DOI: 10.15514/ISPRAS-2021-33(4)-11

Маркирование текстовых документов на экране монитора посредством изменения яркости фона в областях межстрочных интервалов

[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] *Д.О. Обыденков, ORCID: 0000-0002-9296-6333 <obydenkov@ispras.ru>*
[2] *Б.В. Кондратьев, ORCID: 0000-0001-6348-117X <gae@mil.ru>*

[1] *Институт системного программирования им. В.П. Иванникова РАН, 109004, Россия, г. Москва, ул. А. Солженицына, д. 25.*
[2] *Министерство обороны Российской Федерации, 119160, г. Москва, ул. Знаменка, д.19*

Аннотация. Утечка значительной части электронных документов происходит путем фотографирования экрана. Для расследования таких случаев применяются технологии *data leakage prevention* (DLP), в частности, внедрение цифровых водяных знаков (ЦВЗ) в изображение на экране. В статье представлен краткий обзор существующих методов внедрения ЦВЗ. Предложен подход к маркированию изображений текстовых документов, выведенных на экран. Цифровая метка внедряется в межстрочные интервалы текста путем незначительного изменения яркости. ЦВЗ неразличим для восприятия человеком, но может быть запечатлен на цифровую камеру. Разработан алгоритм извлечения цифровой метки из фотографии экрана. Алгоритм не требует изображение исходного документа для успешного извлечения ЦВЗ. Результаты тестирования показали, что цифровая метка устойчива к преобразованиям (атакам), возникающим в ходе фотографирования экрана. Предложен метод, позволяющий оценить корректность извлечения цифровой метки при отсутствии информации о встроенной цифровой метке.

Ключевые слова: защита от утечек информации; цифровой водяной знак; маркирование документов на экране монитора; слепое извлечение ЦВЗ; устойчивость к screen-cam атакам

Для цитирования: Якушев А.Ю., Маркин Ю.В., Фомин С.А., Обыденков Д.О., Кондратьев Б.В. 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. DOI: 10.15514/ISPRAS–2021–33(4)–11

Text documents screen watermarking by changing background brightness in the interline spacing

[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] *S.A. Fomin, ORCID: 0000-0002-1151-2189 <fomin@ispras.ru>*
[1] *D.O. Obydenkov, ORCID: 0000-0002-9296-6333 <obydenkov@ispras.ru>*
[2] *B.V. Kondrat'ev, ORCID: 0000-0001-6348-117X <gae@mil.ru>*

[1] *Ivannikov Institute for System Programming of the RAS, 25, Alexander Solzhenitsyn Str., Moscow, 109004, Russia*
[2] *Ministry of Defence of the Russian Federation, 19, Znamenka Str., Moscow, 119160*

Abstract. One of the most common ways documents leak is taking a picture of document displayed on the screen. For investigation of such cases data leakage prevention technologies including screen watermarking are used. The article gives short review on the problem of screen shooting watermarking and the existing research results. A novel approach for watermarking text images displayed on the screen is proposed. The watermark is embedded as slight changes in luminance into the interline spacing of marked text. The watermark is designed to be invisible for human eye but still able to be detected by digital camera. An algorithm for extraction of watermark from the screen photo is presented. The extraction algorithm doesn't need the original image of document for successful extraction. The experimental results show that the approach is robust against screen-cam attacks, that means that the watermark stays persistent after the process of taking a photo of document displayed on the screen. A criterion for watermark message extraction accuracy without knowledge about the original message is proposed. The criterion represents the probability that the watermark was extracted correctly.

Keywords: data leakage prevention; text documents screen watermarking; screen-cam robust watermarking; blind watermarking method

For citation: Yakushev A.Yu., Markin Yu.V., Fomin S.A., Obydenkov D.O., Kondrat'ev B.V. 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

1. Введение

С развитием информационных технологий происходит постоянное увеличение объемов мирового электронного документооборота. Значительная часть документов содержит конфиденциальную информацию, доступ к которой предназначается ограниченному кругу лиц. Однако компаниям приходится сталкиваться с проблемой утечки таких документов. Зачастую виновниками утечек являются сотрудники компаний, случайно или умышленно способствующие передаче конфиденциальных документов третьим лицам. Так, исследование компании InfoWatch [1] за 2020 год показало, что более 79% утечек информации в российских компаниях происходит по причине действий сотрудников, а не злоумышленников извне. Немалую роль в росте числа утечек документов сыграла пандемия вируса COVID-19 в 2020 году. Многие предприятия перевели большую часть своих сотрудников на удаленный режим работы, что усложнило возможность отслеживания оборота конфиденциальных документов. Значительная доля утечек документов осуществляются путем фотографирования или взятия скриншота экрана.

Для предотвращения утечек конфиденциальной информации используются технологии Data Leakage Prevention (DLP). DLP-решения представляют собой программно-аппаратные комплексы, ведущие мониторинг действий сотрудников, а также определенным образом реагирующие на эти действия. Специализированное ПО на рабочем месте сотрудника может вести журнал используемых приложений, блокировать доступ к сети Интернет, запрещать использование съемных USB-накопителей, записывать действия сотрудника на веб-камеру и т.д. В то же время существующие DLP-системы не позволяют предотвращать утечки,

возникающие в результате фотографирования экрана монитора, на котором отображается содержимое конфиденциального документа.

Для реализации такого рода утечек не нужно обладать специализированными знаниями — достаточно использовать возможности современных смартфонов. Цифровые камеры смартфонов позволяют быстро делать снимки высокого качества, а современные сотовые сети — отправлять снимки любому получателю.

Снимок экрана в дальнейшем может оказаться в публичном доступе. В этом случае его можно использовать для расследования утечки. Один из подходов к последующему проведению расследования состоит во встраивании дополнительной информации в изображение, выводимое на экран. Если эта информация попадет на снимок, эксперт по расследованию сможет ей воспользоваться. Полезными для расследования могут оказаться сведения об устройстве, на котором был открыт документ, учетная запись пользователя устройства, дата и время создания снимка. В совокупности с данными видеонаблюдения и записями в журналах эта информация позволит установить обстоятельства произошедшей утечки и выяснить, какой сотрудник ее допустил.

