов вместо того, чтобы заменять буквы текста, мы будем изменять глифы (то есть конкретные формы букв) их шрифтов для кодирования информации. Основным требованием к методу, в дополнении к требованиям описанным выше, является читаемость исходного документа полностью. То есть алгоритм должен подбирать изменение глифа таким образом, чтобы оно оказывало минимальное влияние на внешний вид шрифта текстового документа, гарантируя при этом, что изменение глифа может быть распознано. Чтобы восстановить встроенную информацию, мы предлагаем архитектуру алгоритма декодирования, который восстанавливает информацию из входного закодированного документа, анализируя глифы. Жизненный цикл цифровой метки предлагаемого метода описан на рисунке 3.



Р и с. 3. Жизненный цикл цифровой метки текстовых документов

F i g. 3. Life cycle of a digital label for text documents

A. Метод внедрения цифровой метки в документ

Предлагаемая система кодирования шрифтов (рис. 4) встраивает в текстовый документ любой тип информации в виде битовой строки. Например, произвольное текстовое сообщение может быть закодировано в битовую строку с использованием стандартного кода ASCII или Unicode. Будем называть такую битовую строку сообщением. Также на этом этапе возможно применение дополнительного кодирования или добавления избыточности для увеличение надежности декодирования метки.

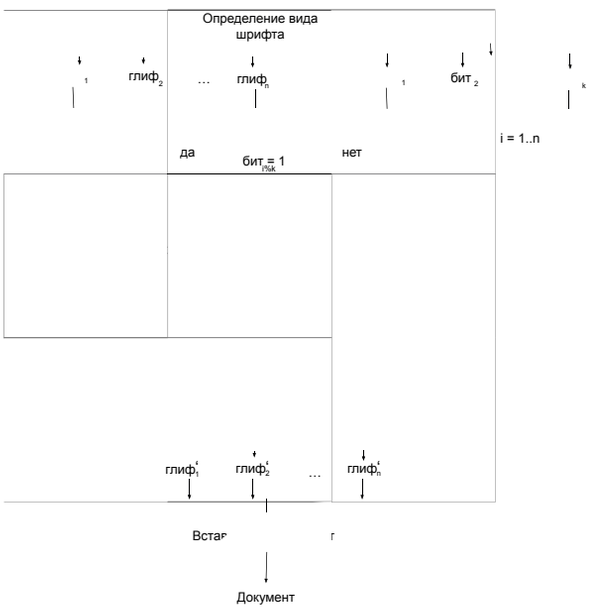
В текстовом документе основными элементами встраивания сообщения будут являться буквы одного шрифта. Идея состоит в том, чтобы изменить глиф каждой буквы, чтобы встроить сообщение. Для этого можно использовать модели генерации шрифтов, которые могут быть использованы для интерполяции в векторном пространстве шрифтов [42-44]. Для изменения глифов, текст внутри документа для начала должен быть распознан и предобработан. Для этого могут использоваться механизмы распознавания текста, которые также способны выделить прямоугольную область каждой буквы. Также на данном этапе важно определить основной шрифт документа, который может быть найден в метаданных документа или распознан с помощью технологии распознавания шрифтов [45]. Далее метод работает в отдельной области каждой буквы и применяет модель, позволяющую изменить очертание буквы определенного шрифта в соответствии с кодируемым битом.

Также в данной архитектуре предлагается кодировать сообщение в текст до тех пор, пока остаются незакодированные глифы. Тем самым это дает возможность декодировать метку из разных областей текста, а также увеличить устойчивость метки к атакам кадрирования при применение подходящих для этого алгоритмов для извлечения метки.

B. Метод извлечения цифровой метки из документа

Чтобы извлечь информацию из закодированного текстово-

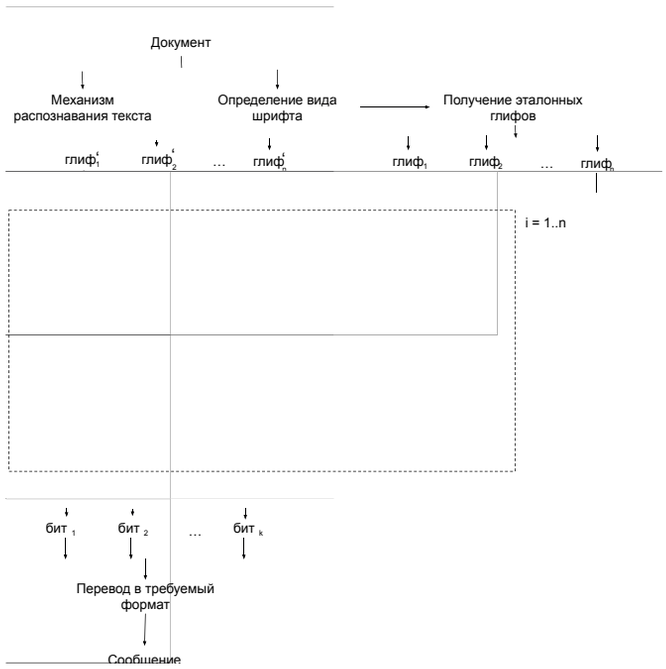
го документа (рис. 5), первым шагом нужно предобработать текст аналогично предобработке текста в этапе внедрения метки.



