DOI: 10.15514/ISPRAS-2021-33(6)-11

Система маркирования документов для проведения расследований при их утечке

[1] *Д.О. Обыденков, ORCID: 0000-0002-9296-6333 <obydenkov@ispras.ru>*
[1] *А.Ю. Якушев, ORCID: 0000-0001-6089-6505 <yakushev@ispras.ru>*
[1] *Ю.В. Маркин, ORCID: 0000-0003-1145-5118 <ustas@ispras.ru>*
[1] *С.А. Фомин, ORCID: 0000-0002-1151-2189 <fomin@ispras.ru>*
[1] *А.Е. Фролов, ORCID: 0000-0001-7616-2354 <aefrolov@ispras.ru>*
[2] *С.В. Козлов, ORCID: 0000-0003-1269-1681 <kozlov_sv@mail.ru>*
[2] *Д.Д. Громей, ORCID: 0000-0003-1066-556X <gromej@academ.msk.rsnet.ru>*
[2] *А.В. Козачок, ORCID: 0000-0002-6501-2008<a.kozachok@academ.msk.rsnet.ru>*
[3] *Б.В. Кондратьев, ORCID: 0000-0001-6348-117X <gae@mil.ru>*

[1] *Институт системного программирования РАН им. В.П. Иванникова, 109004, Россия, г. Москва, ул. А. Солженицына, д. 25*
[2] *Академия Федеральной службы охраны Российской Федерации, 302015, Россия, г. Орёл, ул. Приборостроительная, д. 35*
[3] *Министерство обороны Российской Федерации, 119160, г. Москва, ул. Знаменка, д. 19*

Аннотация. В статье представлена система расследования утечек конфиденциальных текстовых документов, где в качестве каналов утечек рассматриваются изображения, полученные путем сканирования/фотографирования напечатанных документов, а также фотографии документов на экране монитора. Таким образом, система нацелена на защиту каналов, не защищаемых традиционными DLP-решениями. В качестве механизма защиты документов выбрано внедрение цифрового водяного знака (ЦВЗ), предполагающее изменение визуального представления документа. В случае анонимной утечки конфиденциального документа ЦВЗ позволит установить сотрудника, допустившего утечку – намеренно или в результате нарушения протокола безопасности. В статье описана архитектура системы, состоящая из клиентской и серверной частей. На рабочие станции сотрудников устанавливаются компоненты, обеспечивающие внедрение ЦВЗ в документы, отправляемые на печать или выводимые на экран монитора. Факты маркирования документов регистрируются и отправляются на удаленный сервер, использующий данную информацию при расследовании утечек аналитиком службы безопасности. Разработаны алгоритмы маркирования для встраивания ЦВЗ в текстовые документы, выводимые на печать и экран монитора. При маркировании документа на экране монитора ЦВЗ встраивается в межстрочные интервалы: цифровая метка кодируется последовательностью затемненных и осветленных областей. Для встраивания ЦВЗ в печатаемые документы разработано три алгоритма маркирования: на основе горизонтального и вертикального смещения слов, а также посредством изменения яркости отдельных фрагментов слов. Разработана методика тестирования алгоритмов маркирования в условиях, приближенным к условиям эксплуатации, оценена область применимости алгоритмов. Проведен анализ вероятных атак на систему, и сформулирована модель нарушителя.

Ключевые слова: защита от утечек информации; цифровой водяной знак; слепое извлечение ЦВЗ; устойчивость к преобразованиям print-scan, print-cam, screen-cam; обработка изображений

Для цитирования: Обыденков Д.О., Якушев А.Ю., Маркин Ю.В., Фомин С.А., Фролов А.Е., Козлов С.В., Громей Д.Д., Козачок А.В., Кондратьев Б.В. Система маркирования документов для проведения

расследований при их утечке. Труды ИСП РАН, том 33, вып. 6, 2021 г., стр. 161-174. DOI: 10.15514/ISPRAS–2021–33(6)–11

Document Marking System for Leak Investigations

[1] *D.O. Obydenkov, ORCID: 0000-0002-9296-6333 <obydenkov@ispras.ru>*
[1] *A.Yu. Yakushev, ORCID: 0000-0001-6089-6505 <yakushev@ispras.ru>*
[1] *Yu.V. Markin, ORCID: 0000-0003-1145-5118 <ustas@ispras.ru>*
[1] *A.E. Frolov, ORCID: 0000-0001-7616-2354 <aefrolov@ispras.ru>*
[1] *S.A. Fomin, ORCID: 0000-0002-1151-2189 <fomin@ispras.ru>*
[2] *S.V. Kozlov, ORCID: 0000-0003-1269-1681 <kozlov_sv@mail.ru>*
[2] *D.D. Gromey, ORCID: 0000-0003-1066-556X <gromej@academ.msk.rsnet.ru>*
[2] *A.V. Kozachok, ORCID: 0000-0002-6501-2008<a.kozachok@academ.msk.rsnet.ru>*
[3] *B.V. Kondrat'ev, ORCID: 0000-0001-6348-117X <gae@mil.ru>*

[1] *Ivannikov Institute for System Programming of the Russian Academy of Sciences, 25, Alexander Solzhenitsyn st., Moscow, 109004, Russia.*
[2] *Academy of Federal Guard Service, 35, Priborostroitel'naya st., Orel, 302015, Russia*
[3] *Ministry of Defence of the Russian Federation, 19, Znamenka Str., Moscow, 119160, Russia*

Abstract. This paper presents a confidential text documents leakage investigation system, focused on leak channels by documents printing and screen photographing. Internal intruders may print confidential document, take paper copy out of protected perimeter, make document image by scanner and perform anonymous leak. Also, intruders may take a photo of printed confidential document or displayed on workstation screen using personal mobile phone. Described leakage channels are weakly covered by traditional DLP systems that are usually used by enterprises for confidential information leak protection. Digital watermark (DWM) embedding is chosen as a document protection mechanism by implying changing of document image visual representation. In case of confidential document anonymous leak embedded DWM would enable the employee to determine what leak intentionally or by security protocol violation. System architecture consists of different type components. Employees' workstation components provide DWM embedding into documents, which are sent for printing or displayed on screen. Information about watermark embedding is sent to a remote server that aggregates marking facts and provides it to security officer during investigation. Text document marking algorithms are developed, which embed DWM into printed and displayed on screen documents. Screen watermark is embedded into interline space interval, information is encoded by sequence of lightened and darkened spaces. DWM embedding into printed documents is implemented by three algorithms: horizontal and vertical shift based, font fragments brightness changing based. Algorithms testing methodology is developed in view of the production environment, that helped to evaluate the application area of algorithms. Besides, intruder model was formulated, system security was evaluated and determined possible attack vectors.