Встраивание дополнительной информации в изображение относится к технологии *цифровых водяных знаков* (ЦВЗ). ЦВЗ представляет собой незаметное для пользователя внедрение в файл дополнительной информации, которую впоследствии можно будет из этого файла извлечь. В настоящий момент технология внедрения ЦВЗ широко применяется в области защиты авторских прав цифровых данных различных форматов: изображений, аудио- и видеофайлов.

Далее будем отождествлять понятия ЦВЗ и *цифровых меток*, а процесс встраивания ЦВЗ в изображение будем называть *маркированием изображения*. Цифровые водяные знаки на изображениях могут быть видимыми или невидимыми. Видимый ЦВЗ представляет собой изображение, накладываемое на маркируемое изображение так, что ЦВЗ становится заметным для наблюдателя. Такие ЦВЗ используются с целью запрета копирования для сохранения авторских прав. Невидимые ЦВЗ обычно незаметны для невооруженного глаза, но может быть распознан соответствующим алгоритмом извлечения цифровой метки. Невидимые ЦВЗ применяются, когда необходимо скрыть от пользователя факт маркирования. Такие ЦВЗ относятся к стеганографии — способу передачи информации, при котором скрывается сам факт передачи. В данной статье рассматриваются именно невидимые ЦВЗ.

2. Сценарии использования систем внедрения ЦВЗ

Типовой сценарий использования системы, предотвращающей утечки посредством маркирования изображений, как правило состоит из трёх этапов: встраивание ЦВЗ в изображение, преобразование маркированного изображения, извлечение ЦВЗ из преобразованного маркированного изображения. Общая схема представлена на рис. 1.

На этапе встраивания используется алгоритм, формирующий изображение с ЦВЗ по исходному изображению и сообщению, внедряемому в ЦВЗ. В некоторых системах маркирования изображений используются схемы шифрования для защиты цифровой метки. В таких системах при встраивании ЦВЗ применяется криптографический ключ. Результатом работы алгоритма встраивания ЦВЗ является маркированное изображение. Если система маркирования встраивает невидимый ЦВЗ, маркированное изображение должно быть трудно отличимым от исходного, при визуальном восприятии человеком.

На втором этапе маркированное изображение подвергается преобразованиям. Будем называть атакой на цифровую метку любое преобразование изображения со встроенным ЦВЗ. Цифровая метка считается устойчивой к атаке, если после применения этой атаки

сохраняется возможность определения наличия ЦВЗ и его корректного извлечения. Атаки на ЦВЗ делятся на два класса: преднамеренные и непреднамеренные.

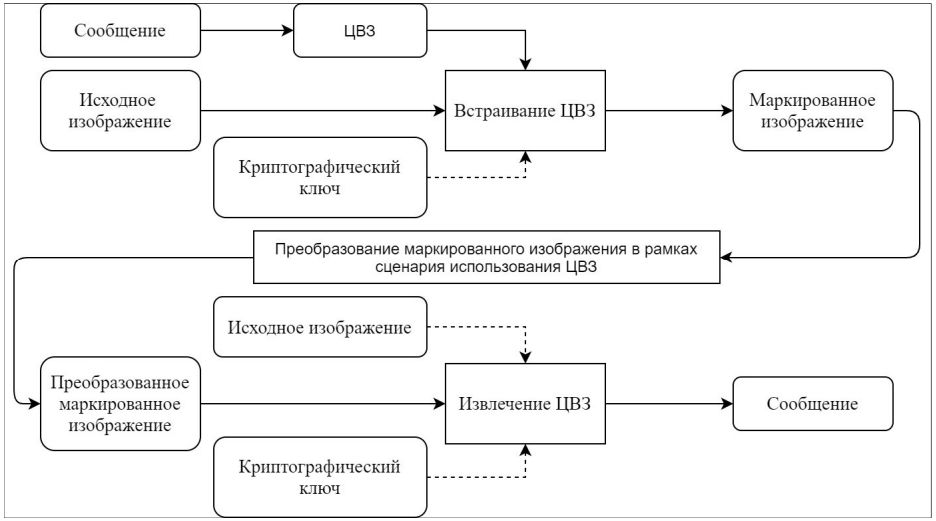


Рис. 1. Общая схема использования ЦВЗ
Fig. 1. General watermark usage scheme

Под преднамеренными атаками подразумеваются преобразования изображения человеком, знающим о наличии встроенного ЦВЗ, с целью удалить цифровую метку или изменить закодированные в ЦВЗ данные. В данной работе такие атаки рассматриваться не будут. Непреднамеренные атаки включают в себя искажения изображения, возникающие в ходе сценариев использования этого изображения. Обычно рассматриваются три сценария: сканирование напечатанных изображений («print-scan»), фотографирование напечатанных изображений («print-cam»), фотографирование изображений, выведенных на экран («screen-cam»).

На последнем, третьем этапе изображение с ЦВЗ, подвергнутое атакам, подается на вход алгоритма извлечения ЦВЗ. В зависимости от того, с какой целью применяется система маркирования изображений, результатом работы алгоритма извлечения ЦВЗ может быть:

- сообщение, внедренное в исходное изображение с помощью ЦВЗ;
- проверка корректности ЦВЗ, показывающая, было ли изображение модифицировано в ходе передачи;
- проверка факта наличия ЦВЗ.

На вход алгоритма извлечения дополнительно могут подаваться: криптографический ключ, исходное изображение, встроенный ЦВЗ. Если при извлечении ЦВЗ исходное изображение не требуется, говорят, что метод работает в режиме *слепого извлечения*.