Р и с. 4. Архитектура метода внедрения метки в документ
F i g. 4. Architecture of the method for embedding a label in a document

В результате метод получает прямоугольные области каждой буквы и пытается восстановить бит сообщения из нее. Метод извлекает бит 0, если глиф распознается как неизмененный и бит 1, если глиф был изменен. Для распознавания могут использоваться различные методы, например, метрики сходства изображений или модели машинного обучения. Данные методы могут использовать исходные, немаркированные глифы для определения расстояния между текущим глифом и эталонным. Исходя из этого расстояния, метод будет принимать решение о значении встроенного бита. Извлеченные биты затем вводятся в схему кодирования для восстановления оригинального сообщения.

Представленные архитектуры могут быть усложнены, например, могут кодироваться не биты, а целые числа. Это может быть реализовано через шаг интерполяции между шрифтами, согласно которому, изменение глифа на i шагов интерполяции будет означать кодирование i-го целого числа. В дополнение, встраивание битов в текст может происходить не в порядке следования глифов, а может быть, например, фрактальным. Порядок встраивания битов должен учитываться при извлечении метки.



Р и с. 5. Архитектура метода извлечения метки в документ
F i g. 5. Architecture of the method for extracting a label into a document

Далее будет приведена реализация предложенной архитектуры цифрового маркирования документов для оценки ее работоспособности и эффективности.

Алгоритмическая основа реализации метода

В данном разделе будут описаны алгоритмы, применяющиеся в реализации метода цифрового маркирования документов, основанного на предложенной архитектуре.

A. Модель глубокой факторизации

В рассматриваемой задаче применяется модель глубокой нейронной сети, используемая в задаче реконструкции шрифтов [42]. После обучения на наборе шрифтов система восстанавливает недостающие символы в наборе ранее невидимых шрифтов, обусловленных небольшим наблюдаемым подмножеством изображений глифов. Существенным свойством данной модели является возможность интерполироваться между одинаковыми буквами разных шрифтов (рис. 6). Данное свойство позволяет нам строить небольшие отклонения от стандартных шрифтов, тем самым изменять глифы букв в соответствии с кодируемым битом.

Модель рассматривает коллекцию шрифтов в виде матрицы X, где каждый столбец соответствует определенному типу буквы, а каждая строка соответствует определенному шрифту (стилю). Каждая запись в матрице x_{ij} представляет собой изображение глифа для символа i в стиле шрифта j, которое рассматривается как черно-белое изображение размера 64 × 64.

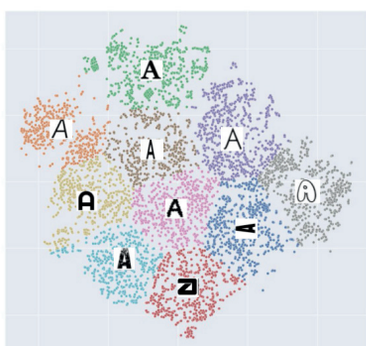
Р и с. 6. Интерполяция букв между двумя шрифтами
F i g. 6. Interpolation of letters between two fonts

Учитывая набор изображений глифов, состоящих из I символов в J шрифтах, модель разделяет структурные атрибуты, специфичные для символов, и стилистические атрибуты, специфичные для шрифтов, на два разных векторных представления. Более конкретно, для каждого шрифта в коллекции переменная шрифта $z_j \in R^k$, выбирается из фиксированного многомерного гауссовского априорного распределения $p(z_j) = N(0,I_k)$. Далее каждое изображение глифа x_{ij} генерируется независимо, обусловленное соответствующей переменной шрифта z_j и вектором параметров, специфичных для символов $e_i \in R^k$, который называется представлением символа. Таким образом, глифы одного и того же символьного типа совместно используют векторное представление символа, в то время как глифы одного и того же шрифта совместно используют переменную шрифта. Этот подход к моделированию можно рассматривать как форму глубокой факторизации матрицы, где содержимое в любой заданной ячейке является функцией векторных представлений соответствующих строк и столбцов.

В соответствии с моделью, распределение вероятностей по каждому изображению, описываемому z_j и e_i, измеряется нейронной сетью, называемой декодером. Распределение декодера определяется как $p(x_{ij}|z_j; e_i, \varphi)$, где φ представляет параметры, общие для всех глифов, которые определяют, как переменные шрифта и символов объединяются для получения изображений глифов. Совместная вероятность в рамках модели задается формулой (4).

$$p(X, Z; E, \Phi) = \prod_{i,j} p(x_{ij}|z_j; e_i, \varphi)p(z_j) \quad (4)$$

Тем самым, модель пытается изучить плавный переход между стилями и буквами и позволяет также выполнить интерполяцию между векторными представлениями шрифтов. В нашем методе мы берем два шрифта из одного семейства шрифтов, которые отличаются по одному свойству, и пропускаем их через кодировщик, чтобы получить скрытую переменную шрифта для каждого. Затем выполняется интерполяция между этими значениями и полученные вектора передаются в декодер для создания новых шрифтов, существующих между рассматриваемыми. На рисунке 7 можно увидеть полученное векторное пространство шрифтов и его разбиение на 10 кластеров (семейств шрифтов).



