

Якушев Алексей Юрьевич

**Методы и средства защиты медиаконтента на основе
цифровых водяных знаков**

Специальность 2.3.5 —
«Математическое и программное обеспечение вычислительных систем,
комплексов и компьютерных сетей»

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте системного программирования им. В.П. Иванникова Российской Академии Наук.

Научный руководитель: кандидат технических наук
Маркин Юрий Витальевич

Официальные оппоненты: **Грибунин Вадим Геннадьевич**,
доктор технических наук, доцент,
Автономная некоммерческая организация «Институт инженерной физики»,
главный научный сотрудник

Евсютин Олег Олегович,
кандидат технических наук, доцент,
Московский институт электроники и математики
им. А.Н. Тихонова Федерального государственного
автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики»,
доцент и ведущий научный сотрудник кафедры
информационной безопасности киберфизических систем

Ведущая организация: Федеральное государственное бюджетное учреждение «Федеральный научный центр научно-исследовательский институт системных исследований национального исследовательского центра «Курчатовский институт»

Защита состоится 17 декабря 2026 г. в 16:15 на заседании диссертационного совета 24.1.120.01 при Федеральном государственном бюджетном учреждении науки Институте системного программирования им. В.П. Иванникова Российской Академии Наук по адресу: 115035, г. Москва, ул. Садовническая, д. 41, ст. 2.

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного учреждения науки Института системного программирования им. В.П. Иванникова Российской академии наук.

Автореферат разослан «__» _____ 2026 г.

Ученый секретарь
диссертационного совета
24.1.120.01,

кандидат физико-математических наук

Турдаков Д.Ю.

Общая характеристика работы

Актуальность темы. Развитие цифровых платформ, социальных сетей, сервисов потокового вещания и генеративных моделей искусственного интеллекта существенно изменило процессы создания и распространения медиаконтента. Изображения, аудио- и видеоданные могут создаваться, обрабатываться и публиковаться с использованием программных средств, а затем практически без потери качества копироваться и распространяться среди неограниченного числа пользователей. В то же время развитие генеративных моделей существенно расширило возможности как по синтезу нового медиаконтента, так и по внесению реалистичных изменений в существующие материалы, что усложняет различие исходного, модифицированного и полностью синтезированного контента. Это приводит к ряду угроз информационной безопасности: несанкционированное копирование и распространение медиаконтента, нарушение авторских и лицензионных прав, подмена отдельных фрагментов изображения, аудиозаписи или видео, создание поддельных материалов, в том числе с применением технологий искусственного интеллекта, а также указание неверного источника происхождения контента. При этом факт модификации медиаданных не всегда удается установить визуально или на слух, а методы выявления подделок, основанные на анализе признаков самого контента, могут терять эффективность по мере совершенствования алгоритмов его генерации и обработки. В связи с этим актуальной является разработка проактивных средств защиты, предусматривающих предварительное внесение в медиаконтент проверяемой информации о его происхождении, владельце, получателе или состоянии целостности.

Дополнительным фактором, определяющим актуальность рассматриваемой задачи, является формирование нормативных требований к обозначению и выявлению материалов, созданных или измененных с применением искусственного интеллекта. В Европейском союзе устанавливаются требования к машиночитаемому обозначению и технической детектируемости синтетического контента, в Китае действует система обязательной явной и неявной идентификации такого контента, а в США и Российской Федерации соответствующие требования продолжают формироваться на уровне законодательства и отраслевых инициатив. При этом нормативное регулирование, как правило, устанавливает требования к возможности определения происхождения материала или факта применения технологий искусственного интеллекта, не всегда конкретизируя технические способы их реализации. Практическое применение таких требований в масштабах современных цифровых платформ обуславливает необходимость разработки автоматизированных средств машиночитаемого обозначения медиаконтента и последующей проверки его происхождения и целостности.

Одним из перспективных технических средств защиты медиаконтента являются цифровые водяные знаки (ЦВЗ), позволяющие внедрять в цифровой

объект дополнительную скрытую информацию. В отличие от внешних метаданных, цифровой водяной знак непосредственно связан с объектом медиа и может сохраняться после преобразований, возникающих при хранении, передаче, перекодировании, сжатии или изменении формата данных. В настоящей работе под маркированием медиаконтента понимается процесс внедрения в изображения, аудио- и видеоданные цифровых водяных знаков, предназначенных для подтверждения происхождения, контроля целостности или отслеживания распространения.

К настоящему времени предложено значительное число методов внедрения ЦВЗ в изображения, аудио- и видеоданные, различающихся по емкости, незаметности, устойчивости, способу внедрения и модели извлечения. Значительный вклад в развитие методов цифровых водяных знаков внесли зарубежные исследователи I. Cox, T. Furon, P. Fernandez, а также отечественные исследователи В. А. Федосеев, В. И. Коржик, О. О. Евсютин и др. Наряду с методами маркирования, основанными на заранее заданных правилах преобразования медиаданных, активно развиваются обучаемые методы, в том числе с использованием нейронных сетей. Отдельное направление связано с формированием ЦВЗ непосредственно в процессе генерации медиаконтента. Однако сравнение существующих методов затрудняется отсутствием единых экспериментальных условий. В различных работах используются разные наборы данных, преобразования и атаки, метрики незаметности и устойчивости ЦВЗ. Вследствие этого заявленные в публикациях характеристики отдельных алгоритмов не всегда позволяют оценить их преимущества и выбрать метод, соответствующий конкретному сценарию применения. Это обуславливает актуальность разработки унифицированной методики и программных средств воспроизводимой оценки методов внедрения ЦВЗ.

Вместе с тем высокая устойчивость ЦВЗ сама по себе не гарантирует целостность медиаконтента и не подтверждает его происхождение. Водяной знак может сохраниться после недопустимого изменения содержимого объекта либо быть перенесен на другой объект. Для контроля целостности необходимо обеспечить проверяемую связь между содержимым медиаданных и внедренным аутентификатором, а также возможность подтверждения его подлинности. Перспективным является совместное применение ЦВЗ, перцептивного хеширования и электронной цифровой подписи. Ключевой научной задачей при этом является построение подписываемого перцептивного представления, сохраняющегося при допустимых преобразованиях и изменяющегося при нарушении содержательной целостности медиаданных.

Методы внедрения цифровых водяных знаков существенно зависят от типа обрабатываемых медиаданных. Различия в структуре, способах обработки и характерных преобразованиях изображений, аудио- и видеоданных требуют учета специфики каждого типа медиаконтента как при разработке методов внедрения ЦВЗ, так и при их экспериментальной оценке. Таким образом, актуальность диссертационной работы определяется необходимостью комплексного развития

методов и программных средств маркирования цифровых медиаданных с учетом современных моделей генерации и преобразования медиаконтента, а также возможности преднамеренного изменения содержимого и подделки ЦВЗ. Решение данной задачи требует унифицированной методики оценки существующих методов, обеспечения проверяемой связи водяного знака с содержимым медиаданных и стороной, сформировавшей ЦВЗ, разработки средств выявления недопустимых изменений, а также разработки методов применения существующих алгоритмов маркирования к видеоданным с учетом их временной структуры.