Цифровая метка обладает рядом свойств, однако в статьях, посвященных технологии внедрения ЦВЗ, акцент главным образом делается на трех свойствах: емкость (*capacity*), незаметность (*imperceptibility*), устойчивость (*robustness*) [2]. Емкость цифровой метки означает количество битов информации, встраиваемой в исходное изображение. Незаметность цифровой метки показывает, насколько сложно человеку определить наличие ЦВЗ в маркированном изображении. Устойчивость цифровой метки характеризует сопротивляемость ЦВЗ изменениям изображения, происходящим между процессом встраивания ЦВЗ и процессом извлечения ЦВЗ. Эти три свойства цифровой метки

взаимоисключают друг друга, поэтому при разработке/применении метода маркирования необходимо достичь компромисса между ними.

3. Атаки на ЦВЗ в сценарии «screen-cam»

Изображение, полученное путем фотографирования экрана, существенно отличается от изображения, выведенного на экран. Такое различие обусловлено большим количеством непреднамеренных атак, возникающих в сценарии «screen-cam». Эти атаки можно разделить на три группы [3] в зависимости от того, в какой момент они возникают: вывод изображения на экран, фотографирование экрана монитора, обработка фотографии (рис. 2).

На качество вывода изображения на экран влияют настройки монитора: яркость, контрастность, цветопередача, гамма-коррекция.

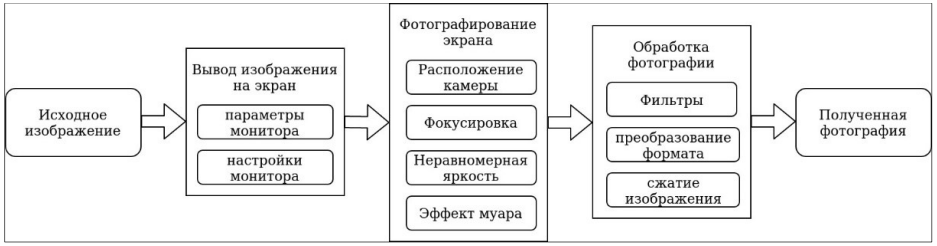


Рис. 2. Атаки на ЦВЗ в сценарии «screen-cam»
Fig. 2. Screen-cam attacks on watermark

В процессе фотографирования экрана возникают дополнительные искажения.

- Расположение камеры относительно экрана: влияет на изменение перспективы, сдвиг, изменение масштаба изображения.
- Фокусировка: если камера расположена под большим углом к плоскости экрана, разные части экрана находятся на разном расстоянии от нее, поэтому часть изображения может быть не в фокусе.
- Неравномерная яркость на фотографии: на яркость фотографии экрана влияет несколько факторов. Помимо того, что сам экран является источником света, на него может падать свет других источников. На экране также могут оказаться тени других объектов. Удаленные части экрана на фотографии будут менее яркими по сравнению с частями, расположенными близко к камере.
- Эффект муара: возникает из-за того, что пиксели экрана и сенсоры камеры расположены периодично. Плоскость экрана и плоскость матрицы камеры во время съемки обычно не строго параллельны друг другу, из-за чего на фотографии появляется нерегулярный узор, распространяющийся по всему изображению.

Полученный снимок подвергается дополнительной обработке на устройстве. Современные смартфоны в процессе обработки применяют к фотографии ряд изменяющих ее фильтров. Также производится преобразование формата и сжатие изображения (например, алгоритмом JPEG). Сжатие уменьшает размер снимка в памяти устройства, но негативно влияет на качество изображения. Оно может быть выполнено автоматически при пересылке фотографии в сети Интернет.

4. Методы маркирования, устойчивые к «screen-cam» атакам

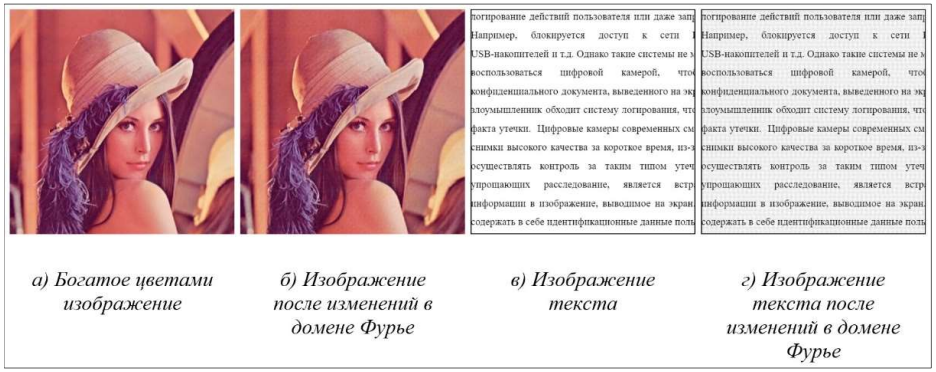
В настоящий момент известно большое число подходов к маркированию изображений. В то же время большинство из них предназначены для использования в сценариях «print-scan» или «print-cam» и требуют значительных изменений для успешной работы в сценарии «screen-

cam». Далее будут рассмотрены методы маркирования изображений, показавших высокую устойчивость к «screen-cam» атакам на ЦВЗ.

В основу метода, описанного в статье [4], положено дискретное косинусное преобразование. Встраивание цифровой метки осуществляется в два этапа. На первом этапе определяются области встраивания ЦВЗ. Для этого используются ключевые точки, полученные с помощью алгоритма I-SIFT [5]. Около каждой ключевой точки выделяется прямоугольная область, в которую внедряется цифровая метка. Кодируемая информация встраивается в домен дискретного косинусного преобразования каждой области. Множественное встраивание применяется для того, чтобы после съемки на камеру была возможность извлечь ЦВЗ из области, наименее пострадавшей от атак на ЦВЗ. В процессе извлечения цифровой метки с фотографии алгоритмом I-SIFT (как и при встраивании) определяются ключевые точки, расположение которых устойчиво к съемке. Цифровая метка извлекается из областей около ключевых точек, определенных на фотографии.