Р и с. 7. Векторное пространство шрифтов, полученное моделью [42]
F i g. 7. Vector font space obtained by the model

B. Коды, исправляющие ошибки

Чтобы противостоять изменениям, которые могут быть умышленно или неумышленно применены к текстовому документу, метод также использует коды, исправляющие ошибки.

Для кодирования исходной последовательности бит сообщения сначала используется код циклической проверки избыточности (CRC) [46], а затем также применяется код Боуза-Чоудхури-Хоквингема (БЧХ) [47] для генерации последовательности с возможностью проверки и исправления ошибок.

Циклическая проверка избыточности (CRC) – это метод, используемый для обнаружения ошибок в цифровых данных. С точки зрения его использования CRC – это хэш-функция, которая обнаруживает случайные изменения в необработанных данных.

Математически CRC можно описать как представление двоичного слова данных как полином над GF(2) и выполнение полиномиального деления на генераторный полином G(x), который обычно называют полиномом CRC. Остаток от операции деления выдает значение обнаружения ошибки.

В теории кодирования коды Бозе-Чаудхури-Хокуэнгема (коды БЧХ) образуют класс циклических кодов с исправлением ошибок, которые также строятся с использованием полиномов над полем Галуа. Одной из ключевых особенностей кодов БЧХ является то, что во время разработки кода существует точный контроль над количеством ошибок символов, исправляемых кодом. В частности, можно создавать двоичные коды БЧХ, которые могут исправлять многобитовые ошибки.

C. Метод распознавания букв в разных шрифтах

Для того чтобы корректно определять встроенный бит каждого символа, сначала нужно определить вид этого символа. Для этого будет использоваться механизм оптического распознавания символов «Tesseract» [48].

Обработка текста происходит в несколько этапов. Первый шаг – это анализ компонентов текста, в котором анализируются контуры компонентов. На этом этапе контуры группируются в связанные объекты. Затем, каждый связанный объект анализируется на предмет некого фиксированного шага (то есть пробельного символа). Текст с фиксированным шагом сразу же разбивается на ячейки символов. Затем распознавание происходит в виде двухпроходного процесса. В первом проходе делается попытка распознать каждое слово по очереди. Каждое возможное слово передается в адаптивный классификатор в качестве обучающих данных. Поскольку адаптивный классификатор, возможно, узнал что-то полезное слишком поздно, чтобы внести свой вклад в верхнюю часть страницы, по странице выполняется второй проход, в котором слова, которые были распознаны недостаточно хорошо, распознаются снова. Заключительная фаза разрешает нечеткие пробелы и проверяет альтернативные гипотезы для высоты текста, чтобы найти текст мелким шрифтом.

Для модели «Tesseract» возможно распознавание не только слов, но и более мелких частей текста, таких как буквы.

D. Метод сравнения изображений

Большинство методов сравнения изображений основаны на количественном определении ошибок между эталонным и рассматриваемым изображением. Общей метрикой является количественная оценка разницы в значениях каждого из соответствующих пикселей между образцом и эталонными изображениями (с использованием, например, среднеквадратичной ошибки).

Система визуального восприятия человека обладает высокой способностью идентифицировать структурную информацию из сцены и, следовательно, определять различия между информацией, извлеченной из эталонной и образцовой сцены. Следовательно, метрика, которая воспроизводит это поведение, будет лучше работать в задачах, связанных с различением рассматриваемого и эталонного изображения.

Показатель Индекса структурного сходства (SSIM) [49] извлекает из изображения 3 ключевых признака:

- Яркость
- Контраст
- Структура

Сравнение между двумя изображениями выполняется на основе этих 3 признаков. Такие признаки кажутся наиболее подходящими для реализации рассматриваемого метода. Для оценки качества изображения полезно применять индекс SSIM локально, а не глобально.

Описание метода цифрового маркирования документов

Рассматриваемый метод встраивания сообщений в документы состоит из нескольких этапов:

1) Предварительное вычисление векторных представлений шрифтов и символов

2) Кодирование встраиваемого сообщения

3) Обнаружение областей глифов во входном документе

4) Встраивание битов закодированного сообщения в глифы входного документа

Во время предварительного вычисления создается набор векторных представлений глифов для широко используемых шрифтов. Его создание происходит с помощью модели, описанной выше.