Keywords: data leakage prevention; blind watermarking methods; print-scan, print-cam, screen-cam watermarking; image processing

For citation: Obydenkov D.O., Yakushev A.Yu., Markin Yu.V., Frolov A.E., Fomin S.A., Kozlov S.V., Gromey D.D., Kozachok A.V., Kondrat'ev B.V. Document Marking System for Leak Investigations. Trudy ISP RAN/Proc. ISP RAS, vol. 33, issue 6, 2021, pp. 161-174 (in Russian). DOI: 10.15514/ISPRAS–2021–33(6)–11

1. Введение

Проблема утечки конфиденциальных документов становится все более актуальной для организаций, работающих над цифровизацией процессов. Особую значимость в организациях представляют следующие данные:

- финансовые и операционные показатели;

- информация о технологиях и процессах;
- сведения о сотрудниках, клиентах, поставщиках.

Данная информация представляет большую ценность для конкурентов организации, и утечка подобной информации может повлечь за собой значительный финансовый и репутационный ущерб организации. Все чаще причиной утечки становятся «инсайдеры» - сотрудники организации, работающие в сговоре с нарушителями внешнего характера. По данным исследования InfoWatch [1] количественное соотношение утечек по типу внешнего и внутреннего нарушителя в мире примерно равно, но в России доля утечек из-за внутренних нарушителей достигает 79% [2].

Исследователи ИБ выделяют ряд каналов утечки информации, среди которых чаще всего используется сетевой канал (79,3% от всех утечек в 2020 году). Утечки осуществляются через сеть Интернет посредством отправки конфиденциальных документов на личную почту, облачный сервис или сетевой ресурс. Для защиты от такого рода утечек применяются DLP-системы (Data Leak Protection). Решения данного типа позволяют предотвращать утечку в режиме реального времени, поскольку осуществляют непрерывный мониторинг периметра - такие системы относят к *активным*. Однако, система может ошибаться и блокировать допустимую активность пользователей, поскольку на практике сложно избежать ложноположительных срабатываний. Поэтому в ряде случаев обосновано использование пассивных DLP-систем, нацеленных на регистрацию инцидентов потенциальных утечек информации. Обнаружение конфиденциального документа может выполняться как по формальным признакам (специальным атрибутам документа), так и на основе анализа содержимого. Зачастую для поиска по содержимому задаются нечеткие критерии соответствия, как например поиск по сигнатурам или регулярным выражениям, более продвинутые методы опираются на лингвистический анализ содержимого или вычисляют цифровой отпечаток документа. И наконец, ряд систем использует OCR (Optical Character Recognition) для поиска конфиденциальных документов [3][4].

Компоненты DLP системы размещаются в различных точках периметра и соответственно берут на себя различные функции. *Агент* размещается на рабочем месте пользователя (*data-in-use*) и охватывает сразу несколько плоскостей возможных утечек. Агент контролирует потоки данных через локальные устройства ввода-вывода: копирование файлов на внешние носители, отправка документов на печать, обмен файлами через устройства Bluetooth и другие. Контроль сетевых взаимодействий также осуществляется в данной точке, поскольку позволяет эффективно контролировать данные, передаваемые через защищенные протоколы (*data-in-motion*): web, почтовые, VoIP и протоколы мессенджеров. Компоненты DLP, размещаемые на сетевом шлюзе, контролируют содержимое сетевого трафика при помощи технологии DPI (Deep Packet Inspection). К компонентам такого типа предъявляются особые требования к производительности, поскольку пропускная способность сетей в корпоративных системах может достигать десятков гигабит. Требования к производительности сильно ограничивают сложность алгоритмов анализа, однако контроль на данном уровне чрезвычайно важен, поскольку позволяет обеспечить защиту от утечек с устройств, неподконтрольных администратору сети, например, с личных мобильных устройств или устройств, используемых в рамках концепции BYOD (Bring Your Own Device). Компоненты поиска размещаются рядом с хранилищем конфиденциальной информации (*data-at-rest*) и выполняют непрерывное сканирование ресурсов организации на предмет небезопасного размещения документа, выполняя таким образом превентивную защиту от утечек. При эксплуатации таких систем особое значение имеет генерация и визуализация отчетов о событиях, необходимые для работы аналитика службы безопасности.

Лидеры рынка DLP решений совмещают в себе несколько типов компонентов и обеспечивают комплексный контроль цифровых ресурсов организации, однако актуальной остается проблема утечек с использованием бумажных копий и личных мобильных

телефонов. Распечатанный документ выходит из-под контроля системы DLP, что позволяет потенциальному "инсайдеру" вынести бумажную копию за пределы защищаемого контура и анонимно разместить скан или фотографию конфиденциального документа в сети Интернет. Также фотографии, сделанные личным мобильным телефоном с экрана рабочего компьютера, не могут отслеживаться системами DLP. На российском рынке присутствуют решения Trace Doc [5], EveryTag [6] и SafeCopy [7], нацеленные на борьбу с утечками данного типа. Однако использованные разработчиками алгоритмы маркирования требуют для расследования утечки оригинальную версию документа, что была отправлена на печать или сфотографирована с экрана. Данное требование подразумевает создание централизованного защищенного хранилища оригинальных документов и сведений о маркировании. Соответственно, если нужная версия документа отсутствует в хранилище, то извлечение метки не представляется возможным.