В работе [3] применяется схожая идея поиска наиболее подходящих областей для встраивания ЦВЗ. Центрами этих областей являются точки, определенные модифицированным детектором Харисса-Лапласа [6]. Для каждой такой точки определяется вектор, задающий поворот и размер квадрата, выступающего в качестве области для встраивания. Цифровая метка внедряется в домен дискретного преобразования Фурье найденных квадратных областей с использованием псевдослучайной последовательности, расположенной на окружности, соответствующей средним частотам домена Фурье. Положение центров квадратов, направление и размер векторов, задающих эти квадраты (длину стороны и угол поворота), а также средние частоты в домене Фурье обладают высокой устойчивостью к атакам сценария «screen-cam», что позволяет успешно извлекать встроенный ЦВЗ.

Рассмотренные методы обладают общей чертой: цифровая метка, внедряемая в изображение, встраивается в домене преобразований. В статье [7] также предлагается подход, выполняющий встраивание ЦВЗ в домене преобразования Фурье. Цифровая метка, встроенная в домене преобразования, создает характерный шум на изображении. Этот шум незаметен на богатых цветами и деталями изображениях, но хорошо различим на однотонных изображениях, в частности, на текстовых документах. В работе [4] показано, что предложенный в [4] ЦВЗ, встроенный в домене дискретного косинусного преобразования, хорошо заметен на изображении текстового документа. Похожий эффект возникает после внесения изменений в домене Фурье изображения текста на белом фоне (рис. 3).



а) Богатое цветами изображение | *б) Изображение после изменений в домене Фурье* | *в) Изображение текста* | *г) Изображение текста после изменений в домене Фурье*

Рис. 3. Сравнение изображений после внесения изменений в домене Фурье
Fig. 3. Images comparison after changes in DFT domain

Метод, предложенный в статье [8], предназначен для маркирования любых изображений, выведенных на экран, в том числе текстовых документов. Идея подхода основана на том, что

зрительная система человека слабо восприимчива к небольшому непрерывному изменению яркости, в то время как камера способна его различить. Цифровая метка встраивается путем уменьшения или увеличения яркости областей на экране. Маска яркости, накладываемая на изображение на экране, рассчитывается заранее и зависит от кодируемой информации, но не от содержимого изображения на экране. Такой подход позволяет использовать маску в режиме реального времени, не расходуя при этом вычислительные ресурсы системы на расчет ЦВЗ. Авторам удалось добиться работы метода в режиме слепого извлечения. Каждому биту информации сопоставляется круговой маркер в накладываемой маске. Область в центре круга делается ярче или темнее чем область у его границы – в зависимости от значения кодируемого им бита. Незаметность ЦВЗ достигается за счет плавного изменения яркости в круге. Однако маркеры остаются заметными на белом фоне (наиболее распространенном для документов). В ходе извлечения ЦВЗ из фотографии экрана определяется положение кругов и для каждого круга вычисляется разность яркости в центре и на границе.

5. Разработка метода маркирования текстовых документов на экране

Проведенный анализ существующих методов маркирования текстовых документов указывает на необходимость разработки собственных алгоритмов внедрения и извлечения ЦВЗ. Под текстовым документом далее подразумевается изображение, содержащее несколько строк текста, расположенных периодично на однотонном фоне (как правило, черный текст на белом фоне). При этом строки текста могут быть ориентированы не горизонтально, а под некоторым углом.

На основе анализа существующих методов маркирования были выделены следующие требования к системе:

- Цифровая метка в документе должна быть незаметной для пользователя.
- ЦВЗ должен встраиваться в документ в режиме реального времени: это значит, что метка должна соответствовать содержимому изображения на экране в момент получения снимка экрана.
- ЦВЗ должен быть устойчивым к атакам, возникающим в сценарии «screen-cam».
- Цифровая метка должна содержаться во всех текстовых документах, отображенных на экране, вне зависимости от формата файла документа и приложения, посредством которого осуществляется работа над документом (например, документы Microsoft Word, документы в формате PDF, изображения сканированных документов).
- Цифровая метка должна извлекаться в случае, когда на фотографии запечатлена только часть экрана, содержащая текстовый документ.
- Алгоритм извлечения ЦВЗ должен быть способен извлекать метку, получив на вход только фотографию экрана (режим слепого извлечения).

5.1 Структура цифровой метки

Исходя из требований к ЦВЗ, был разработан подход к маркированию текстовых документов, выводимых на экран монитора. Емкость сообщения, внедряемого в ЦВЗ, составляет 32 бита. Цифровая метка встраивается в межстрочные интервалы маркируемого текста, в качестве последовательности светлых и темных прямоугольных областей (рис. 4). Будем называть эти области *маркерами*. В начало, середину и конец межстрочного интервала встраиваются специальные двойные маркеры, состоящие из двух частей разных цветов (выделены синей рамкой на рисунке), причем маркер середины «противоположен» маркерам начала и конца.

Эти маркеры используются для определения положения маркеров, кодирующих биты сообщения, в процессе извлечения ЦВЗ из фотографии экрана.

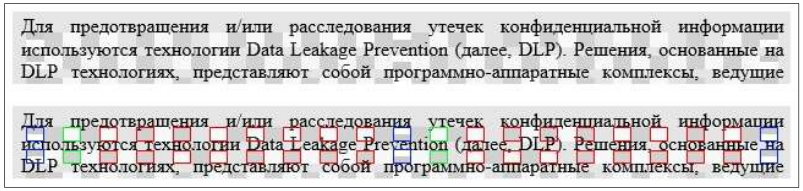


Рис. 4. Структура цифровой метки
Fig. 4. Watermark structure

Каждому биту сообщения ставится в соответствие кодирующий его маркер (выделены красной рамкой на рисунке). Темные маркеры кодируют «1», светлые – «0». Внедряемое сообщение делится на 4 части по 8 бит. Каждая четверть сообщения встраивается межу соседними двойными маркерами.

Перед четвертью сообщения добавляется дополнительный бит (маркер, выделенный зеленой рамкой), указывающий, к какой половине сообщения относится следующая за ним часть сообщения. Таким образом, двойной маркер и дополнительный бит, предшествующие последовательности из 8 маркеров, кодирующих биты сообщения, однозначно определяют, к какой четверти сообщения относится эта последовательность.