Затем для генерации последовательности с возможностью исправления ошибок и проверки, к встраиваемому сообщению сначала применяется код циклической проверки избыточности (CRC), а затем используется код БЧХ. После кодирования, в режиме реального времени, происходит извлечение из входного документа текстового содержимого и расположения букв. Предлагаемый метод также должен знать исходный шрифт текста. Это может быть задано пользователем или автоматически получено из метаданных векторного графического документа (например, PDF). Если документ представлен в виде пиксельного изображения, для распознавания шрифта текста можно использовать новейшие методы [50].

Для изменения глифов текста, мы будем использовать модель глубокой факторизации для типографского анализа [42]. Принимая в качестве входных данных набор изображений глифов шрифтов, модель отделяет стилистические особенности шрифтов от структуры каждого символа. То есть модель генерирует векторное представление стиля каждого шрифта, а также параметризует структуру каждого символа в шрифте. Важно отметить, что в результате работы модели, каждый вектор стиля является общим для всех символов в шрифте, в то время как представления символов являются общими для символов одного и того же типа во всех шрифтах. Эта генеративная модель предварительно вычисляется один раз для каждого шрифта. Затем это позволяет нам изменять глиф каждой текстовой буквы. Тем самым, в режиме онлайн, предоставив текстовый документ (или текстовую область или абзацы), наш метод может изменять глифы букв в документе, чтобы встроить сообщение. Чтобы вставить нужный бит в букву шрифта, мы ищем векторное представление его шрифта и находим нужный глиф. Затем мы интерполируемся на значение λ_i, где i - значение зашифрованного бита в сторону другого шрифта из этого же семейства. Полученный интерполяцией глиф масштабируется, чтобы поместить его в ограничивающую рамку исходной буквы (которая обнаруживается методом распознавания текста), и затем используется для замены исходной буквы входного документа.

Извлечение закодированного сообщения из документа также происходит в несколько этапов:

1) Предобработка текста
2) Обнаружение областей глифов во входном документе
3) Поиск алгоритмом сравнения изображений наиболее близкого значения бита
4) Декодирование полученного сообщения

Чтобы извлечь целые числа из пиксельного изображения, мы извлекаем текстовое содержимое и идентифицируем обла-

сти отдельных букв с помощью инструмента распознавания текста. После того, как мы обрезаем области букв, они обрабатываются алгоритмом сравнения: эталонное изображение, содержащее бит , сранивается с текущим глифом и выбирается бит 0 или 1, для которого значение SSIM больше.
Извлеченные биты затем вводятся в схему кодирования с исправлением ошибок для восстановления сообщения. Из-за избыточности данных, даже если некоторые символы распознаются ошибочно (например, из-за плохого качества изображения), эти ошибки будут исправлены, и мы все равно сможем правильно восстановить сообщение.

Экспериментальное исследование

Для обучения модели глубокой нейронной сети мы используем набор данных Capitals64, который содержит черно-белые изображения размера 64×64 английских заглавных букв в 10682 различных шрифтах. Данный набор данных разбивается на обучающую, валидационную и тестовую выборки в пропорции 70%, 15%, 15%. Мы обучаем нашу модель максимизировать логарифм вероятности 4.1, используя алгоритм оптимизации Адам [51] с размером шага 10^{-5}. Чтобы повысить надежность кодировщика, мы случайным образом отбрасываем глифы во время обучения с вероятностью 0,7 (отбрасывая также все символы шрифта глифа). Реализация выполнена с помощью библиотеки PyTorch [52] версии 1.11.0.
В нашем эксперименте был выбран размер сообщения равный 11 символам ASCII. Код, исправляющий ошибки – это БЧХ который может исправлять 10 ошибок, в то время как длина битов CRC составляет 17 бит. Таким образом, все сообщение состоит из 241 бит. Размер сообщения может быть разным в разных сценариях применения предложенного метода, от него будет также зависеть количество глифов, нужных для кодирования сообщения.

Таблица 2. **Области применения цифрового маркирования**
Table 2. **Areas of application of digital marking**

Шрифт		PSNR	SSIM	Точность	Ошибка
Courier Prime Bold Italic	0	361.2	1	-	-
	1	31.18	0.99	100%	0
	2	22.13	0.95	100%	0
	3	17.93	0.91	100%	0
	4	16.09	0.88	100%	0
	5	14.91	0.86	100%	0
Coving21	0	361.2	1	-	-
	1	44.11	0.99	100%	0
	2	38.03	0.99	100%	0
	3	34.46	0.99	100%	0
	4	31.9	0.99	100%	0
	5	29.95	0.99	100%	0
Creator CreditsB Bita	0	361.2	1	-	-
	1	34.17	0.99	100%	0
	2	26.79	0.98	100%	0
	3	22.32	0.96	100%	0
	4	20.1	0.95	100%	0
	5	18.4	0.93	100%	0
FiraSans-Regular	0	361.2	1	-	-
	1	49.63	0.99	100%	6
	2	43.71	0.99	100%	6
	3	40.29	0.99	100%	0
	4	37.88	0.99	100%	0
	5	36.09	0.99	100%	0
VDS Compen-sated Light-Italic	0	361.2	1	-	-
	1	37.47	0.99	100%	0
	2	31.39	0.99	100%	0
	3	27.84	0.98	100%	0
	4	25.53	0.97	100%	0
	5	23.8	0.96	100%	0