В данной работе рассматривается система, нацеленная на борьбу с анонимными утечками конфиденциальных документов с использованием бумажных носителей и фотографий экрана. Основной упор делается на расследование утечки постфактум и выявление внутреннего нарушителя, осуществившего утечку или нарушившего протокол безопасности, что привело к утечке информации. Согласно выбранной концепции сотруднику службы безопасности для расследования инцидента утечки достаточно изображения документа, несанкционированным образом покинувшего защищаемый периметр. Статья организована следующим образом: в первом разделе обосновывается мотивация разработки системы, второй раздел описывает архитектуру системы, третий раздел включает краткое описание разработанных алгоритмов, в четвертом выполняется анализ разработанной системы. Система маркирования состоит из множества компонентов, подробное описание наиболее значимых компонентов системы содержится в статьях [8,12,26,27]. Данная статья описывает взаимодействие между отдельными компонентами и особенности работы системы в целом.

2. Архитектура системы

В качестве ключевого принципа функционирования системы был выбран подход встраивания *цифрового водяного знака* (ЦВЗ) в документ. В документ внедряется битовая последовательность заданной длины – цифровая метка, уникальным образом идентифицирующая распространителя. Механизм выдачи цифровых меток должен позволять однозначно определить источник утечки. Процесс *внедрения* ЦВЗ в документ обозначается в статье как *маркирование* – преобразование документа согласно определенному алгоритму маркирования. Для *извлечения* метки требуется *маркированный документ*. Процесс восстановления документа к состоянию до встраивания метки или близкое к этому обозначается как *стирание* метки. Отметим, что для извлечения или стирания цифровой метки оригинальный документ не требуется. Это одно из ключевых преимуществ разработанной системы.

Компоненты системы (рис. 1) разделяются на два класса: клиентские и серверные. Первые устанавливаются на рабочую станцию пользователя и обеспечивают маркирование печатаемых документов и документов, выводимых на экран. Встраиваемая метка генерируется односторонней криптографической функцией на основе сведений о рабочей станции и учетной записи *пользователя*. Цифровая метка встраивается в документы, отправляемые на печать или отображаемые на экране монитора рабочей станции. Информация о маркировании регистрируется и передается на удаленный сервер, агрегирующий информацию о маркировании со всех рабочих станций периметра. В дальнейшем при утечке конфиденциальной информации *аналитик службы безопасности* использует специальный web-интерфейс анализа инцидента для установления виновника утечки. Аналитик определяет пользователя, допустившего утечку, и рабочую станцию, которую он использовал.

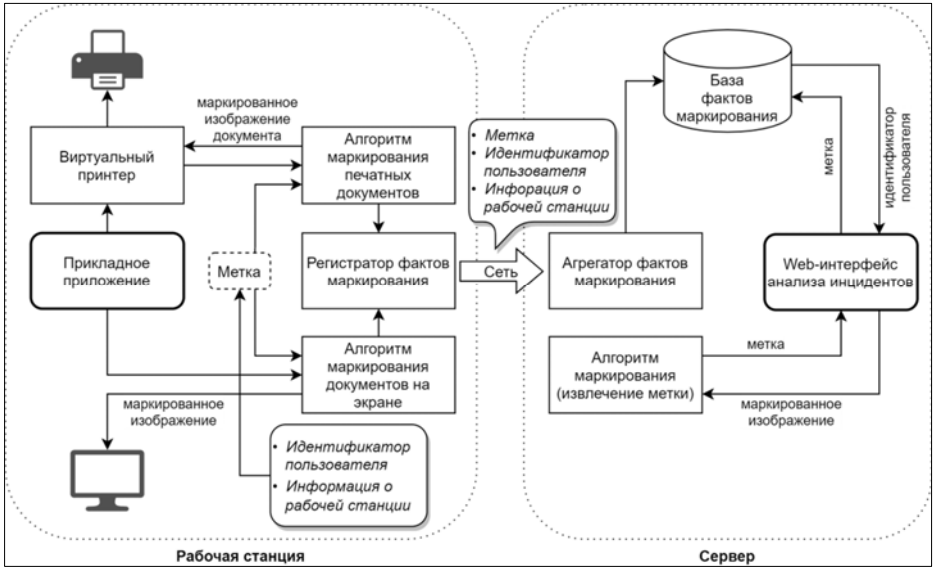

Рис. 1. Архитектура системы
Fig. 1. System architecture

Рассмотрим подробнее процессы, происходящие на рабочей станции при отправке документа на печать (рис. 2).

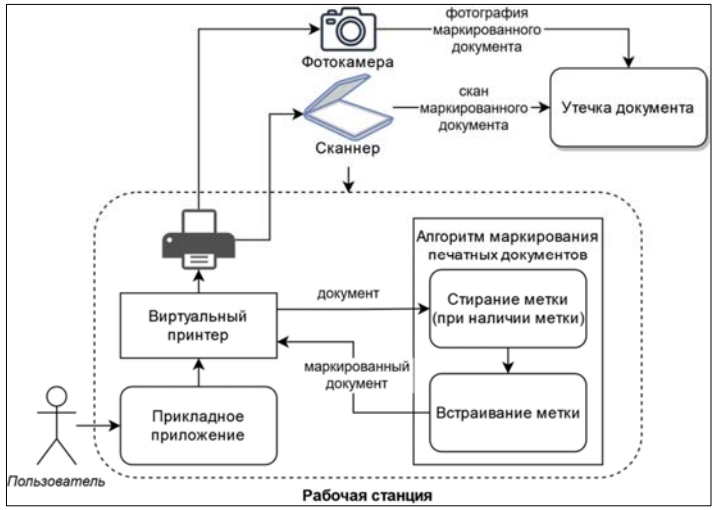

Рис. 2. Маркирование документов при печати
Fig. 2. Printed documents marking

При установке системы на рабочую станцию устанавливается и конфигурируется виртуальный принтер, который обрабатывает документы и перенаправляет их на физический принтер [8]. Виртуальный принтер берет на себя важные функции: разбивает

многостраничный документ на отдельные страницы, растеризует каждую из них, при необходимости корректирует ориентацию страницы и маркирует каждую страницу документа посредством *алгоритма маркирования*. При печати документа из прикладного приложения, как например, Word или Adobe Reader, пользователь должен выбрать виртуальный принтер как устройство печати. Соответственно, при установке системы на рабочую станцию должны применяться политики, запрещающие печать через иные устройства помимо виртуального принтера.