Между каждой парой маркеров, кодирующих биты сообщения, добавляется промежуточный маркер. Если соседние значимые маркеры одного цвета, промежуточный маркер между ними заполняется цветом, дополняющим до белого, а если разного — средним значением цветов соседних значимых маркеров. Такое чередование цветов используется в процессе извлечения метки.

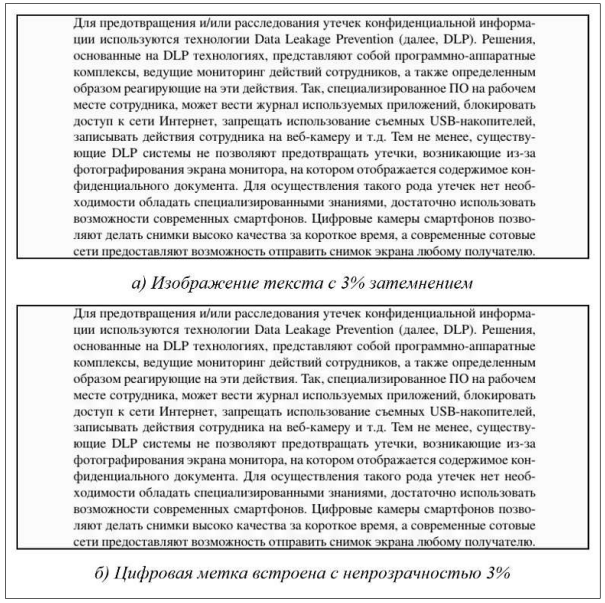


Рис. 5. Пример маркированного текста
Fig. 5. Example of watermarked text image

Для повышения незаметности метки переходы цветом между маркерами сглаживаются посредством фильтра Гаусса, размер ядра которого подбирается исходя из размера (высоты и ширины) межстрочного интервала. Поскольку фон текстовых документов преимущественно белого цвета, общая яркость изображения понижается так, чтобы разность яркости между светлыми маркерами и белым фоном совпадала с разностью яркости между белым фоном и темными маркерами. При этом цифровая метка становится менее заметной на белом фоне. На незаметность цифровой метки влияет интенсивность маркеров, задаваемая уровнем непрозрачности. Так, при уровне непрозрачности 0% ЦВЗ будет отсутствовать, а при непрозрачности 100% темные маркеры будут абсолютно черными, а светлые – белыми (рис. 5). При промежуточных значениях непрозрачности цвет маркера будет зависеть от фона текста.

Предложенная структура ЦВЗ имеет ряд преимуществ.

- Цифровая метка, встроенная в межстрочный интервал, не влияет на качество отображаемого текста.
- Плавное изменение яркости слабо заметно для восприятия, но хорошо различимо цифровой камерой.
- Для полного встраивания 32 битного сообщения достаточно трех строк текста.
- Если текст состоит из большого числа строк, метка может быть встроена несколько раз, а ЦВЗ может быть извлечен из фотографии только части текста.
- Допустимы отклонения высоты определенного межстрочного интервала от его фактической высоты.
- Поскольку информация, встроенная в один межстрочный интервал, однозначно соответствует части сообщения и не зависит от других межстрочных интервалов, допустимы ошибки определения межстрочных интервалов на этапах встраивания/извлечения (разбиение межстрочного интервала на несколько интервалов, объединение межстрочных интервалов и строки между ними в один межстрочный интервал, определение ложных межстрочных интервалов).

5.2 Алгоритм встраивания цифровой метки

Для наложения маски с областями измененной яркости на изображение на экране используется подход, описанный в [9]: создается окно, обладающее свойствами частичной визуальной прозрачности и «прозрачности» нажатия клавиш. Это окно всегда находится на вершине стека отображаемых окон, что позволяет встраивать ЦВЗ в любой момент времени. Далее будем называть это окно оверлей. С помощью оверлея можно отобразить любое изображение, содержащее 4 канала: 3 цветовых канала красного, зеленого и синего цветов, а также альфа-канал, задающий непрозрачность каждого пикселя изображения в оверлее. Итоговые значения цветовых каналов изображения, выводимого на экран, получаются как линейная комбинация изображения на оверлее и изображения, составленного другими окнами.

Положение окон на экране, а также их содержимое меняется в зависимости от действий пользователя. Содержимое оверлея необходимо регулярно обновлять, чтобы цифровая метка соответствовала измененному изображению. Процесс встраивания ЦВЗ состоит из следующих этапов:

1) получение списка окон и их положения в стеке;
2) получение скриншотов видимых частей окон, подлежащих маркированию;
3) определение областей с текстом на полученных изображениях;
4) определение углов поворота текстовых областей;
5) определение положения межстрочных интервалов;

6) отображение цифровой метки на оверлее в областях, соответствующие межстрочным интервалам нижележащего маркируемого текста.

На первом этапе посредством оконного менеджера ОС определяется список открытых окон, их положение в стеке окон, координаты прямоугольников окон. Из этого списка исключаются окно-оверлей, а также окна, заведомо не содержащие текст. На основе полученной информации определяется, какие окна видны пользователю. Для дальнейшей обработки получаются скриншоты видимых частей окон.

Второй этап – определение областей с текстом на полученных скриншотах. Для этого используется предварительно обученная нейронная сеть. В основу нейронной сети положена архитектура U-Net [10]. С целью ускорения работы алгоритма было сокращено число слоев нейронной сети, а также число каналов в промежуточных слоях.

Угол поворота текстовых областей определяется при помощи преобразования Хафа [11], применяемого для идентификации прямых на изображении. Направление линий, найденных таким образом на изображении текста, преимущественно совпадает с направлением строк текста. Это позволяет достаточно быстро определить угол наклона текста с достаточно высокой точностью.