Целью диссертационной работы является разработка и экспериментальное обоснование методов и создание программных средств защиты медиаконтента на основе цифровых водяных знаков, обеспечивающих воспроизводимую оценку методов маркирования, контроль целостности изображений и аудиоданных, а также эффективное применение методов маркирования изображений к видеоданным.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Провести анализ существующих методов маркирования медиаконтента, подходов к контролю его целостности и средств сравнительной оценки и определить основные ограничения существующих решений.
2. Разработать методику и программные средства унифицированной воспроизводимой оценки методов внедрения цифровых водяных знаков в медиаконтент, включающие единую модель экспериментального конвейера, наборы данных, модели атак и метрики незаметности и устойчивости для медиаданных различного типа.
3. Разработать и экспериментально исследовать метод контроля целостности изображений, устойчивый к допустимым преобразованиям и обеспечивающий выявление недопустимых модификаций.
4. Разработать и экспериментально исследовать метод контроля целостности аудиоданных, устойчивый к допустимым преобразованиям и обеспечивающий выявление и локализацию недопустимых модификаций.
5. Разработать и экспериментально исследовать методику переноса алгоритмов внедрения и извлечения цифровых водяных знаков, предназначенных для изображений, на видеоданные, позволяющую сократить вычислительные затраты за счет использования временной избыточности видеопоследовательности без изменения исходных алгоритмов.

Объект и предмет исследования. Объектом исследования являются процессы цифровой обработки медиаданных при внедрении в них цифровых водяных знаков.

Предметом исследования являются методы, алгоритмы и программные средства обработки медиаданных, предназначенные для унифицированной экспериментальной оценки методов маркирования, контроля целостности изображений и аудиоданных и переноса методов маркирования изображений на видеоданные.

Научная новизна:

1. Разработана методика унифицированной воспроизводимой оценки методов маркирования медиаконтента, основанная на едином представлении экспериментального процесса для методов с различными принципами формирования и проверки ЦВЗ и различных типов медиаданных при сохранении модально-зависимых воздействий, метрик и способов представления данных.
2. Разработан метод контроля целостности изображений, в котором предложен способ повышения побитовой устойчивости подписываемого перцептивного представления за счет контролируемой модификации низкочастотных коэффициентов изображения относительно порогов бинаризации без изменения исходного значения перцептивного хеш-кода.
3. Разработан метод контроля целостности аудиоданных, в котором независимое формирование и проверка контентно-зависимых аутентификаторов для последовательных временных фрагментов используются для построения карты целостности аудиозаписи, позволяющей выявлять и локализовать содержательные модификации.
4. Разработана методика переноса методов маркирования изображений на видеоданные, основанная на однократной обработке агрегированного представления группы последовательных кадров и распределении сформированной модификации между кадрами, что снижает вычислительные затраты без изменения исходного алгоритма и необходимости его дополнительного обучения.

Теоретическая и практическая значимость. Теоретическая значимость диссертационной работы заключается в развитии методов оценки и применения цифровых водяных знаков к медиаданным различных типов. В работе предложена унифицированная методика оценки методов внедрения цифровых водяных знаков в медиаконтент, позволяющая сопоставлять алгоритмы в воспроизводимых условиях. Разработанные методы интеграции цифровых водяных знаков с электронной цифровой подписью и перцептивным хешированием расширяют научные представления о построении схем маркирования медиаданных с аутентификацией, устойчивых к допустимым преобразованиям и чувствительных к недопустимым модификациям.

Практическая значимость диссертационной работы состоит в разработке программных средств и алгоритмов, применимых в системах защиты медиаконтента, сервисах генерации изображений, аудио- и видеоданных, системах

потокowego вещания, а также в средствах проверки происхождения и целостности цифрового контента. Разработанная система WIBE может использоваться исследователями и разработчиками для воспроизводимого тестирования методов внедрения цифровых водяных знаков. Предложенные методы контроля целостности изображений и аудиоданных могут применяться для проверки подлинности медиаконтента и выявления недопустимых изменений. Разработанная методика переноса методов маркирования изображений на видеоданные позволяет применять существующие алгоритмы к видеопоследовательностям без изменения их архитектуры и дополнительного обучения при снижении вычислительных затрат за счет групповой обработки кадров.

Результаты диссертационной работы внедрены в практику работы ООО «ЯНДЕКС» при проведении исследований и разработке средств маркирования и проверки медиаконтента. Практическое применение получили разработанные методика и программные средства воспроизводимой оценки методов маркирования, а также алгоритмические решения для маркирования изображений и аудиоданных, что подтверждено актом о внедрении.

Методология и методы исследования. Методологическую основу работы составляют системный подход к построению средств защиты медиаконтента, моделирование угроз и нарушителя, а также экспериментальная оценка разработанных методов. В работе использованы методы цифровой обработки изображений, аудиосигналов и видеоданных, методы маркирования, перцептивного хеширования, помехоустойчивого кодирования, машинного обучения и статистического анализа. Экспериментальное исследование основано на применении преобразований и преднамеренных атак к маркированному медиаконтенту, оценке незаметности и устойчивости цифровых водяных знаков.

Основные положения, выносимые на защиту:

1. Методика унифицированной воспроизводимой оценки методов внедрения цифровых водяных знаков в медиаконтент, обеспечивающая единую модель экспериментального конвейера при использовании модально-зависимых наборов данных, моделей атак и метрик незаметности и устойчивости.
2. Метод контроля целостности изображений на основе подписываемого перцептивного представления, обеспечивающий устойчивость к допустимым преобразованиям и выявление недопустимых модификаций.
3. Метод контроля целостности аудиоданных на основе независимой проверки подписываемого перцептивного представления временных фрагментов, обеспечивающий устойчивость к допустимым преобразованиям, а также выявление и локализацию недопустимых модификаций.
4. Методика переноса методов маркирования изображений на видеоданные на основе групповой обработки последовательных кадров, обеспечивающая снижение вычислительных затрат за счет использования

временной избыточности видеопоследовательности без изменения исходных алгоритмов маркирования.

Соответствие паспорту специальности. Приведенные положения соответствуют следующим направлениям исследований паспорта специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»:

- **п.4** Интеллектуальные системы машинного обучения, управления базами данных и знаний, инструментальные средства разработки цифровых продуктов.
- **п.7** Модели, методы, архитектуры, алгоритмы, форматы, протоколы и программные средства человеко-машинных интерфейсов, компьютерной графики, визуализации, обработки изображений и видеоданных, систем виртуальной реальности, многомодального взаимодействия в социокберфизических системах.
- **п.8** Модели и методы создания программ и программных систем для параллельной и распределенной обработки данных, языки и инструментальные средства параллельного программирования.
- **п.10** Оценка качества, стандартизация и сопровождение программных систем.

Достоверность результатов обеспечивается экспериментальной проверкой разработанных методов на общедоступных наборах медиаданных, сравнением с существующими решениями и использованием количественных показателей качества медиаданных, устойчивости ЦВЗ и точности выявления модификаций.

Апробация работы. Результаты работы обсуждались на следующих конференциях:

- II форум «Технологии доверенного искусственного интеллекта», Москва, 27 мая 2024 г.
- Всероссийская конференция «Методы и технические средства обеспечения безопасности информации», Санкт-Петербург, 24–27 июня 2024 г.
- VI форум «Безопасность информационных технологий в цифровой экономике», Москва, 25 сентября 2024 г.
- Международная конференция «Иванниковские чтения», Иркутск, 26–27 июня 2025 г.
- 40th IEEE/ACM International Conference on Automated Software Engineering, Сеул, 16–20 ноября 2025 г.

Личный вклад. Все выносимые на защиту результаты получены лично автором.

Публикации. Основные результаты по теме диссертационной работы представлены в 10 научных работах, из которых 8 опубликованы и 2 приняты к публикации. Среди опубликованных работ 4 статьи изданы в журналах, рекомендованных ВАК, 2 работы — в изданиях, индексируемых в Web of Science и

Scopus. Получено 8 свидетельств о государственной регистрации программ для ЭВМ.