Перед внедрением метки выполняется стирание ранее встроенной метки, если она присутствует. Это возможно, если пользователь отправил на печать отсканированный документ с внедренной меткой. В этом случае необходимо заменить метку в документе на метку данного пользователя. В случае успешного встраивания цифровой метки страница отправляется на печать. Если происходит ошибка или встраивание метки в данную страницу невозможно, то на печать отправляется исходная (немаркированная) страница. Стратегия полного запрета печати немаркированных документов может создать значительные сложности для пользователей, а также сфокусировать их внимание на различиях между распечатанными документами и документами, отправленными на печать. Информация о маркировании – факт – сохраняется в лог-файл и отправляется на удаленный сервер. Факт маркирования включает:

- Тип маркирования (документ выведен на печать или на экран);
- Встроенная в документ цифровая метка;
- Дата и время маркирования;
- Идентификатор пользователя;
- Информация о рабочей станции:
  - MAC-адрес;
  - IP-адрес;
  - серийный номер жесткого диска.

Рис. 3. Маркирование документов, выводимых на экран
Fig. 3. Marking of documents displayed on the screen

Схожим образом построен процесс маркирования выводимой на экран информации (рис. 3). Алгоритм маркирования экрана взаимодействует с операционной системой, от которой

получает информацию о состоянии графического интерфейса, положении и содержании окон графических приложений. На основе этой информации программа маркирования экрана генерирует ЦВЗ, который накладывается поверх всех остальных графических приложений.

Сервер получает уведомления о событиях маркирования. Для экономии дискового хранилища и оптимизации поиска при проведении расследования выполняется агрегация фактов маркирования - объединение множества одинаковых событий маркирования в одно (маркирование документа на экране) путем замены отметок времени на временной интервал. Для расследования произошедшей утечки документа за пределы периметра требуется наличие изображения маркированного документа. Аналитик загружает изображение через специальный web-интерфейс, после чего запускается программа извлечения метки из изображения. По извлеченной метке выполняется поиск в базе данных, и в качестве результата аналитику выдается идентификатор пользователя и его рабочей станции.

3. Алгоритмы маркирования

3.1. Алгоритмы маркирования документов на экране

В рамках работы над системой маркирования документов, выводимых на экран монитора, были рассмотрены существующие решения внедрения цифровой метки в изображения на экране, устойчивые к искажениям, возникающим при фотографировании экрана. Первый подход [9], предложенный в 2018 году, основан на изменении яркости областей на экране. Каждому биту цифровой метки ставится в соответствие круговая область, яркость которой повышается или понижается в зависимости от значения бита. Недостатком метода является высокая заметность круговой области, если метка встраивается в одноцветное изображение. В основу двух других методов [10, 11] положена идея поиска таких признаков областей для встраивания метки, что эти признаки сохраняются на фотографии экрана. Для первого алгоритма [10] признаки областей задаются положением особых точек алгоритма I-SIFT. Для второго метода [11] поиск осуществляется с помощью детектора Харриса-Лапласа. Цифровая метка встраивается в домен преобразования найденных областей – дискретного косинусного преобразования или дискретного преобразования Фурье. Методы, внедряющие метку в домен преобразований, обладают общим недостатком: цифровая метка незаметна на изображениях, богатых цветами, но хорошо заметна на изображениях текстовых документов, как правило, черно-белых (пример в статье [12]).

Проведенный анализ существующих подходов показал необходимость разработать собственный метод встраивания цифровой метки в изображение документа, выводимого на экран монитора. Было принято решение встраивать цифровую метку путем изменения яркости областей на экране. С целью уменьшить заметность метки на одноцветных областях метод должен изменять яркость только в тех областях на экране, на которых присутствует текст.

К разрабатываемой системе маркирования выдвигается ряд требований. Метод должен быть устойчив к искажениям, возникающим при фотографировании экрана. Цифровая метка должна быть незаметна для пользователя устройства. Метка должна встраиваться в режиме реального времени. Маркироваться должен любой документ на экране независимо от его формата – текстовый файл, изображение документа и др. Извлечение цифровой метки должно проводиться в условиях отсутствия оригинального документа.

Разработанный алгоритм внедрения метки в текстовые документы [12], выводимые на экран монитора, работает следующим образом. Цифровая метка внедряется в межстрочные интервалы текста. Метка является последовательностью светлых и темных прямоугольных областей, кодирующих биты 0 и 1 соответственно. В каждый межстрочный интервал встраивается 16 бит, поэтому для встраивания цифровой метки размером 32 бита достаточно

двух межстрочных интервалов (или трех строк текста). Отображение метки производится при помощи частично прозрачного окна-оверлея, накладываемого поверх остальных окон, открытых на рабочем столе устройства. Содержимое окна-оверлея регулярно обновляется в соответствии с расположением окон с текстом на экране так, что последовательности прямоугольных областей соответствуют межстрочным интервалам маркируемого текста.