Качество документов с водяными метками (невидимость метки) оценивается по разнице между документом до и после сокрытия секретной информации. В качестве метрики, мы используем пиковое отношение сигнала к шуму (PSNR) и индекс структурного сходства (SSIM) для оценки качества документа. В таблице 2 представлены 5 случайно выбранных шрифтов, которые были проинтерполированы с другим шрифтом из своего семейства на шаг интерполяции. Также в таблице указана точность (кол-во битов верно распознанных после декодирования метки) и кол-во ошибочно распознанных битов до применения кодов БЧХ для обнаружения ошибки.
Можно заметить, что во всех случаях сообщение было декодировано корректно, а полученные ошибки небольшие и присутствуют только на небольших шагах интерполяции. Метрика PSNR резко падает при внедрении метки, но метрика SSIM показывает, что структура текста в виде изображения практически не изменяется и падает с увеличением шага интерполяции. Тем самым, чем меньше шаг интерполяции, тем более невидима метка и текст сохраняет свою читабельность. При этом с увеличением шага интерполяции метка распознается лучше. При проектировании системы цифрового маркирования нужно учитывать эту особенность и подбирать нужный шаг исходя из основных задач системы. При этом в общих случаях оптимальные шаги интерполяции могут быть вычислены до введения модели маркирования в работу и применяться индивидуально для каждого шрифта.

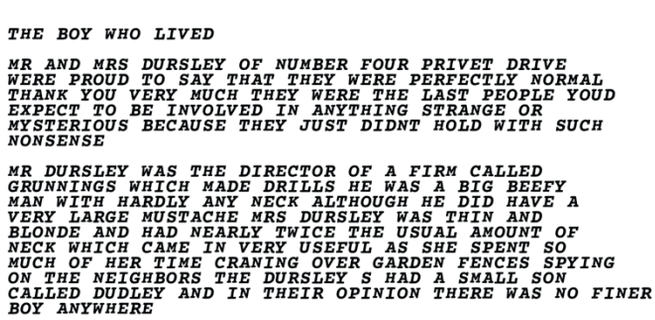
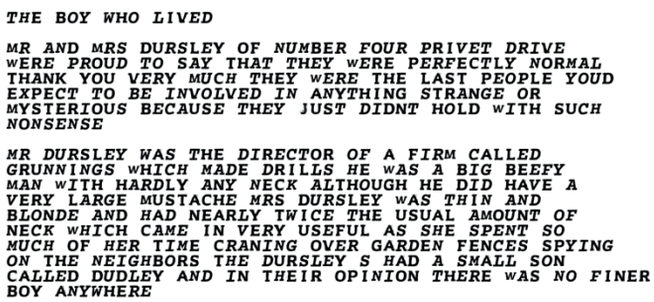


Рис. 8. Внедренная в текст метка на шаге интерполяции
Fig. 8. Text-embedded label in the interpolation step



Р и с. 9. Внедренная в текст метка на шаге интерполяции

F i g. 9. Text-embedded label in the interpolation step

На рис. 8, 9 показан пример документов с водяными метками, на разных шагах интерполяции, соответствующие шрифту Courier Prime Bold Italic.

Заключение

Приложения для нанесения и извлечения цифровых водяных меток получают широкое распространение в современном обществе. В этой работе была предложена архитектура нанесения и извлечения цифровых меток для защиты подлинных документов. Представленный подход основан на изменение глифов символов, который мало изучен в литературе. Предлагаемая архитектура позволяет создавать невидимые цифровые метки, не влияющие на содержание текста, и извлекать их при работе с документами. В частности, был рассмотрен один из вариантов реализации данной архитектуры, основанный на генеративной модели шрифтов, который показал высокую точность при обнаружении скрытой информации, а также относительную невидимость метки.

References

[1] Hiary H., Ng K. Watermark: From Paper Texture to Digital Media. *First International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS'05).* IEEE Computer Society, USA; 2005. 4 p. (In Eng.) doi: https://doi.org/10.1109/AXMEDIS.2005.50

[2] Tsolis D., Nikolopoulos S., Drossos L., Sioutas S., Papatheodorou T. Applying robust multibit watermarks to digital images. *Journal of Computational and Applied Mathematics*. 2009; 227(1):213-220. (In Eng.) doi: https://doi.org/10.1016/j.cam.2008.07.035

[3] Zhang X., Wang S., Zhang K. Multi-bit Watermarking Scheme Based on Addition of Orthogonal Sequences. In: Gorodetsky V., Popyack L., Skormin V. (eds.). *Computer Network Security. MMM-ACNS 2003. Lecture Notes in Computer Science.* Vol. 2776. Springer, Berlin, Heidelberg; 2003. p. 407-418. (In Eng.) doi: https://doi.org/10.1007/978-3-540-45215-7_35