В работах [1; 2] автором предложены и разработаны методы внедрения цифровых водяных знаков в изображения текстовых документов. В работах [3; 4] автор принимал участие в разработке и реализации системы маркирования текстовых документов. В работе [5] автором исследованы подходы к применению цифровых водяных знаков для защиты видеоконтента. В работах [6; 7] автором предложены архитектура программной системы и подход к унифицированной воспроизводимой оценке методов маркирования, разработан ряд компонентов системы. В работе [8] автором предложен и разработан метод контроля целостности изображений CPW и способ повышения побитовой устойчивости подписываемого перцептивного представления.

Содержание работы

Во введении обоснована актуальность разработки методов и программных средств защиты медиаконтента на основе цифровых водяных знаков. Сформулированы цель и задачи исследования, определены объект и предмет исследования, представлены научная новизна, теоретическая и практическая значимость полученных результатов, положения, выносимые на защиту, а также сведения об апробации и публикациях.

Первая глава посвящена анализу существующих методов защиты медиаконтента на основе цифровых водяных знаков. Рассматриваются основные принципы маркирования изображений, аудио- и видеоданных, методы контроля целостности медиаконтента и средства сравнительной оценки методов маркирования.

В начале главы вводятся основные понятия маркирования медиаконтента и схемы внедрения, обнаружения и извлечения ЦВЗ. Рассматриваются методы обнаружения наличия ЦВЗ (*zero-bit*) и передачи сообщений (*multi-bit*), слепые и неслепые схемы проверки, а также устойчивые, хрупкие и полухрупкие ЦВЗ. В качестве основных характеристик методов выделяются незаметность, емкость, устойчивость и вычислительная эффективность. Под воздействиями в работе понимаются как типовые преобразования медиаданных, возникающие при их обработке, хранении и передаче, так и преднамеренные атаки.

В разделе 1.2 проводится анализ методов маркирования изображений. Классические методы основаны на явно заданных правилах изменения пикселей или коэффициентов преобразований. Рассматриваются подходы на основе младших значащих битов (LSB), дискретного косинусного, вейвлет- и Фурье-преобразований, расширения спектра (DCT CAISS) и модуляции индекса квантования (QIM).

Нейросетевые методы определяют правила внедрения и извлечения в процессе обучения и позволяют учитывать предполагаемые преобразования маркированного изображения. Рассмотрены подходы, ориентированные на

повышение устойчивости к JPEG-сжатию (MBRS), геометрическим преобразованиям (DWSF, SyncSeal), редактированию изображений (Robust-Wide), локальное обнаружение ЦБЗ (WAM), работу с изображениями высокого разрешения (TrustMark, PixelSeal) и передачу сообщений большой длины (ChunkySeal).

Отдельное направление составляют методы формирования ЦБЗ непосредственно в процессе генерации изображений диффузионными моделями. В частности, рассмотрены подходы Stable Signature, Tree-Ring, RingID, METR, Gaussian Shading и их дальнейшие развития. Маркирование начального шума диффузионной модели позволяет формировать ЦБЗ без дополнительной обработки получаемого изображения, однако требует доступа к процессу генерации и последующего приближенного восстановления начального шума при проверке.

В разделе 1.3 анализируются классические и нейросетевые методы маркирования аудиоданных. Классические методы используют временное, частотное и временно-частотное представления сигнала. Рассматриваются подходы на основе LSB, QIM, расширения спектра, дискретного косинусного и вейвлет-преобразований, а также признаки, формируемые по значениям спектральных коэффициентов.

Современные нейросетевые методы маркирования аудиоданных используют обучаемые кодировщики и декодеры и позволяют учитывать при обучении шум, фильтрацию, сжатие и другие преобразования сигнала. Рассмотрены методы WavMark, AudioSeal, SilentCipher, AWARE и другие подходы. Часть методов использует повторное внедрение сообщения и специальные механизмы поиска его временного положения, другие обеспечивают локальную оценку наличия ЦБЗ во времени. Отдельное направление связано с сохранением ЦБЗ при клонировании голоса и нейросетевом синтезе речи. Анализ показывает, что для аудиоданных наряду с незаметностью и устойчивостью к обработке необходимо учитывать устойчивость к нарушениям временной структуры и возможность получения локального результата обнаружения ЦБЗ в аудиозаписи.

В разделе 1.4 рассматриваются методы маркирования видеоданных, которые во многих случаях развивают и адаптируют подходы, первоначально предназначенные для изображений. Выделяются независимое покадровое маркирование, распространение модификации одного кадра на соседние кадры (VideoSeal), совместная обработка нескольких кадров или их признаков (ItoV, PixelSeal), а также специализированные методы для видео (RivaGAN, DVMark, REVMark). Покадровый подход наиболее универсален, но требует выполнения алгоритма для каждого кадра и не использует временную избыточность видео. Групповая обработка и использование общих признаков позволяют учитывать связанность соседних кадров и снижать вычислительные затраты, однако обычно требуют изменения архитектуры исходного метода или его дополнительного обучения. В результате анализа установлено, что существующие подходы не обеспечивают одновременно универсальность покадрового применения, учет

содержимого нескольких кадров и снижение вычислительных затрат без изменения исходного метода маркирования изображений.

Отдельно рассматривается задача контроля целостности медиаконтента. Анализируются хрупкие и полухрупкие ЦВЗ, нейросетевые методы выявления и локализации изменений (SepMark, EditGuard, OmniGuard), а также подходы, использующие контентно-зависимые аутентификационные данные и цифровую подпись (MetaSeal). Показано, что успешное обнаружение устойчивого ЦВЗ само по себе не подтверждает сохранность содержимого объекта: водяной знак может быть перенесен на другой объект или сохранен после недопустимой модификации. Для защиты от таких сценариев рассматривается формирование аутентификационных данных в зависимости от содержимого объекта с их последующей защитой цифровой подписью.

В заключительной части главы проводится анализ существующих программных средств для оценки методов маркирования. Для изображений рассматриваются StirMark, WAVES, W-Bench, MarkDiffusion и OmniSeal Bench, для аудиоданных — AudioMarkBench, RAW-Bench, VoxWatermark, DeepMark Benchmark и OmniSeal Bench. Современные средства учитывают широкий спектр стандартных преобразований, специализированных атак, генеративного редактирования, нейронной обработки и различных моделей доступа нарушителя. При этом большинство решений ориентировано на отдельные модальности, классы методов или экспериментальные постановки.

Проведенный анализ выявил ограничения существующих подходов к сравнительной оценке методов маркирования, контролю целостности медиаконтента и применению методов маркирования изображений к видео, определившие направления последующих исследований.

Вторая глава посвящена разработке методики и программных средств унифицированной воспроизводимой оценки методов маркирования медиаконтента. Основная задача состоит в формировании единой организации экспериментального процесса, применимой к методам с различными принципами формирования и проверки ЦВЗ и к различным типам медиаданных. Общими являются организация эксперимента, последовательность его стадий и принципы подключения компонентов, тогда как конкретные представления медиаданных, воздействия и метрики определяются исследуемой модальностью.

В разделе 2.1 предлагается модель унифицированной экспериментальной оценки методов маркирования. В рамках разработанной методики эксперимент задается совокупностью

$$\mathcal{E} = (\mathcal{D}, \mathcal{W}, \mathcal{A}, \mathcal{Q}, \mathcal{R}),$$

где $\mathcal{D}$ — набор исходных медиаданных, $\mathcal{W}$ — множество исследуемых методов маркирования, $\mathcal{A}$ — множество воздействий, $\mathcal{Q}$ — набор метрик качества и незаметности, $\mathcal{R}$ — набор метрик обнаружения и извлечения ЦВЗ. Общая схема разработанной методики представлена на рисунке 1.