Для работы алгоритма извлечения метки требуется только фотография документа, выведенной на экран монитора. На снимке определяются межстрочные интервалы текста. В основу алгоритма извлечения метки положен факт, что цифровая камера хорошо различает плавные переходы яркости. Переходы яркости в межстрочных интервалах на фотографии соответствуют переходам яркости в цифровой метке, внедренной на маркированный документ, что позволяет определить значения битов встроенной метки. Это позволяет алгоритму извлекать цифровую метку из фотографии без изображения оригинала документа, выведенного на экран монитора. Разработанный алгоритм не только определяет значения битов встроенного сообщения, но также дает вероятностную оценку корректности полученных значений. Извлечение цифровой метки может проводится многократно с перебором параметров с целью достичь максимального значения этой оценки.

Разработанный алгоритм извлечения был протестирован на предмет устойчивости к искажениям, возникающим при фотографировании экрана. Цифровая метка успешно извлекалась при фотографировании экрана на различных расстояниях (30 см, 50 – 100 см) и углах (0º – 45º) между камерой и экраном, а также при JPEG-сжатии фотографии высокой степени, вплоть до коэффициента качества JPEG 15. Цифровая метка частично извлекалась из фотографий, сделанных на расстоянии 40 см, при углах между камерой и экраном не более 60º, а также при коэффициенте качества JPEG не менее 10.

3.2. Алгоритмы маркирования документов при печати

Задача маркирования документов при печати широко освещена во множестве публикаций. Существует большое разнообразие методов внедрения метки, опирающихся как на область преобразований (*transform domain*), так и на пространственную область (*spatial domain*). К первым относятся подходы на основе дискретного преобразования Фурье (DFT) [13], дискретно-косинусного (DCT) [14] или вейвлет преобразований (DWT). Вторая группа методов предполагает модификацию оригинальных документов с использованием информации о структуре, например, данные о местоположении слов, строк или текстовое содержимое. Эту группу можно разделить на подгруппу лингвистических методов, изменяющих семантические и/или синтаксические свойства текстового содержимого документа, а также структурные методы, изменяющие параметры визуального представления документа, но не изменяющие смысл/содержимое текста [15]. Семантические методы могут корректировать правописание, заменять слова на синонимы и аббревиатуры [16, 17]. Синтаксические методы существенно не изменяют смысл текста, а используют его свойства и особенности, как например, заменяют символы букв одного алфавита на визуально схожие символы букв другого алфавита [18]. Одни из самых ранних работ по маркированию документов использовали структурный подход и основывались на вертикальном смещении текстовых строк вверх или вниз [19, 20]. Позднее подход со смещением применялся для горизонтального смещения отдельных слов [21, 22]. Также было опубликовано большое количество работ, посвященных структурному кодированию (термины внедрение и кодирование ЦВЗ следует считать равнозначными). В них используются свойства форматирования текста: размер, цвет, особенности начертания шрифта и другие свойства: например, методы кодирования на основе высвечивания контуров символов [23] или искажения шрифтов [24]. Для арабских языков разработан метод, использующий особенности начертания букв при кодировании: сдвиги точек/увеличение длины черты в определенных словах [25].

По результатам анализа особенностей различных методов и техник внедрения цифровых меток в документы при печати было принято решение об использовании в качестве базового метода для решения поставленной задачи подход на основе структурного кодирования. Применение лингвистических методов невозможно, поскольку в результате внедрения изменяется содержимое документов. Подходы, изменяющие область преобразования, существенно ухудшают качество документов и делают внедренную цифровую метку заметной. Методы, предложенные в [23, 24], предъявляют значительные требования к разрешению и качеству изображений маркированных документов при извлечении, что ограничивает применение методов на практике.

В рамках выбранной схемы функционирования системы были сформулированы требования к алгоритму маркирования. Должны поддерживаться следующие операции обработки документа: внедрение, извлечение, стирание метки. Последняя не встречалась в публикациях по данной тематике. В то же время необходимость поддержки операции стирания метки существенно ограничивает возможные преобразования над документом при встраивании метки, поскольку в этом случае маркирование документа должно быть обратимым. Другим значимым ограничением является возможность извлечения метки из маркированного документа без оригинального документа. Также при разработке методов маркирования учитывались требования к работоспособности рассмотренных подходов при значительных искажениях, возникающих при многократной печати, сканировании и фотографировании документов.

Было разработано три алгоритма маркирования:

1) на основе горизонтального смещения слов [26];
2) на основе вертикального смещения слов [27];
3) на основе перечеркивания слов [27].

Первый алгоритм развивает идеи, положенные в методики кодирования ЦВЗ на основе смещения, представленные в работах [19-22]. Ключевая идея механизма кодирования – горизонтальное смещение слов. Документ разбивается на множество строк, строки посредством жадного алгоритма разбиваются на блоки последовательно расположенных в строке слов, разделенных четырьмя или двумя пробелами. При внедрении метки в документ слова горизонтально смещаются таким образом, чтобы величина одного из пробелов в блоке увеличилась, а величина остальных пробелов уменьшилась так, чтобы общая длина блока осталась неизменной. Позиция удлиненного пробела в блоке позволяет кодировать цифровую метку: для блоков из четырех пробелов – 2 бита, для блоков из двух пробелов – 1 бит. Стирание ранее внедренной метки выполняется одновременно с встраиванием метки и не требует дополнительных операций, однако, исходное положение слов в документе восстановлению не подлежит.

Рис. 4. Пример базовой линии и медианы слова
Fig. 4.Word baseline and median example

В основе второго алгоритма маркирования – вертикальное смещение слов. Заметим, что упомянутые ранее работы использовали для кодирования информации вертикально смещенные строки [19, 20] или горизонтально смещенные слова [21, 22]. Разработанный алгоритм имеет большую информационную емкость по сравнению с алгоритмами, использующими вертикальное смещение строк. Для кодирования одного бита информации слово вертикально смещается вверх или сохраняет исходного положение. Первое и последнее

слова в строке никогда не смещаются и задают «нулевой уровень» смещения, то есть уровень несмещенных слов. Извлечение метки выполняется построчно с использованием алгоритма Витерби, позволяющего определить наиболее вероятную последовательность смещений слов в строке. Относительное положение слов определяется путем сравнения их базовых линий (рис. 4). При стирании метки выполняется обратное вертикальное смещение слов.