[4] Rahman A.U., Sultan K., Musleh D., Aldhafferi N., Alqahtani A., Mahmud M. Robust and Fragile Medical Image Watermarking: A Joint Venture of Coding and Chaos Theories. *Journal of Healthcare Engineering*. 2018; 2018:8137436. (In Eng.) doi: https://doi.org/10.1155/2018/8137436

[5] Tao H., Chongmin L., Zain J.M., Abdalla A.N. Robust Image Watermarking Theories and Techniques: A Review. *Journal of Applied Research and Technology*. 2014; 12(1):122-138. (In Eng.) doi: https://doi.org/10.1016/S1665-6423(14)71612-8

[6] Anbarjafari G., Ozcinar C. Imperceptible non-blind watermarking and robustness against tone mapping operation attacks for high dynamic range images. *Multimedia Tools and Applications*. 2018; 77(18):24521-24535. (In Eng.) doi: https://doi.org/10.1007/s11042-018-5759-1

[7] Pradhan C., Rath S., Bisoi A. Non Blind Digital Watermarking Technique Using DWT and Cross Chaos. *Procedia Technology*. 2012; 6:897-904. (In Eng.) doi: https://doi.org/10.1016/j.protcy.2012.10.109

[8] Shahid A., et al. A Non-Blind Watermarking Scheme for Gray Scale Images in Discrete Wavelet Transform Domain using Two Subbands. *International Journal of Computer Science Issues*. 2012; 9(5-1):101-109. Available at: https://ijcsi.org/papers/IJCSI-9-5-1-101-109.pdf (accessed 23.12.2021). (In Eng.)

[9] Memon N., Wong P.W. Protecting Digital Media Content. *Communications of the ACM*. 1998; 41(7):35-43. (In Eng.) doi: https://doi.org/10.1145/278476.278485

[10] Kwok S.H., Yang C.C., Tam K.Y. Watermark design pattern for intellectual property protection in electronic commerce applications. *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*. IEEE Computer Society; 2000. p. 1-10. (In Eng.) doi: https://doi.org/10.1109/HICSS.2000.926859

[11] Hartung F., Kutter M. Multimedia watermarking techniques. *Proceedings of the IEEE*. 1999; 87(7):1079-1107. (In Eng.) doi: https://doi.org/10.1109/5.771066

[12] Topkara M., Taskiran C.M., Delp III E.J. Natural language watermarking. In: Delp III E.J., Wong P.W. *Proceedings of SPIE: Security, Steganography, and Watermarking of Multimedia Contents VII.* Vol. 5681. SPIE; 2005. p. 441-452. (In Eng.) doi: https://doi.org/10.1117/12.593790

[13] Mali M.L., Patil N.N., Patil J.B. Implementation of Text Watermarking Technique Using Natural Language Watermarks. *2013 International Conference on Communication Systems and Network Technologies.* IEEE Computer Society; 2013. p. 482-486. (In Eng.) doi: https://doi.org/10.1109/CSNT.2013.106

[14] Jalil Z., Mirza A.M., Sabir M. Content based Zero-Watermarking Algorithm for Authentication of Text Documents. *International Journal of Computer Science and Information Security*. 2010; 7(2):212-217. (In Eng.)

[15] Alotaibi R.A., Elrefaei L.A. Utilizing Word Space with Pointed and Un-pointed Letters for Arabic Text Watermarking. *2016 UKSim-AMSS 18th International Conference on Computer Modelling and Simulation (UKSim).* IEEE Computer Society; 2016. p. 111-116. (In Eng.) doi: https://doi.org/10.1109/UKSim.2016.34

[16] Sarkar T., Sanyal S. Digital Watermarking Techniques in Spatial and Frequency Domain. arXiv:1406.2146v2. 2014. (In Eng.) doi: https://doi.org/10.48550/arxiv.1406.2146

[17] Saqib M., Naaz S. Spatial and Frequency Domain Digital Image Watermarking Techniques for Copyright Protection. *International Journal of Engineering Science and Technology*. 2017; 9(06):691-699. (In Eng.)

[18] Bamatraf A., Ibrahim R., Salleh M.N.B.M. Digital watermarking algorithm using LSB. *2010 International Conference on Computer Applications and Industrial Electronics*. IEEE Computer Society; 2010. p. 155-159. (In Eng.) doi: https://doi.org/10.1109/ICCAIE.2010.5735066

[19] Dixit A., Dixit R. A Review on Digital Image Watermarking Techniques. *International Journal of Image, Graphics and Signal Processing.* 2017; 9(4):56-66. (In Eng.) doi: https://doi.org/10.5815/ijigsp.2017.04.07

[20] Samcovic A., Turan J. Digital Image Watermarking by Spread Spectrum. *Proceedings of the 11th Conference on 11th WSEAS International Conference on Communications – Volume 11 (ICCOM'07).* World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA; 2007. p. 29-32. (In Eng.)