Области межстрочных интервалов определяются по особым точкам алгоритма FAST [12]. Метод схож с методом определения горизонтального профиля страницы, описанным в [13]. При подсчете горизонтального профиля учитываются не черные пиксели, а особые точки. Такой подход позволяет определять области межстрочных интервалов вне зависимости от цвета текста и цвета фона.

5.3 Алгоритм извлечения цифровой метки

В сценарии «screen-cam», для которого была разработана предлагаемая система, изображение маркированного документа выводится на экран монитора, после чего экран фотографируется. Полученный снимок подвергается процедуре извлечения цифровой метки. Извлечение ЦВЗ проводится в несколько этапов:

1) коррекция перспективы и обрезка фотографии экрана;
2) определение областей с текстом на фотографии;
3) определение межстрочных интервалов текста;
4) поиск двойных маркеров в межстрочных интервалах;
5) извлечение битов сообщения, закодированных в маркерах ЦВЗ.

В задаче расследования утечек текстовых документов время и вычислительные ресурсы, затрачиваемые на процесс извлечения цифровой метки, имеют меньшее значение, чем точность извлеченного сообщения. Это позволяет выполнять извлечение ЦВЗ многократно с подбором параметров с целью получения наиболее успешных результатов. Некоторые этапы извлечения могут быть проведены как автоматически, так и вручную. К ним относятся этапы 1–3. При фотографировании экрана камера может быть расположена под углом к плоскости экрана, поэтому на первом этапе извлечения необходимо произвести коррекцию перспективы и обрезку фотографии. На следующих двух этапах на скорректированной фотографии определяются области с текстом и межстрочные интервалы. На этих этапах может быть применен такой же метод, что и при встраивании ЦВЗ.

Для определения положения цифровой метки в межстрочном интервале производится поиск двойных маркеров. При этом учитывается, что верхняя и нижняя половины всех маркеров, кроме двойных, одинакового цвета. Путем сравнения нижней и верхней половин межстрочного интервала на фотографии определяется, в каких точках межстрочного интервала расположены маркеры начала, середины и конца ЦВЗ. По ним можно точно определить координаты центров маркеров, кодирующих биты сообщения.

В процессе извлечения отдельных битов сообщения используется тот факт, что соседние с кодирующим промежуточные маркеры отличаются от него цветом. Функция яркости от горизонтальной координаты в межстрочном интервале на фотографии имеет локальные минимумы в темных маркерах, кодирующих бит «1», и локальные максимумы в светлых маркерах, кодирующих бит «0». Характер экстремума можно определить по знаку второй производной: положительное значение для локального минимума и отрицательное для локального максимума. Пусть $I(x, y)$ — функция яркости скорректированной фотографии экрана, (x_i, y_i) — координаты центра маркера i-го бита сообщения $M = m_1 \dots m_N$. Значение извлекаемого бита определяется по правилу:

$$s_i = \frac{\partial^2}{\partial x^2} I(x_i, y_i),$$

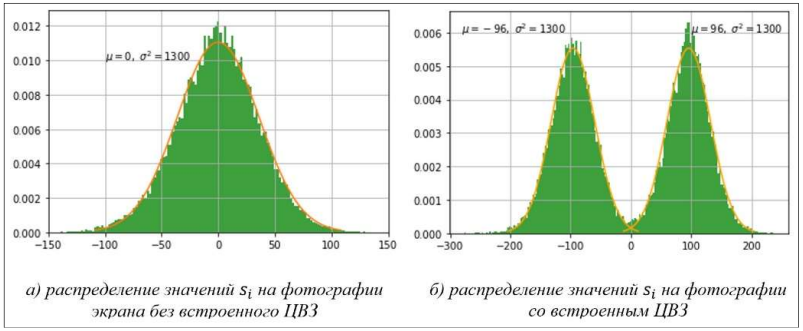
$$m_i = \begin{cases} 0, \text{если } s_i < 0 \\ 1, \text{иначе} \end{cases}.$$

5.4 Алгоритм извлечения цифровой метки

Обычно в процессе извлечения ЦВЗ отсутствует возможность сравнения извлеченного сообщения с встроенным посредством ЦВЗ. В таком случае, если результаты, полученные в ходе извлечения с различными параметрами, не совпадают, невозможно определить, какое из извлеченных сообщений содержит меньше ошибок по отношению к встроенному сообщению. В данном подразделе предложена метрика оценки корректности результата работы алгоритма извлечения цифровой метки. Её построение основано на следующих предположениях.

- $s_i, i = 1 \dots N$ — независимые одинаково распределенные случайные величины;
- Значения второй производной в точках межстрочного интервала на фотографии без встроенного ЦВЗ удовлетворяют нормальному распределению $\mathcal{N}(0, \sigma^2)$;
- Если i-ый бит m_i встроенного сообщения равен 0, что равносильно встраиванию светлого маркера, то значение второй производной в центре соответствующего маркера смещается на константу $-\mu, s_i \sim \mathcal{N}(-\mu, \sigma^2)$. Если же бит сообщения равен 1, то $s_i \sim \mathcal{N}(\mu, \sigma^2)$.

Приведенные предположения были проверены экспериментально. Результаты эксперимента представлены на рис.6.