В третьем алгоритме кодирования изменяется яркость отдельных фрагментов слов. Вдоль горизонтальной оси слова между базовой линий и медианой (рис. 4) выделяется прямоугольная область, и там, где она пересекает символы текста – буквы и некоторые знаки препинания, например, вопросительный/восклицательный знаки – изменяется яркость. Визуально данный эффект похож на перечеркивание слова осветляющим маркером. Обозначим данное преобразование как нанесение *перечеркивающей линии*. При внедрении цифровой метки один бит кодируется одним или несколькими словами для минимизации ошибки. Извлечение метки из документа выполняется построчно, каждое слово в строке проверяется на наличие перечеркивающей линии, для чего применяется обученная нейронная сеть на основе архитектуры U-Net [28]. Стирание метки также выполняется посредством нейронной сети, восстанавливающей оригинальное визуальное представление каждого слова по отдельности.

Для разработанных алгоритмов маркирования была разработана методика тестирования, приближенная к условиям реальной эксплуатации системы. Методика предполагает проверку работы в трех сценариях (*П* – печать, *С* – сканирование, *Ф* – фотографирование): *П-С*, *П-С-П-С*, *П-Ф*. На основе открытых источников был сформирован набор изображений документов, содержащих текст различного форматирования, изображения и таблицы. В общей сложности каждый алгоритм маркирования проходит порядка 400 тестов.

Алгоритм №1 демонстрирует наилучшую незаметность метки согласно результатам экспертной оценки. Наилучшую точность извлечения показывает алгоритм №3 (табл. 1). При оценке оценки точности извлечения метки рассчитывались следующие метки:

- *BER (Bit Error Rate)* – отношение числа неверно извлеченных бит цифровой метки к общему числу бит метки для документов, в которых присутствует достаточное для внедрения метки количество машинописного текста (70% документов);
- *Полнота* – доля извлеченных меток с $BER = 1$.

Наибольшее значение имеет метрика *полноты*, так как данная метрика позволяет оценить эффективность алгоритма при проведении расследования утечки. Помимо развития каждого из алгоритмов в отдельности, планируется проведение работ по их совместному применению.

Табл. 1. Сравнительная таблица оценки точности алгоритмов маркирования
Table 1. Comparative table of accuracy evaluation of marking algorithms

Тестовый сценарий	*Алгоритм №1*		*Алгоритм №2*		*Алгоритм №3*	
	BER-32	*Полнота*	*BER-32*	*Полнота*	*BER-32*	*Полнота*
П-С	0.8565	0.6622	0.9771	0.8378	0.9991	0.9910
П-С-П-С	0.8110	0.5930	0.9150	0.5964	0.9974	0.9728
П-Ф	0.7297	0.4247	0.8294	0.2857	0.9948	0.9277

4. Анализ атак на систему

В данном разделе рассмотрены основные угрозы и способы противодействия разработанной системе с позиции потенциального внутреннего нарушителя. Атаки на систему можно разделить на следующие категории:

- *Искажение маркированного изображения* – изображение документа перед анонимной публикацией изменяется с целью затруднить или сделать невозможным извлечение ЦВЗ;

- *Подмена метки* – замена встраиваемой в документ метки на отличную от метки пользователя;
- *Обход системы маркирования* – предотвращение внедрения ЦВЗ в документ.

Наиболее обширной областью атак является категория искажения изображения. Стоит отметить, что атаки данного типа могут иметь непреднамеренный характер, поскольку преобразование из цифрового представления (растровое изображение, готовое к печати) в аналоговое (бумажный носитель или экран монитора), а затем снова в цифровое (растровое изображение, полученное сканером или фотоаппаратом) неизбежно ведет к потере части информации. Подобная угроза учитывалась при разработке – тестирование системы показало устойчивость алгоритмов к данной атаке. Тем не менее, искажения могут вноситься в изображение намеренно, причем некоторые искажения обратимы, другие – нет. К обратимым искажениям относятся: искажение перспективы, нарушение контраста и баланса белого. К необратимым искажениям можно отнести искажения, нацеленные на уменьшение общего числа точек растрового изображения (снижение разрешения) или ухудшение цветности (конвертация цветовой схемы). Обратимые искажения аналитик может скорректировать посредством графического редактора, необратимые – ухудшают точность извлечения метки, но в то же время снижают ценность утечки.

Существует возможность утечки текстового содержания документа. Для этого нарушителю потребуется переписать документ вручную или распознать содержимое методами OCR. Трудоемкость организации такой утечки значительно выше, чем утечка изображения. В данном сценарии следует отметить, что ценность утечки может быть ничтожно малой, поскольку в переписанном вручную документе отсутствуют такие элементы, как печать или подпись.

Другая возможная атака – имитация алгоритма стирания с целью полного удаления ЦВЗ или замены метки на метку другого пользователя. Такая атака потребует обратной разработки алгоритма маркирования. Искажение имитации на изображении маркированного документа можно получить при помощи графического редактора или специально созданной программы. Помимо высокой трудоемкости процесса создания имитации, возникает вероятность неполного стирания или замены метки, поскольку на практике крайне сложно полностью скрыть следы редактирования изображения.

Категория атак, связанных с подменой метки на этапе генерации метки в системе, требует доступа к учетной записи и рабочей станции другого пользователя. Как правило, подобные атаки совершаются по неосторожности сотрудников (как пример, стикер с логином и паролем на мониторе) или методами социальной инженерии (рассылка через корпоративную почту). Реализация данной атаки усложняется тем, что требует от злоумышленника подмены не только учетной записи пользователя, но и рабочей станции.