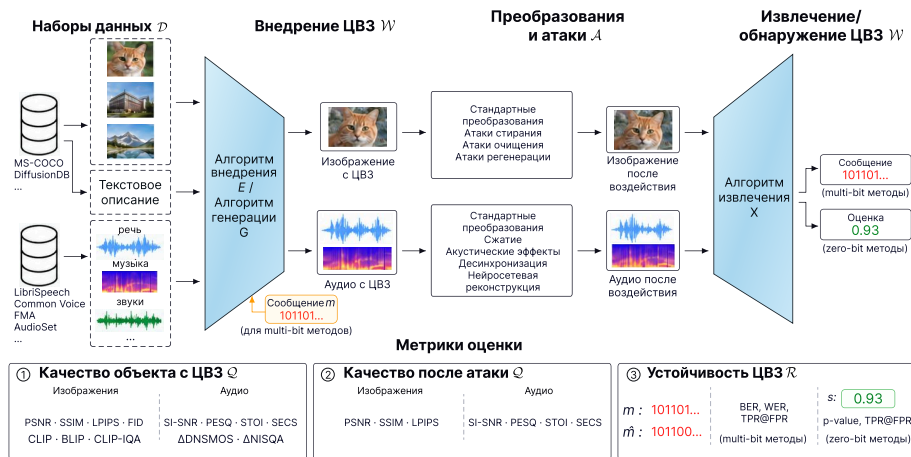


Рисунок 1 — Схема унифицированной экспериментальной оценки методов маркирования медиаконтента

Эксперимент включает формирование маркированного объекта, применение воздействия, обнаружение или извлечение ЦВЗ и расчет соответствующих метрик. Такое представление позволяет исследовать в рамках одной модели различные методы при сохранении модально-зависимых данных, воздействий и метрик.

Предложенная модель учитывает различия между методами, внедряющими ЦВЗ в уже сформированный объект, и методами, формирующими маркированный контент непосредственно в процессе генерации. Для первой группы возможно непосредственное сопоставление исходного и маркированного объектов. При маркировании в процессе генерации соответствующий немаркированный объект в общем случае отсутствует, поэтому наряду с безэталонными метриками могут использоваться метрики соответствия условию генерации и метрики сравнения распределений объектов.

Для оценки устойчивости используются метрики, учитывающие тип проверяемого ЦВЗ. Для multi-bit методов используются BER и доля сообщений с ошибками, а для задачи обнаружения — TPR при заданном FPR. Для сопоставления zero-bit и multi-bit методов используется задача обнаружения ЦВЗ. Для multi-bit методов решение об обнаружении принимается по числу совпадающих битов исходного и извлеченного сообщений. Также предложен теоретический расчет FPR по биномиальной модели. Для немаркированных объектов предполагаются независимые совпадения битов с вероятностью 0,5. Применимость модели проверена экспериментально на методе TrustMark.

В разделе 2.2 рассматривается программная реализация предложенной методики в разработанной системе WIBE. Архитектура системы основана на разделении ядра экспериментального конвейера и подключаемых компонентов, реализующих методы маркирования, наборы данных, воздействия и метрики.

В рамках экспериментального конвейера последовательно выполняются стадии внедрения ЦВЗ, применения воздействий, извлечения ЦВЗ, расчета метрик и агрегации результатов.

Для передачи результатов между стадиями используется объект `Context`, содержащий текущее состояние обработки отдельного объекта. Реализации методов маркирования, наборы данных, воздействия и метрики подключаются к конвейеру как независимые компоненты, благодаря чему состав исследуемых методов и условий эксперимента может расширяться без изменения ядра системы.

Воспроизводимое описание условий эксперимента отделено от программного кода и выполняется декларативно с использованием конфигурационных файлов. В них задаются состав и параметры экспериментальных компонентов, а также начальное состояние генераторов псевдослучайных чисел. Для интеграции программных реализаций с несовместимыми зависимостями разработан менеджер виртуальных окружений `wibench-venv`. WIBE также поддерживает сохранение промежуточных состояний объектов, раздельное выполнение стадий экспериментального конвейера, параллельную обработку объектов и сохранение результатов вместе с параметрами соответствующих экспериментальных конфигураций.

Для экспериментальной апробации разработанной методики в разделах 2.3 и 2.4 проведена сравнительная оценка методов маркирования изображений и аудиоданных. Для изображений в WIBE интегрированы 32 конфигурации методов, из которых 26 предназначены для внедрения ЦВЗ в сформированные изображения, а 6 — для маркирования в процессе генерации. Эксперименты проводились на наборах `MS-COCO` и `DiffusionDB`.

При сравнении оценивались незаметность, емкость, вычислительная эффективность и устойчивость ЦВЗ. Полученные результаты показали, что увеличение емкости не обязательно сопровождается ростом перцептивных искажений: `PixelSeal` и `MBRS` поддерживают сообщения длиной 256 бит, а `ChunkySeal` — 1024 бита при сравнительно небольших значениях метрики незаметности `LPIPS`.

Устойчивость ЦВЗ исследовалась для четырех классов воздействий: стандартных преобразований, атак стирания, очищения и регенерации. Для сопоставления устойчивости использовался показатель $\text{TPR}@10^{-4}\text{FPR}$. Результаты показали, что устойчивость существенно зависит от класса воздействия. Высокая устойчивость к стандартным преобразованиям не гарантирует устойчивость при преднамеренных атаках. Наиболее заметные различия наблюдались при регенерации изображений: среди методов маркирования в процессе генерации `Tree-Ring` и `Gaussian Shading` сохраняли высокую вероятность обнаружения ЦВЗ, тогда как `Stable Signature` оказался более чувствительным к этому классу воздействий. Эффективность атак оценивалась совместно с величиной вносимых перцептивных искажений. Атаки `Flux Rinsing` и `Flux Regen` сильнее подавляли ЦВЗ, но сопровождалась сравнительно высокими значениями метрики `LPIPS`,

тогда как атаки SADRE и Diffusion Regen обеспечивали заметное подавление при меньших искажениях.

Применимость той же методики к другой модальности исследована на 14 классических и нейросетевых методах маркирования аудиоданных. Экспериментальная выборка включала 1000 речевых аудиозаписей из наборов LibriSpeech, VCTK, Common Voice, Golos и AISHELL. Для сопоставления устойчивости zero-bit и multi-bit методов использовался показатель $TPR@10^{-3}FPR$.

Воздействия на аудиоданные были разделены на пять классов: стандартные преобразования, сжатие с потерями, акустические эффекты, атаки десинхронизации, а также очищение и нейросетевую реконструкцию. Результаты показали существенную зависимость устойчивости от типа воздействия: наиболее высокие значения показателя в совокупности рассмотренных воздействий получены для DNN-Audio и AWARE, тогда как для значительной части методов наиболее сложными оказались акустические эффекты и временная десинхронизация.

Третья глава посвящена методу контроля целостности изображений CPW (*Cryptographic Perceptual Watermarking*), предназначенному для публичной проверки соответствия изображения ранее аутентифицированному содержимому без доступа к исходному изображению и закрытому ключу.