а) распределение значений s_i на фотографии экрана без встроенного ЦВЗ

б) распределение значений s_i на фотографии со встроенным ЦВЗ

Рис. 6. Распределение значений s_i на фотографии экрана
Fig. 6. Experimental distribution of s_i

По фотографии экрана, поданной на извлечение, определяется выборка значений $\widehat{s_i}, i = 1 \dots N$. По этой выборке можно оценить значения μ и σ^2. Для всех i производится выбор из двух гипотез:

$$\begin{cases} H_0: s_i \sim \mathcal{N}(-\mu, \sigma^2), m_i = 0 \\ H_1: s_i \sim \mathcal{N}(\mu, \sigma^2), m_i = 1 \end{cases}.$$

Решающее правило строится на основе равноценности гипотез, а именно равенстве ошибок первого и второго рода (минимаксное решающее правило). В силу симметрии распределений относительно 0:

$$m_i = \theta(\widehat{s_i}) = \begin{cases} 0, f(\widehat{s_i}|H_0) > f(\widehat{s_i}|H_1) \\ 1, \text{иначе} \end{cases} = \begin{cases} 0, \widehat{s_i} < 0 \\ 1, \text{иначе} \end{cases},$$

где

$$f(x|H_0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x+\mu)^2}{2\sigma^2}},$$

$$f(x|H_1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Для каждого полученного бита сообщения можно оценить вероятность, с которой выбранная гипотеза справедлива. По формуле Байеса, при условии $\mathbb{P}(H_0) = \mathbb{P}(H_1)$:

$$\mathbb{P}(H_0|\widehat{s_i}) = \frac{f(s_i|H_0)\cdot\mathbb{P}(H_0)}{f(s_i|H_0)\cdot\mathbb{P}(H_0)+f(s_i|H_1)\cdot\mathbb{P}(H_1)} = \frac{f(s_i|H_0)}{f(s_i|H_0)+f(s_i|H_1)}.$$

Определим вероятности p_i, соответствующие решающему правилу θ:

$$p_i = \max\{\mathbb{P}(H_0|\widehat{s_i}), \mathbb{P}(H_1|\widehat{s_i})\}.$$

Из предположения о независимости s_i, получаем вероятность того, что сообщение было извлечено верно:

$$\mathbb{P}(m_1, \cdots, m_n|\widehat{s_1}, \cdots, \widehat{s_N}) = \prod_{i=1}^{N} p_i.$$

Полученная вероятность позволяет оценить, насколько точно прошло извлечение ЦВЗ: если значение близко к 1 (0.9 и выше), можно считать, что извлеченное сообщение совпадает с сообщением, встроенным при помощи ЦВЗ. Приведенная оценка была получена экспериментально. Если значение вероятности меньше 0.9, извлечение ЦВЗ следует считать неудачным: рекомендуется изменить параметры алгоритма и повторить процедуру извлечения цифровой метки.

6. Тестирование разработанного метода маркирования

Для проверки устойчивости предложенного метода маркирования текстовых документов к атакам сценария «screen-cam» был проведен ряд экспериментов по встраиванию цифровой метки в изображение документа на экране и извлечению этой метки из фотографии экрана. Цифровая метка встраивалась в текст, состоящий из 15 строк, что соответствует 14 межстрочным интервалам. Кегль шрифта 14 пт, множитель межстрочного интервала 1.15, масштаб текста 100%. Текст отображался на экране монитора Acer VG272U диагональю 27 дюймов, с разрешением 2560×1440 пикселей, типом матрицы IPS. Снимки экрана выполнялись на камеру смартфона Samsung Galaxy S8, обладающую характеристиками: 12 мегапикселей, апертура f/1.7, фокусное расстояние 26 мм. В межстрочные интервалы текста внедрялась цифровая метка, состоящая из 32 бит. Всего метка была встроена 7 раз, в пары подряд идущих интервалов. Фотографии подвергались процедуре извлечения метки. Извлеченная метка сравнивалась со встроенной на экран.

Для оценки извлекаемости ЦВЗ применяется мера Bit Error Rate (BER). BER вычисляется как отношение числа неверно извлеченных битов к общему числу битов цифровой метки. Если при извлечении эта величина равна нулю, это означает, что ЦВЗ был извлечен без ошибок. Извлекаемость ЦВЗ сравнивалась тремя оценками: оценкой BER-32, вычисляемой после объединения значений цифровых меток, встроенных в разные межстрочные интервалы, оценкой BER-224, подсчет которой проводился при предположении, что цифровые метки, встроенные в разные межстрочные интервалы, независимы, и оценкой вероятности $\mathbb{P}(M|\widehat{s_1}, \cdots, \widehat{s_{32}})$, рассчитываемой без учета знания встроенного сообщения. Если значение

последней оценки близко к 1 в случае, если извлечение проходит успешно, и близко к 0, если сообщение извлечено с ошибками, это означает, что она применима для принятия решения об успешности извлечения в условиях отсутствия информации о встроенном сообщении.

В каждом эксперименте было сделано по 10 фотографий. В приведенных ниже таблицах указано число фотографий, на которых цифровая метка была обнаружена, а также средние значения оценок по всем таким фотографиям.

6.1 Непрозрачность маркеров

Ранее отмечалась необходимость достижения компромисса между устойчивостью цифровой метки к атакам и ее незаметностью. На незаметность ЦВЗ оказывает влияние непрозрачность маркеров: чем больше непрозрачность маркеров, тем они более заметны. При уменьшении непрозрачности маркеров снижается заметность цифровой метки, но при этом снижается и различимость маркеров на фотографии на этапе извлечения. В табл. 1 представлены результаты эксперимента по изменению непрозрачности маркеров цифровой метки. В этом эксперименте камера располагалась на расстоянии 60 см от экрана параллельно его плоскости. В следующих экспериментах маркеры метки встраивались с непрозрачностью 3%.

Табл. 1. Извлекаемость цифровой метки в зависимости от непрозрачности маркеров
Table 1. Extraction rate with different watermark opacity

Непрозрачность маркеров	Метка обнаружена	BER-32	BER-224	$\mathbb{P}(\boldsymbol{M}\|\widehat{\boldsymbol{s_1}},\cdots,\widehat{\boldsymbol{s_{32}}})$
6%	10/10	0%	0%	1
5%	10/10	0%	0.2%	1
4%	10/10	0%	0.8%	1
3%	10/10	0%	2.3%	1
2%	6/10	0.5%	7%	0.94
1%	0/10	-	-	-

6.2 Расположение камеры относительно экрана

Предложенный метод был протестирован на устойчивость к атакам пространственного расположения камеры относительно экрана. Было проведено два эксперимента. В первом эксперименте изучалось влияние расстояния от камеры до экрана на извлекаемость цифровой метки при расположении камеры параллельно плоскости экрана. Результаты представлены в табл. 2. Метка извлекается с ошибками, когда камера расположена на расстояниях 40 см и 25 см от экрана. Это связано с тем, что на таких расстояниях эффект муара оказывает большое влияние на извлекаемость ЦВЗ.