Для реализации атак, подразумевающих обход или отключение системы маркирования, потребуется получение прав суперпользователя. Не секрет, что существует множество уязвимостей повышения прав пользователя. Однако защита от подобных угроз лежит не на разработчиках описываемой системы, а на ИБ специалистах и системных администраторах, ответственных за своевременную установку обновлений безопасности на рабочие станции периметра организации.

По результатам анализа возможных атак на систему была уточнена модель нарушителя. С бóльшей вероятностью нарушитель непреднамеренный – невнимательный или неосведомленный пользователь. Теоретически возможна атака на систему, при которой нельзя установить виновника утечки, однако трудоемкость ее реализации высока, требует от атакующего проведения исследования алгоритмов маркирования, выполняется вручную и плохо поддается автоматизации.

5. Заключение

Разработанная система маркирования документов, выводимых на печать или экран, позволяет противостоять угрозам, с которыми не способны справиться традиционные DLP–системы. Существующие на рынке решения имеют ограничения, затрудняющие внедрение решения в ряд организаций. Представленная система лишена этих ограничений: при извлечении цифровой метки не требуется оригинальный документ, а также предусмотрена функциональность внедрения цифровой метки в маркированные ранее документы. Разработанная система имеет высокую эффективность и показала свою применимость на практике.

Список литературы / References

[1]. Исследование утечек информации ограниченного доступа в 2020 году. InfoWatch. 2021, 40 стр. / Research on restricted information leaks in 2020. InfoWatch. 2021, 40 p. Available at: https://www.infowatch.ru/analytics/analitika/issledovanie-utechek-informatsii-ogranichennogo-dostupa-v-2020-godu, accessed 24.10.2021.

[2]. Утечки информации ограниченного доступа: отчет за 9 месяцев 2020 г. Экспертно-аналитический центр InfoWatch, 2020 г. / Restricted information leaks: report for 9 months of 2020. InfoWatch Analytical Center, 2020 (in Russian). Available at: https://www.infowatch.ru/analytics/analitika/utechki-informatsii-ogranichennogo-dostupa-otchet-za-9-mesyatsev-2020, accessed 24.10.2021.

[3]. McAfee Data Loss Prevention. Available at: https://docs.mcafee.com/bundle/data-loss-prevention-11.4.x-product-guide, accessed 24.10.2021.

[4]. Symantec Data Loss Prevention. Available at: https://www.broadcom.com/products/cyber-security/information-protection/data-loss-prevention, accessed 24.10.2021.

[5]. Trace Doc. Available at: https://secretgroup.ru/trace-doc, accessed 10.08.2021.

[6]. Unique Interface. EveryTag. Available at: https://everytag.ru/ui, accessed 10.08.2021.

[7]. Safe Copy. Available at: https://www.niisokb.ru/products/safecopy, accessed 10.08.2021.

[8]. Козлов С.В., Копылов С.А. и др. Реализация маркирования в подсистеме печати ОС семейства Windows на основе виртуального XPS-принтера. Труды ИСП РАН, том 32, вып. 5, 2020 г., стр. 95-110 / Kozlov S.V., Kopylov S.A. et al. Implementing watermarking based on a virtual XPS printer for Windows operating systems. rudy ISP RAN/Proc. ISP RAS, vol. 32, issue 5, 2020, pp. 95-110 (in Russian). DOI: 10.15514/ISPRAS–2020–32(5)–7.

[9]. Gugelmann D., Sommer D. et al. Screen Watermarking for Data Theft Investigation and Attribution. In Proc. of the 10th International Conference on Cyber Conflict (CyCon), 2018, pp. 391-408.

[10]. Fang H., Zhang W. et al. Screen-Shooting Resilient Watermarking. IEEE Transactions on Information Forensics and Security, vol. 14, no. 6, 2019, pp. 1403-1418.

[11]. Chen W., Ren N. et al. Screen-Cam Robust Image Watermarking with Feature-Based Synchronization. Applied Sciences, vol. 10, no. 21, 2020, article no. 7494.

[12]. Якушев А.Ю., Маркин Ю.В. и др. Маркирование текстовых документов на экране монитора посредством изменения яркости фона в областях межстрочных интервалов. Труды ИСП РАН, том 33, вып. 4, 2021 г., стр. 147-162 / Yakushev A.Yu., Markin Yu.V. et al. Text documents screen watermarking by changing background brightness in the interline spacing. Trudy ISP RAN/Proc. ISP RAS, vol. 33, issue 4, 2021. pp. 147-162 (in Russian). DOI: 10.15514/ISPRAS–2021–33(4)–11

[13]. Pramila A., Keskinarkaus A., Seppänen T. Multiple domain watermarking for print-scan and JPEG resilient data hiding. Lecture Notes in Computer Science, vol. 5041, 2007, pp. 279-293.

[14]. Dong P., Galatsanos N.P. Affine transformation resistant watermarking based on image normalization. In Proc. of the International Conference on Image Processing, 2002, pp. 489-492.

[15]. Ahvanooey M.T., Li Q.et al. Modern text hiding, text steganalysis, and applications: a comparative analysis. Entropy, vol. 21, no. 4, 2019, article no. 355.

[16]. Topkara M., Topkara U., Atallah M.J. Words are not enough: sentence level natural language watermarking. In Proc. of the 4th ACM International Workshop on Contents Protection and Security, 2006, pp. 37-46.

[17]. Topkara U., Topkara M., Atallah M.J. The hiding virtues of ambiguity: quantifiably resilient watermarking of natural language text through synonym substitutions. In Proc. of the 8th Workshop on Multimedia and Security, 2006, pp. 164-174.

[18]. Shirali-Shahreza M. A new Persian/Arabic text steganography using “La” word. In Advances in computer and information sciences and engineering, Springer, 2008, pp. 339-342.

[19]. Low S. H., Maxemchuk N. F. et al. Document marking and identification using both line and word shifting. In Proc. of INFOCOM'95, 1995, pp. 853-860.