В разделе 3.1 формулируется задача контроля целостности. Для маркированного изображения I_w , открытого ключа K_{pub} и процедуры верификации V требуется

$$V(T(I_w), K_{pub}) = \begin{cases} 1, & T \in T_A, \\ 0, & T \in T_N, \end{cases}$$

где T_A и T_N — множества допустимых и недопустимых преобразований соответственно.

К допустимым относятся JPEG-сжатие, фильтрация, шум, изменение яркости и контрастности и масштабирование с сохранением соотношения сторон; к недопустимым — локальные вставка, удаление и клонирование фрагментов изображения. В модели нарушителя предполагается знание алгоритмов и параметров метода, наличие открытого ключа и возможности многократного выполнения процедуры верификации, но отсутствие доступа к соответствующему закрытому ключу. Целью нарушителя является формирование содержательно измененного изображения, успешно проходящего проверку.

В разделе 3.2 представлен разработанный метод CPW, объединяющий перцептивное хеширование, электронную цифровую подпись и цифровой водяной знак. Общая схема метода приведена на рисунке 2. При формировании маркированного изображения контентно-зависимое перцептивное представление подписывается закрытым ключом, а полученная подпись защищается помехоустойчивым кодом и внедряется в виде ЦВЗ. Тем самым перцептивное представление связывает аутентификатор с содержимым изображения, а цифровая подпись — с заданным открытым ключом.

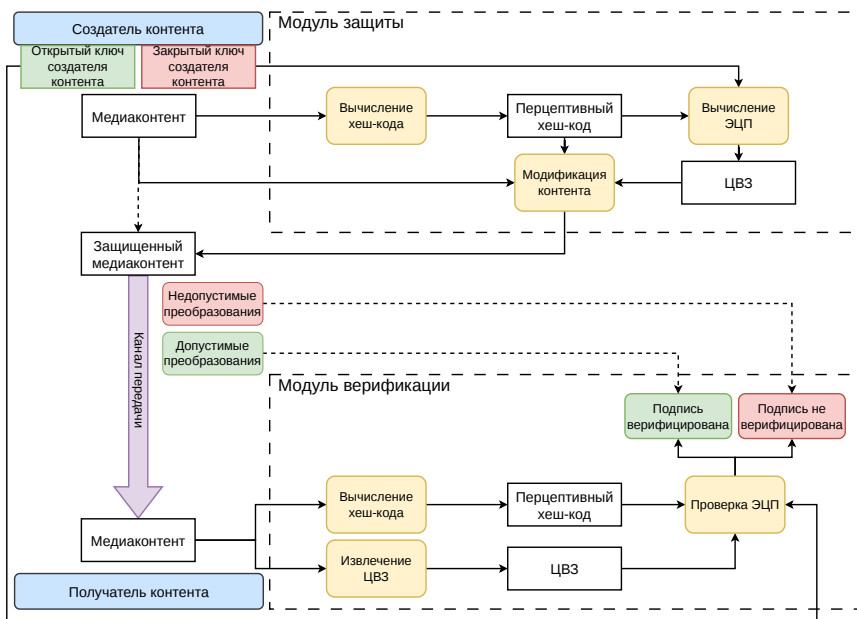


Рисунок 2 — Схема метода CPW для контроля целостности изображений

Для формирования подписываемого представления разработана перцептивная хеш-функция EPHash (*Enhanced Perceptual Hash*). Изображение переводится в градации серого, масштабируется и разбивается на непересекающиеся блоки. В каждом блоке вычисляется дискретное косинусное преобразование (ДКП), выбранные низкочастотные коэффициенты которого бинаризируются относительно их медианы. Полученные биты объединяются в хеш-код изображения.

Для повышения побитовой устойчивости хеш-кода предложен способ контролируемой модификации низкочастотных коэффициентов ДКП относительно порога бинаризации. Если расстояние коэффициента $C_{u,v}$ от медианного значения m меньше заданного отступа $\delta_{\min}$, коэффициент изменяется в сторону уже определенного значения соответствующего бита:

$$C'_{u,v} = \begin{cases} \max(C_{u,v}, m + \delta_{\min}), & b_{u,v} = 1, \\ \min(C_{u,v}, m - \delta_{\min}), & b_{u,v} = 0. \end{cases}$$

Такая модификация увеличивает расстояние коэффициентов от порогов бинаризации, не изменяя сформированного значения хеш-кода.

Значение $\delta_{\min}$ подбирается отдельно для каждого блока при ограничении на минимально допустимое качество восстановленного изображения. Тем самым повышение побитовой устойчивости достигается за счет однократной модификации изображения на этапе формирования аутентификатора, тогда как при

последующей проверке EPHash вычисляется без изменения исследуемого изображения.

Для формирования аутентификатора значение EPHash подписывается алгоритмом ECDSA на кривой `secp256k1`. Полученная 512-битная подпись защищается кодом БЧХ, после кодирования длина внедряемого сообщения составляет 1024 бита.

Для внедрения 1024-битного аутентификатора рассмотрены две стратегии: непосредственное внедрение всей последовательности с использованием ChunkySeal и ее разделение на четыре 256-битных фрагмента для независимого внедрения методами PixelSeal, VideoSeal и MBRS в 4 квадранта изображения. Метод маркирования является заменяемым компонентом CPW и отвечает за передачу аутентификатора.

При верификации из исследуемого изображения извлекается и декодируется аутентификатор, независимо вычисляется текущее значение EPHash и выполняется проверка цифровой подписи с использованием открытого ключа. Положительный результат подтверждает соответствие аутентификатора текущему перцептивному представлению изображения и заданному открытому ключу.

В разделе 3.3 приведено экспериментальное исследование разработанного метода. Оценивались побитовая устойчивость EPHash, незаметность формирования маркированного изображения, устойчивость полной процедуры CPW к допустимым преобразованиям и способность выявлять содержательно значимые модификации. Основная часть экспериментов проводилась на выборке из 1000 изображений из набора MS-COCO, а для исследования недопустимых модификаций использовались 926 пар изображений из набора IMD2020.

Побитовая устойчивость EPHash сравнивалась с перцептивными хеш-функциями aHash, pHash, wHash и PDQ. Результаты показали, что предложенный способ формирования отстupa от порогов бинаризации повышает вероятность точного воспроизведения хеш-кода. EPHash демонстрировал более высокую устойчивость к масштабированию, размытию и JPEG-сжатию. При добавлении гауссовского шума наилучшие результаты получены для EPHash и wHash.

Сравнение реализованного подхода CPW производилось с методом MetaSeal. Для сравнения незаметности методов использовались метрики PSNR, SSIM и LPIPS. Полученные результаты приведены в таблице 1.

Таблица 1 — Сравнение незаметности различных вариантов CPW и метода MetaSeal

Метод	PSNR, дБ ↑	SSIM ↑	LPIPS ↓
CPW (MBRS)	40,7	0,982	0,011
CPW (VideoSeal)	41,1	0,985	0,009
CPW (PixelSeal)	41,2	0,985	0,009
CPW (ChunkySeal)	41,4	0,986	0,011
MetaSeal	35,1	0,929	0,098

Исследование полной процедуры верификации показало, что ее устойчивость определяется совместно свойствами EPHash и используемого метода маркирования. Вариант CPW (MBRS) демонстрировал наиболее высокую устойчивость к JPEG-сжатию и сохранял успешную верификацию при снижении качества JPEG до 40, однако был менее устойчив к размытию. CPW (ChunkySeal) лучше переносил размытие и уменьшение разрешения, но оказался чувствительным к JPEG-сжатию. Наиболее сбалансированный профиль устойчивости среди рассмотренных вариантов показал вариант CPW (VideoSeal).