[21] Hartung F.H., Su J.K., Girod B. Spread Spectrum Watermarking: Malicious Attacks and Counterattacks. Security and Watermarking of Multimedia Contents. *Proceedings of SPIE: Security and Watermarking of Multimedia Contents.* Vol. 3657. SPIE; 1999. p. 147-158. (In Eng.) doi: https://doi.org/10.1117/12.344665

[22] Thanki R., Trivedi R., Kher R., Vyas D. Digital Watermarking Using White Gaussian Noise (WGN) in Spatial Domain. *Proceeding of International Conference on Innovative Science and Engineering Technology (ICISET- 2011).* Vol. 1. p. 38-42. (In Eng.)

[23] Thanki R.M., Kher R.K., Vyas D.D. Robustness of Correlation Based Watermarking Techniques Using WGN against Different Order Statistics Filters. *International Journal of Computer Science and Telecommunications*. 2011; 2(4):45-49. Available at: https://www.ijcst.org/Volume2/Issue4/p9_2_4.pdf (accessed 23.12.2021). (In Eng.)

[24] George L.E., Mohammed F.G., Taqi I.A. Effective Image Watermarking Method Based on DCT. *Iraqi Journal of Science.* 2015; 56(3B):2374-2379. Available at: https://www.iasj.net/iasj/download/ebbc3e942c16498f (accessed 23.12.2021). (In Eng.)

[25] Pithiya P.M., Desai H.L. DCT Based Digital Image Watermarking, De- watermarking & Authentication. *International Journal of Latest Trends in Engineering and Technology.* 2013; 2(3):213-219. Available at: https://www.ijltet.org/pdfviewer.php?id=875&j_id=2633 (accessed 23.12.2021). (In Eng.)

[26] Islam S. M. M., Debnath R., Hossain S. K. A. DWT Based Digital Watermarking Technique and its Robustness on Image Rotation, Scaling, JPEG compression, Cropping and Multiple Watermarking. *2007 International Conference on Information and Communication Technology.* IEEE Computer Society; 2007. p. 246-249. (In Eng.) doi: https://doi.org/10.1109/ICICT.2007.375386

[27] Savakar D.G., Pujar S. Digital Image Watermarking at Different Levels of DWT using RGB Channels. *International Journal of Recent Technology and Engineering*. 2020; 8(5):559-570. (In Eng.) doi: https://doi.org/10.35940/ijrte.D6821.018520

[28] Liu R., Tan T. An SVD-based watermarking scheme for protecting rightful ownership. *IEEE Transactions on Multimedia.* 2002; 4(1):121-128. (In Eng.) doi: https://doi.org/10.1109/6046.985560

[29] Jarušek R., Volna E., Kotyrba M. Neural Network Approach to Image Steganography Techniques. In: Matoušek R. (ed.) *Mendel 2015. ICSC-MENDEL 2016. Advances in Intelligent Systems and Computing.* Vol. 378. Springer, Cham; 2015. p. 317-327. (In Eng.) doi: https://doi.org/10.1007/978-3-319-19824-8_26

[30] Tang W., Tan S., Li B., Huang J. Automatic Steganographic Distortion Learning Using a Generative Adversarial Network. *IEEE Signal Processing Letters*. 2017; 24(10):1547-1551. (In Eng.) doi: https://doi.org/10.1109/LSP.2017.2745572

[31] Khan I., Verma B., Chaudhari V.K., Khan I. Neural network based steganography algorithm for still images. *INTERACT-2010.* IEEE Computer Society; 2010. p. 46-51. (In Eng.) doi: https://doi.org/10.1109/INTERACT.2010.5706192

[32] Islam M., Roy A., Laskar R.H. SVM-based robust image watermarking technique in LWT domain using different sub-bands. *Neural Computing and Applications*. 2020; 32(5):1379-1403. (In Eng.) doi: https://doi.org/10.1007/s00521-018-3647-2

[33] Yen C.-T., Huang Y.-J. Frequency domain digital watermark recognition using image code sequences with a back-propagation neural network. *Multimedia Tools and Applications.* 2016; 75(16):9745-9755. (In Eng.) doi: https://doi.org/10.1007/s11042-015-2718-y

[34] Sun L., Xu J., Liu S., et al. A robust image watermarking scheme using Arnold transform and BP neural network. *Neural Computing and Applications*. 2018; 30(8):2425-2440. (In Eng.) doi: https://doi.org/10.1007/s00521-016-2788-4

[35] Mun S.-M., Nam S.-H., Jang H., Kim D., Lee H.-K. Finding robust domain from attacks: A learning framework for blind watermarking. *Neurocomputing.* 2019; 337:191-202. (In Eng.) doi: https://doi.org/10.1016/j.neucom.2019.01.067

[36] Pibre L., Pasquet J., Ienco D., Chaumont M. Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover sourcemismatch. *Proceedings IS&T Int'l. Symp. on Electronic Imaging: Media Watermarking, Security, and Forensics.* Society for Imaging Science and Technology; 2016. (In Eng.) doi: https://doi.org/10.2352/ISSN.2470-1173.2016.8.MWSF-078