[20]. Alattar A. M., Alattar O. M. Watermarking electronic text documents containing justified paragraphs and irregular line spacing. Security, Steganography, and Watermarking of Multimedia Contents VI, vol. 5306, 2004, pp. 685-695.

[21]. Brassil J. T., Low S. Electronic marking and identification techniques to discourage document copying. IEEE Journal on Selected Areas in Communications, vol. 13, no. 8, 1995, pp. 1495-1504.

[22]. Kim Y.W., Moon K.A., Oh I. S. A Text Watermarking Algorithm based on Word Classification and Inter-word Space Statistics. In Proc. of the Seventh International Conference on Document Analysis and Recognition, 2003, pp. 775-779.

[23]. Tan L., Hu K. et al. Print-scan invariant text image watermarking for hardcopy document authentication. Multimedia Tools and Applications, vol. 78, no. 10, 2018, pp. 13189-13211.

[24]. Xiao C., Zhang C., Zheng C. Fontcode: Embedding information in text documents using glyph perturbation. ACM Transactions on Graphics (TOG), vol. 37, no. 2, 2017, pp. 1-16.

[25]. Gutub A., Fattani M. A novel Arabic text steganography method using letter points and extensions. In Proc. of the WASET International Conference on Computer, Information and Systems Science and Engineering (ICCISSE), 2007, pp. 28-31.

[26]. Козачок А.В., Копылов С.А. и др. Алгоритм маркирования текстовых документов на основе изменении интервала между словами, обеспечивающий устойчивость к преобразованию формата. Труды ИСП РАН, том 33, вып. 4, 2021 г., стр. 131-146 / Kozachok A.V., Kopylov S.A. et al. Text documents marking algorithm based on interword distances shifting invariant to format conversion. Trudy ISP RAN/Proc. ISP RAS, vol. 33, issue 4, 2021. pp. 131-146 (in Russian). DOI: 10.15514/ISPRAS–2021–33(4)–10.

[27]. Обыденков Д.О., Фролов А.Е. и др. Методы маркирования текстовых документов при печати посредством вертикального сдвига и изменения яркости фрагментов слов. Труды ИСП РАН, том 33, вып. 5, 2021 г., стр. 65-82 / Obydenkov D. O., Frolov A.E. et al. Printed text documents watermarking based on vertical word shift and word fragments brightness changing. Trudy ISP RAN/Proc. ISP RAS, vol. 33, issue 5, 2021, pp. 65-82 (in Russian). DOI: 10.15514/ISPRAS–2021–33(5).

[28]. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation Lecture Notes in Computer Science, vol. 9351, 2015, pp. 234-241.

Информация об авторах / Information about authors

Дмитрий Олегович ОБЫДЕНКОВ – аспирант. Научные интересы: методы сокрытия и защищенной передачи информации, компьютерные сети, технологии анализа сетевого трафика.

Dmitry Olegovich OBYDENKOV is a graduate student. Scientific interests: methods for information hiding and secure transmission, computer networks, technologies of network traffic analysis.

Алексей Юрьевич ЯКУШЕВ – студент. Научные интересы: стеганография, обработка цифровых изображений, алгоритмы машинного обучения.

Aleksey Yur'evich YAKUSHEV is a student. Scientific interests: steganography, digital image processing, machine learning algorithms.

Юрий Витальевич МАРКИН – научный сотрудник, кандидат технических наук. Область научных интересов: информационная безопасность, анализ сетевого трафика, обработка изображений, алгоритмы машинного обучения.

Yury Vital'evich MARKIN is a researcher, PhD in Technical Sciences. Scientific interests: information security, network traffic analysis, image processing, machine learning algorithms.

Станислав Александрович ФОМИН – ведущий программист. Область научных интересов: теория сложности, алгоритмы дискретной оптимизации, верификация ПО, архитектура информационных систем.

Stanislav Alexandrovich FOMIN – leading programmer. Research interests: complexity theory, discrete optimization algorithms, information systems architecture.

Александр Евгеньевич ФРОЛОВ – студент-магистр. Научные интересы: стеганография, методы анонимизации сетевого трафика, компьютерные сети, методы машинного обучения.

Alexander Evgenevich FROLOV is a master student. Research interests: steganography, traffic anonymization methods, computer networks, machine learning.

Сергей Викторович КОЗЛОВ – кандидат технических наук. Сфера научных интересов: информационная безопасность, защита от несанкционированного доступа, особенности построения и функционирования операционных систем, средства и методы программирования.

Sergey Viktorovich KOZLOV – Candidate of Technical Sciences. Research interests: information security, protection from unauthorized access, construction and functioning features of operating systems, programming tools and methods.

Дмитрий Дмитриевич ГРОМЕЙ – сотрудник Академии Федеральной службы охраны Российской Федерации). Область научных интересов: общая теория систем, системы управления базами данных, машинное обучение, архитектура операционных систем, сетевые технологии.

Dmitry Dmitrievich GROMEY – employer of the Academy of Federal Security Guard Service of the Russian Federation. Research interests: general system theory, database management system, machine learning, operating system architecture, computer network.

Александр Васильевич КОЗАЧОК – доктор технических наук, доцент, сотрудник Академии Федеральной службы охраны Российской Федерации. Его научные интересы включают: информационная безопасность, защита от несанкционированного доступа, математическая криптография, теоретические проблемы информатики.

Alexander Vasilievich KOZACHOK – Doctor of Technical Sciences, Associated Professor. Employer of the Academy of Federal Guard Service. His research interests include: information security, unauthorized access protection, mathematical cryptography and theoretical problems of computer science.

Борис Владимирович КОНДРАТЬЕВ, сфера научных интересов: безопасность информации, защита информации от несанкционированного доступа и утечки по техническим каналам, построение информационных систем в защищенном исполнении, сертификация программного обеспечения по требованиям безопасности информации.

Boris Vladimirovich KONDRAT’EV, scientific interests: information security, protection of information from unauthorized access and leakage through technical channels, building information systems in a secure design, certification of software for information security requirements.