Способность выявлять содержательно значимые изменения исследовалась на наборе IMD2020, включающем вставку, удаление и клонирование фрагментов изображений. Для всех вариантов CPW доля выявленных нарушений приближается к единице при средней площади модифицированной области не менее 1,5 %, тогда как для MetaSeal сопоставимый результат достигается при площади порядка 10 % (рисунок 3).

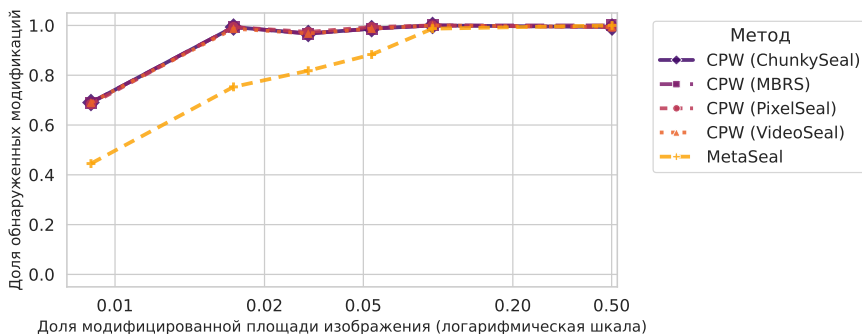


Рисунок 3 — Зависимость доли выявленных нарушений целостности от площади модифицированной области изображения

В разделе 3.4 рассматриваются ограничения разработанного метода. Устойчивость EPHash к целенаправленному поиску второго прообраза формально не доказывается, а рассматриваемая реализация не поддерживает ряд геометрических преобразований. Кроме того, CPW формирует итоговый результат проверки целостности, но не локализует модифицированную область.

Четвертая глава посвящена разработке и экспериментальному исследованию метода контроля целостности аудиоданных SuptoMark. Метод должен не только выявлять нарушение целостности, но и определять временные участки, для которых она не может быть подтверждена.

В разделе 4.1 процедура верификации аудиосигнала A' определяется как

$$\mathcal{V}(A', K_{\text{pub}}) = (v_1, v_2, \dots, v_N), \quad v_i \in \{0, 1\},$$

где каждому v_i соответствует временной фрагмент аудиозаписи. Последовательность результатов образует карту целостности.

К допустимым преобразованиям относятся сжатие с потерями, пере-дискретизация, переквантование, умеренное добавление шума, частотная фильтрация и изменение амплитуды. В качестве недопустимых модификаций рассматриваются преобразование голоса, замена временных участков синтетическим содержимым и нейросетевая реконструкция.

В разделе 4.2 представлен разработанный метод CryptoMark, основанный на независимой аутентификации последовательных временных фрагментов аудиосигнала. Общая схема метода приведена на рисунке 4.

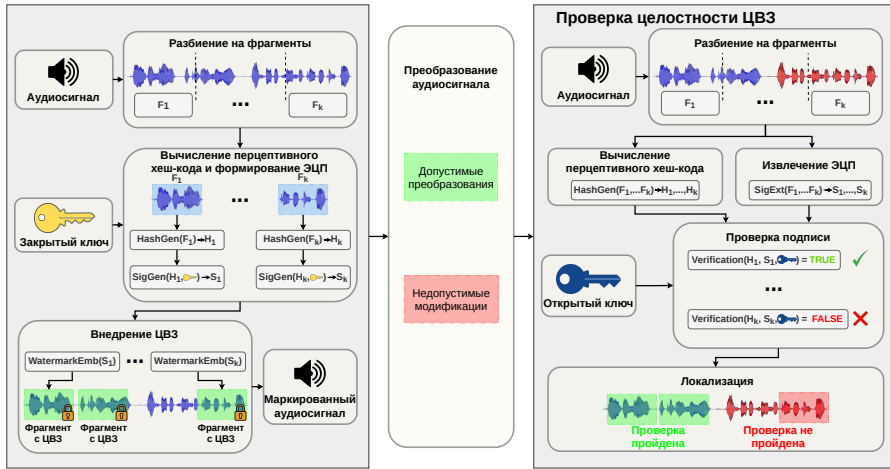


Рисунок 4 — Схема метода CryptoMark для контроля целостности аудиоданных

Аудиосигнал разбивается на неперекрывающиеся фрагменты длиной L_f отсчетов. Для каждого фрагмента независимо формируется и внедряется контентно-зависимый аутентификатор, а при проверке формируется локальный результат верификации. Временное разрешение локализации составляет $\Delta t = L_f / f_s$, где f_s — частота дискретизации. Уменьшение L_f повышает точность локализации, но сокращает объем сигнала, доступный для надежного внедрения аутентификатора.

Перцептивное представление H_i каждого фрагмента формируется путем бинаризации коэффициентов одномерного ДКП относительно медианного значения. Для повышения побитовой устойчивости выполняется итерационная коррекция спектральных коэффициентов с моделированием допустимого преобразования, при которой исходное значение хеш-кода сохраняется.

Для формирования аутентификатора значение H_i подписывается алгоритмом ECDSA. Полученная 512-битная цифровая подпись защищается кодом

Рида—Соломона. После помехоустойчивого кодирования длина внедряемой последовательности составляет 1024 бита на фрагмент.

Для внедрения аутентификатора используется модифицированный вариант метода FSVC. Внедрение ЦВЗ осуществляется путем модификации спектральных компонентов аудиозаписи. Область внедрения выбирается отдельно для каждого фрагмента с учетом распределения энергии его спектральных компонентов. При недостаточной энергии отдельных компонентов интенсивность внедрения дополнительно уменьшается.

При верификации для каждого фрагмента независимо вычисляется перцептивное представление, извлекается цифровая подпись и выполняется проверка с использованием открытого ключа. Полученные локальные результаты формируют карту целостности аудиозаписи.

В разделе 4.3 приведена экспериментальная оценка CryptoMark, включающая исследование незаметности маркирования, устойчивости к допустимым преобразованиям, способности выявлять содержательно значимые модификации и точности их временной локализации. Во всех экспериментах использовалась единая конфигурация CryptoMark с частотой дискретизации 22050 Гц и длительностью независимо проверяемого фрагмента около 4,46 с.

Незаметность CryptoMark сравнивалась с методами AudioSeal и SilentCipher с использованием метрик SI-SNR, PESQ и STOI. Полученные результаты приведены в таблице 2. Емкость указана без учета кодов коррекции.

Таблица 2 — Оценка качества маркированных аудиоданных

Метод	Емкость	SI-SNR, дБ ↑	PESQ ↑	STOI ↑
AudioSeal	16 бит	25,47	4,38	0,99
CryptoMark	512 бит / фрагмент	25,56	4,41	0,99
SilentCipher	40 бит	34,98	4,57	0,99

Для оценки контроля целостности использовались 177 аудиозаписей, сформированных из поднаборов LibriSpeech test-clean и test-other. Рассматривались 14 конфигураций допустимых преобразований, а в качестве недопустимых модификаций использовались два варианта kNN-VC и реконструкция с помощью Vocos.

В эксперименте к положительному классу относились аудиозаписи, подвергнутые допустимым преобразованиям, а к отрицательному — аудиозаписи с недопустимыми модификациями. Оценкой сохранности аудиозаписи служила доля временных фрагментов с успешно проверенной цифровой подписью. При превышении этой долей заданного порога запись признавалась прошедшей контроль целостности. Для каждого порога вычислялись доля правильно принятых объектов положительного класса (TPR) и доля ошибочно принятых объектов отрицательного класса (FPR). Путем изменения порога строилась ROC-кривая и определялась площадь под ней (ROC AUC).