[37] Doerr G., Dugelay J.-L. Collusion issue in video watermarking. *Proceedings of SPIE: Security, Steganography, and Watermarking of Multimedia Contents VII.* Vol. 5681. SPIE; 2005. p. 685-696. (In Eng.) doi: https://doi.org/10.1117/12.585783

[38] Xie G., Shen H. Robust wavelet-based blind image watermarking against geometrical attacks. *2004 IEEE International Conference on Multimedia and Expo (ICME)* (IEEE Cat. No.04TH8763). Vol. 3. IEEE Computer Society; 2004. p. 2051-2054. (In Eng.) doi: https://doi.org/10.1109/ICME.2004.1394668

[39] Venturini I. Counteracting Oracle attacks. *Proceedings of the 2004 workshop on Multimedia and security (MM&Sec'04).* Association for Computing Machinery, New York, NY, USA; 2004. p. 187-192. (In Eng.) doi: https://doi.org/10.1145/1022431.1022464

[40] Zhang X., Wang S. Invertibility attack against watermarking based on forged algorithm and a countermeasure. *Pattern Recognition Letters*. 2004; 25(8):967-973. (In Eng.) doi: https://doi.org/10.1016/j.patrec.2004.02.007

[41] Kutter M., Voloshynovskiy S.V., Herrigel A. Watermark Copy Attack. *Proceedings of SPIE: Security and Watermarking of Multimedia Contents II.* Vol. 3971. SPIE; 2000. p. 371-380. (In Eng.) doi: https://doi.org/10.1117/12.384991

[42] Srivatsan N., Barron J., Klein D., Berg-Kirkpatrick T. A Deep Factorization of Style and Structure in Fonts. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).* Hong Kong, China. Association for Computational Linguistics; 2019. p. 2195-2205. (In Eng.) doi: https://doi.org/10.18653/v1/D19-1225

[43] Campbell N.D.F., Kautz J. Learning a manifold of fonts. *ACM Transactions on Graphics*. 2014; 33(4):91. (In Eng.) doi: https://doi.org/10.1145/2601097.2601212

[44] Cu V.L., Burie J. -C., Ogier J. -M., Liu C. -L. Hiding Security Feature Into Text Content for Securing Documents Using Generated Font. *2019 International Conference on Document Analysis and Recognition (ICDAR).* IEEE Computer Society; 2019. p. 1214-1219. (In Eng.) doi: https://doi.org/10.1109/ICDAR.2019.00196

[45] Zramdini A., Ingold R. Optical font recognition using typographical features. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1998; 20(8):877-882. (In Eng.) doi: https://doi.org/10.1109/34.709616

[46] Peterson W.W., Brown D.T. Cyclic Codes for Error Detection. *Proceedings of the IRE.* 1961; 49(1):228-235. (In Eng.) doi: https://doi.org/10.1109/ JRPROC.1961.287814

[47] Bose R.C., Ray-Chaudhuri D.K. On a class of error correcting binary group codes. *Information and Control*. 1960; 3(1):68-79. (In Eng.) doi: https://doi.org/10.1016/S0019-9958(60)90287-4

[48] Smith R. An Overview of the Tesseract OCR Engine. *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007).* Vol. 2. IEEE Computer Society; 2007. p. 629-633. (In Eng.) doi: https://doi.org/10.1109/ICDAR.2007.4376991

[49] Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004; 13(4):600-612. (In Eng.) doi: https://doi.org/10.1109/TIP.2003.819861

[50] Chen G., et al. Large-Scale Visual Font Recognition. *2014 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE Computer Society; 2014. p. 3598-3605. (In Eng.) doi: https://doi.org/10.1109/CVPR.2014.460

[51] Kingma D.P., Ba J.L. Adam: A Method for Stochastic Optimization. arXiv:1412.6980v9. 2014. (In Eng.) doi: https://doi.org/10.48550/arxiv.1412.6980

[52] Canziani A., Paszke A., Culurciello E. An Analysis of Deep Neural Network Models for Practical Applications. arXiv:1605.07678v4. 2016. (In Eng.) doi: https://doi.org/10.48550/arXiv.1605.07678

Поступила 23.12.2021; одобрена после рецензирования 16.02.2022; принята к публикации 27.02.2022.
Submitted 23.12.2021; approved after reviewing 16.02.2022; accepted for publication 27.02.2022.

Об авторе:

Гуртова Кристина Сергеевна, магистрант кафедры информационной безопасности, факультет вычислительной математики и кибернетики, ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова» (119991, Российская Федерация, г. Москва, ГСП-1, Ленинские горы, д. 1), **ORCID: https://orcid.org/0000-0003-3863-6760,** kristinagurtov@yandex.ru

Автор прочитал и одобрил окончательный вариант рукописи.

About the author:

Kristina S. Gurtova, Master degree student of the Chair of Information Security, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University (1 Leninskie gory, Moscow 119991, GSP-1, Russian Federation), **ORCID: https://orcid.org/0000-0003-3863-6760,** kristinagurtov@yandex.ru

The author has read and approved the final manuscript.