Табл. 2. Влияние расстояния от камеры до экрана на извлекаемость ЦВЗ
Table 2. Extraction rate with different capture distances

Расстояние, см	Метка обнаружена	BER-32	BER-224	$\mathbb{P}(\boldsymbol{M}\|\widehat{\boldsymbol{s_1}},\cdots,\widehat{\boldsymbol{s_{32}}})$
100	10/10	0.6%	9.6%	0.7
80	10/10	0%	6.9%	1
60	10/10	0%	2.3%	1
50	10/10	0%	2.5%	1
40	9/10	3.3%	13%	0.61
30	10/10	0%	1.9%	1
25	8/10	12%	24%	0.27

Во втором эксперименте снимки экрана были сделаны под углом к плоскости экрана на фиксированном расстоянии. С целью повышения точности оценки влияния величины угла, съемка проводилась на расстоянии 50 см от камеры до центра экрана. Результаты представлены в табл. 3. Цифровая метка успешно извлекается при углах не более 45º, а также сохраняет возможность частичного извлечения вплоть до угла 60º между камерой и плоскостью экрана.

Табл. 3. Извлекаемость ЦВЗ при разных углах между камерой и плоскостью экрана
Table 3. Extraction rate with different capture angles

Величина угла	Метка обнаружена	BER-32	BER-224	$\mathbb{P}(\boldsymbol{M}\|\widehat{\boldsymbol{s_1}},\cdots,\widehat{\boldsymbol{s_{32}}})$
0º	10/10	0%	2.3%	1
15º	10/10	0%	2.5%	1
30º	10/10	0%	5.4%	0.89
45º	10/10	0.9%	7.6%	0.89
60º	10/10	1.5%	10%	0.64
75º	4/10	7.5%	24%	0.17

6.2 Сжатие фотографии

Выше упоминалось, что в сценарии «screen-cam» фотографии экрана могут подвергаться воздействию алгоритмов сжатия.

Одним из распространенных алгоритмов сжатия изображений является алгоритм JPEG.

Был проведен эксперимент по проверке предложенного метода на устойчивость сжатию изображения.

Фотографии экрана, полученные при расположении камеры на расстоянии 60 см параллельно плоскости экрана, подвергались разной степени сжатия по алгоритму JPEG. Результаты представлены в табл. 4. Предложенный метод показывает высокую устойчивость к атаке сжатия фотографии.

Табл. 4. Влияние степени сжатия JPEG на извлекаемость ЦВЗ
Table 4. Extraction rate with different JPEG compression

Коэффициент качества JPEG	Метка обнаружена	BER-32	BER-224	$\mathbb{P}(\boldsymbol{M}\|\widehat{\boldsymbol{s_1}},\cdots,\widehat{\boldsymbol{s_{32}}})$
100	10/10	0%	2.1%	1
60	10/10	0%	2%	1
40	10/10	0%	3.2%	1
20	10/10	0%	7.6%	1
15	10/10	0%	9.9%	0.98
10	8/10	0.4%	22%	0.82

7. Заключение

В статье представлен метод маркирования текстовых документов, выводимых на экран монитора – алгоритмы встраивания и извлечения ЦВЗ. Описан подход, позволяющий оценить точность извлеченной цифровой метки при отсутствии встроенного сообщения. В ходе проведенного тестирования разработанного метода оценивалась точность ЦВЗ при различных параметрах: заметность, положение камеры, степень JPEG-сжатия. Результаты экспериментов показали, что разработанный метод устойчив к основным атакам сценария «screen-cam» и может использоваться на практике.

Список литературы / References

[1]. 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).
[2]. Pramila A. Reading watermarks with a camera phone from printed images. PhD Thesis. University of Oulu, Oulu, 2018, 86 p.
[3]. 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.
[4]. 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.
[5]. Song J., Lu X., Wang W., Chen C. ISIFT: Improving the Performance of SIFT for Mirror Images. In Proc. of the 2nd IEEE International Conference on Computer and Communications (ICCC), 2016, pp. 742-745.
[6]. Mikolajczyk K., Schmid C. An Affine Invariant Interest Point Detector. Lecture Notes in Computer Science, vol. 2350, 2002, pp. 128–142.
[7]. Chen W., Ren N. et al. Joint Image Encryption and Screen-Cam Robust Two Watermarking Scheme. Sensors, vol. 21, no. 3, 2021, article no. 701.
[8]. 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.
[9]. Piec M., Rauber A. Real-Time Screen Watermarking Using Overlaying Layer. In Proc. of the Ninth International Conference on Availability, Reliability and Security, 2014, pp. 561-570.
[10]. 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.
[11]. Hough P.V.C. Method and Means for Recognizing Complex Patterns. U.S. Patent 30696541962, 1992.
[12]. Mathur G., Rikhari M.S. Text Detection in Document Images: Highlight on Using FAST Algorithm. International Journal of Advanced Engineering Research and Science, vol. 4, no. 3, 2017, pp. 275–284.
[13]. Low S., Brassil J.T., Maxemchuk N.F., O'Gorman L. Document Marking and Identification Using Both Line and Word Shifting. In Proc. of the INFOCOM'95, 1995, pp. 853-860.

Информация об авторах / Information about authors

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.

Дмитрий Олегович ОБЫДЕНКОВ – аспирант. Научные интересы: методы сокрытия и защищённой 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.

Борис Владимирович КОНДРАТЬЕВ – сотрудник МО РФ. Сфера научных интересов: безопасность информации, защита информации от несанкционированного доступа и утечки по техническим каналам, построение информационных систем в защищённом исполнении, сертификация программного обеспечения по требованиям безопасности информации.

Boris Vladimirovich KONDRAT'EV is an employee of the Ministry of Defence of the Russian Federation. 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.