CryptoMark сохранял положительный результат проверки для большинства допустимых преобразований. Безошибочная классификация получена для MPEG-сжатия во всем исследованном диапазоне битрейтов, переквантования, передискретизации и изменения амплитуды. Наибольшее снижение устойчивости наблюдалось при частотной фильтрации и добавлении шума.

Для всех исследованных содержательно значимых модификаций получены $FPR = 0$ и $ROC AUC = 1$: ни одна модифицированная запись не была ошибочно признана прошедшей контроль целостности.

Отдельно исследовалась способность CryptoMark локализовать содержательные модификации во времени. Для этого непрерывный участок маркированной аудиозаписи заменялся результатом преобразования исходного содержимого методом kNN-VC. Качество локализации оценивалось с использованием Accuracy и Intersection over Union (IoU), а результаты сравнивались с AudioSeal. При расчете IoU положительному классу соответствовали неизменные участки аудиозаписи. Полученные зависимости представлены на рисунке 5.

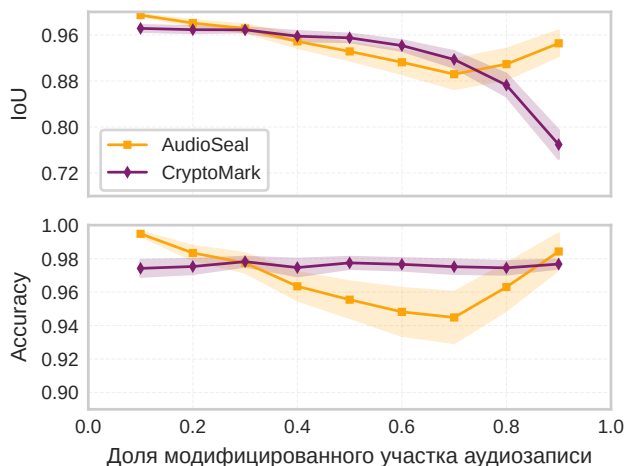


Рисунок 5 — Зависимость качества локализации от доли модифицированного участка аудиозаписи

Значение Accuracy для CryptoMark оставалось высоким во всем рассматриваемом диапазоне длительности модифицированного участка. По метрике IoU оба метода демонстрировали близкие результаты при малой длительности модификации, тогда как в среднем диапазоне CryptoMark обеспечивал более высокое перекрытие предсказанной и фактической областей.

В разделе 4.4 рассматриваются ограничения метода. Текущая реализация не содержит механизма восстановления временной синхронизации, а точность локализации ограничена длиной независимо проверяемого фрагмента.

Пятая глава посвящена разработке и экспериментальному исследованию методики переноса методов маркирования изображений на видеоданные.

В разделе 5.1 ставится задача построения методики маркирования видео на основе исходных алгоритмов внедрения E и извлечения X . Методика должна обеспечивать сокращение вычислительных затрат по сравнению с независимой покадровой обработкой видео. При этом требуется сохранить устойчивость ЦВЗ, ограничить вносимые перцептивные искажения и исключить необходимость изменения и дополнительного обучения исходных алгоритмов.

В разделе 5.2 представлена разработанная методика, основанная на совместной обработке групп последовательных кадров. Для каждой группы формируется агрегированное представление

$$\bar{I}_g = \frac{1}{N_g} \sum_{i=1}^{N_g} I_{g,i},$$

где N_g — число кадров в группе. К полученному изображению однократно применяется исходный алгоритм внедрения E , после чего определяется сформированная модификация

$$D_g = E(\bar{I}_g, m) - \bar{I}_g,$$

которая распределяется между кадрами группы:

$$I'_{g,i} = \text{clip}(I_{g,i} + \beta_{g,i} D_g, 0, 1), \quad \frac{1}{N_g} \sum_{i=1}^{N_g} \beta_{g,i} = 1.$$

Коэффициенты $\beta_{g,i}$ задают временное распределение модификации, m — внедряемое сообщение. В базовом варианте $\beta_{g,i} = 1$. При извлечении маркированные кадры группы также усредняются, после чего к полученному изображению однократно применяется исходный алгоритм извлечения X . При обработке видеопоследовательности из T кадров число вызовов каждого из исходных алгоритмов сокращается с T до $\lceil T/N \rceil$, где N — заданный размер группы.

В разделе 5.3 приведена экспериментальная оценка разработанной методики. Исследования проводились на 1000 видеопоследовательностях набора YouTube-VOS. Рассматривались нейросетевые методы VideoSeal и TrustMark, а также классический метод DCT CAISS. В качестве базового варианта использовалось независимое покадровое маркирование. С ним сравнивались три стратегии, сокращающие число вызовов исходного алгоритма: маркирование только первого кадра группы, распространение модификации первого кадра на остальные кадры и предлагаемая групповая обработка агрегированного представления. Основной размер группы составлял $N = 30$ кадров.

Устойчивость извлечения оценивалась после H.264-сжатия с параметром CRF=23 по метрике BER, пространственная незаметность — по PSNR и LPIPS.

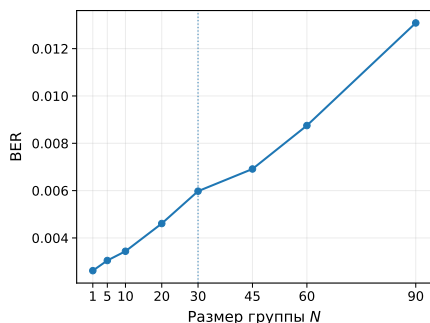
Для оценки изменений временной согласованности соседних кадров использовалась метрика TLP100. Результаты сравнения стратегий представлены в таблице 3.

Таблица 3 — Сравнение стратегий переноса методов маркирования изображений на видеоданные

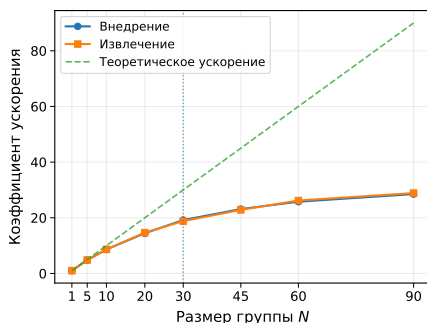
Стратегия	BER ↓	PSNR ↑, дБ	LPIPS ↓	TLP100 ↓	Вызовы внедрения ↓
VideoSeal					
Покадровая	0,003	44,9	0,0037	0,045	T
Первый кадр группы	0,312	59,7	0,0001	0,014	$\lceil T/N \rceil$
Перенос модификации	0,062	45,0	0,0049	0,070	$\lceil T/N \rceil$
Групповая агрегация	0,006	47,5	0,0022	0,027	$\lceil T/N \rceil$
TrustMark					
Покадровая	0,003	44,1	0,0027	0,026	T
Первый кадр группы	0,271	58,9	0,0001	0,010	$\lceil T/N \rceil$
Перенос модификации	0,030	44,2	0,0032	0,038	$\lceil T/N \rceil$
Групповая агрегация	0,007	44,3	0,0024	0,026	$\lceil T/N \rceil$
DCT CAISS					
Покадровая	0,000	41,6	0,0783	0,523	T
Первый кадр группы	0,333	56,4	0,0026	0,414	$\lceil T/N \rceil$
Перенос модификации	0,000	41,7	0,0773	0,294	$\lceil T/N \rceil$
Групповая агрегация	0,000	43,4	0,0567	0,223	$\lceil T/N \rceil$

Высокие показатели незаметности при маркировании только первого кадра обусловлены отсутствием ЦВЗ в остальных кадрах группы и сопровождаются значительным ростом BER. При одинаковом числе вызовов исходного алгоритма предлагаемая групповая агрегация не уступает переносу модификации первого кадра ни по одной из рассматриваемых метрик. По сравнению с независимым покадровым маркированием предлагаемая методика сокращает число вызовов исходных алгоритмов и улучшает показатели пространственной незаметности для всех трех методов. При этом BER не изменяется для DCT CAISS и незначительно возрастает для VideoSeal и TrustMark, тогда как временные искажения уменьшаются либо остаются на прежнем уровне.

Отдельно исследовано влияние размера группы кадров на устойчивость при H.264-сжатии с параметром CRF=23 и вычислительную эффективность. Его увеличение сокращает число вызовов исходных алгоритмов и уменьшает пространственные искажения, однако сопровождается постепенным ростом доли ошибочно извлеченных бит сообщения. Зависимости устойчивости извлечения и вычислительной эффективности от размера группы представлены на рисунке 6.



а) Зависимость устойчивости ЦВЗ к сжатию кодеком H.264 от размера группы кадров



б) Зависимость ускорения от размера группы кадров

Рисунок 6 — Влияние размера группы кадров на устойчивость извлечения и вычислительную эффективность предложенной методики

Для VideoSeal при размере группы из 30 кадров ускорение внедрения составило 19,2 раза, а извлечения — 18,8 раза относительно покадровой обработки. Значение $N = 30$ использовалось в последующих экспериментах как компромисс между устойчивостью извлечения, незаметностью маркирования и вычислительной эффективностью.

Исследовано влияние временного распределения модификации между кадрами группы. Равномерное распределение обеспечивает наименьшие пространственные искажения, однако может приводить к локальным изменениям на границах соседних групп. Использование оконных функций позволяет сгладить такие переходы. Среди исследованных вариантов наиболее сбалансированные результаты получены для окна Тьюки с $\alpha = 0,25$.

Устойчивость группового маркирования оценивалась при различной интенсивности H.264-сжатия. При слабом и умеренном сжатии устойчивость группового варианта оставалась близкой к покадровому применению исходных методов. При сильном сжатии различие увеличивалось, и групповая обработка характеризовалась более высоким BER. Величина снижения устойчивости при этом зависела от используемого метода маркирования.

Также исследовано влияние нарушения временной синхронизации между этапами внедрения и извлечения. Рассогласование границ групп приводило к росту BER и снижению уверенности извлечения, причем наибольшее ухудшение наблюдалось при сдвиге, близком к половине размера группы. Для восстановления фазы группировки предложена стратегия перебора возможных временных сдвигов с выбором положения границ групп, обеспечивающего максимальную уверенность извлечения ЦВЗ.

В разделе 5.4 рассматриваются ограничения методики. Ее характеристики зависят от исходного метода маркирования и размера группы, а сильные воздействия могут дополнительно снижать устойчивость. При вставке или удалении кадров может потребоваться повторная синхронизация последующих групп.

В **заключении** приведены основные результаты работы, которые заключаются в следующем:

1. Разработана методика унифицированной воспроизводимой оценки методов маркирования медиаконтента и создана реализующая ее программная система WIBE. Методика обеспечивает единую организацию экспериментального процесса для различных типов медиаданных при использовании модально-зависимых наборов данных, преобразований, атак и метрик. Экспериментальная апробация на методах маркирования изображений и аудиоданных показала необходимость совместного учета незаметности, емкости, устойчивости и характера рассматриваемых воздействий.
2. Разработан и экспериментально исследован метод CPW контроля целостности изображений на основе подписываемого перцептивного представления. Предложенный способ повышения побитовой устойчивости перцептивного хеш-кода позволяет сохранять подписываемое представление после рассматриваемых допустимых преобразований. Метод обеспечивает проверяемую связь между содержимым изображения, внедренным аутентификатором и заданным открытым ключом и позволяет выявлять рассматриваемые содержательно значимые модификации.
3. Разработан и экспериментально исследован метод CryptoMark контроля целостности аудиоданных на основе независимого формирования и проверки контентно-зависимых аутентификаторов последовательных временных фрагментов. Пофрагментная организация позволяет формировать карту целостности аудиозаписи, выявлять содержательно значимые модификации и локализовать их во времени при сохранении работоспособности для большинства исследованных допустимых преобразований.
4. Разработана и экспериментально исследована методика переноса методов маркирования изображений на видеоданные, основанная на однократной обработке агрегированного представления группы последовательных кадров и распределении сформированной модификации между кадрами. Методика позволяет применять существующие методы без изменения их архитектуры и дополнительного обучения и снижает вычислительные затраты по сравнению с независимой покadroвой обработкой. Экспериментально исследованы компромиссы между размером группы кадров, незаметностью, устойчивостью извлечения и вычислительной эффективностью, а также влияние временного распределения модификации и нарушения синхронизации.

Публикации автора по теме диссертации

1. Маркирование текстовых документов на экране монитора посредством изменения яркости фона в областях межстрочных интервалов / А. Ю. Якушев [и др.] // Труды Института системного программирования РАН. — 2021. — Т. 33, № 4. — С. 147—162.
2. Docmarking: Real-time screen-cam robust document image watermarking / A. Yakushev [и др.] // 2022 Ivannikov Ispras Open Conference (ISPRAS). — IEEE, 2022. — С. 142—150.
3. Система маркирования документов для проведения расследований при их утечке / Д. О. Обыденков [и др.] // Труды Института системного программирования РАН. — 2021. — Т. 33, № 6. — С. 161—174.
4. Предотвращение анонимных утечек конфиденциальных документов / Д. О. Обыденков [и др.] // Материалы 33-й научно-технической конференции «Методы и технические средства обеспечения безопасности информации». — Санкт-Петербург, 2024. — С. 114—115.
5. Водяные знаки как инструмент защиты видеоконтента / А. Ю. Якушев [и др.] // Материалы 33-й научно-технической конференции «Методы и технические средства обеспечения безопасности информации». — Санкт-Петербург, 2024. — С. 117—118.
6. WIBE: Watermarks for generated Images—Benchmarking & Evaluation / A. Yakushev [и др.] // 2025 40th IEEE/ACM International Conference on Automated Software Engineering (ASE). — IEEE, 2025. — С. 4033—4036.
7. WARP: A Unified Benchmark for Invisible Image Watermarking – Robustness and Protection Against Attacks / K. Abud [и др.] // Proceedings of the 34th ACM International Conference on Multimedia (MM '26). — ACM, 2026. — Принято к публикации.
8. Контроль целостности изображений на основе цифровых водяных знаков, перцептивного хеширования и цифровой подписи / А. Ю. Якушев [и др.] // Труды Института системного программирования РАН. — 2026. — Т. 38, № 5. — С. 133—154.
9. AWA: выравнивание весов на основе активаций для повышения устойчивости статических методов маркирования нейронных сетей / А. А. Акименков [и др.] // Труды Института системного программирования РАН. — 2026. — Т. 38, № 4. — С. 225—244.
10. Exploring Weaknesses of Generative Image Watermarks against Latent Frequency Masking / K. Aistov [и др.] // 2026 IEEE International Conference on Data Mining (ICDM). — IEEE, 2026. — Принято к публикации.

Якушев Алексей Юрьевич

Методы и средства защиты медиаконтента на основе цифровых водяных знаков

Автореф. дис. на соискание ученой степени канд. техн. наук

Подписано в печать _____._____._____. Заказ № _____

Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.

Типография _____