На правах рукописи

БАШМАКОВ Даниил Андреевич

МЕТОДЫ И АЛГОРИТМЫ ВЫЯВЛЕНИЯ ВСТРОЕННЫХ СООБЩЕНИЙ В ПРОСТРАНСТВЕННОЙ ОБЛАСТИ НЕПОДВИЖНЫХ ИЗОБРАЖЕНИЙ ПРИ МАЛОЙ ПОЛЕЗНОЙ НАГРУЗКЕ

Специальность 05.13.19 – Методы и системы защиты информации, информационная безопасность

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата технических наук

Санкт-Петербург – 2018

Работа выполнена на кафедре Проектирования и безопасности компьютерных систем Федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

Научный руководитель: **КОРОБЕЙНИКОВ Анатолий Григорьевич**,
д. т. н., профессор, профессор кафедры Проектирования и безопасности компьютерных систем федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

Официальные оппоненты: **СОКОЛОВ Сергей Сергеевич**,
д.т.н., доцент, заведующий кафедрой Комплексного обеспечения информационной безопасности федерального государственного бюджетного образовательного учреждения высшего образования "Государственный университет морского и речного флота имени адмирала С.О. Макарова"

СИНЮК Александр Демьянович,
д.т.н., доцент, профессор кафедры Безопасности инфокоммуникационных систем специального назначения Военной академии связи имени Маршала Советского Союза С.М. Буденного

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Поволжский государственный технологический университет»

Защита диссертации состоится "20" декабря 2018 г. в 15:30 часов на заседании совета по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук Д 002.199.01, созданного на базе Федерального государственного бюджетного учреждения науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН) по адресу:
199178, Санкт-Петербург, 14-а линия В.О., 39, комн. 401.
Факс: (812)-328-44-50 тел: (812)-328-34-11.

С диссертацией и авторефератом можно ознакомиться на сайте Федерального государственного бюджетного учреждения науки Санкт-Петербургского института информатики и автоматизации Российской академии наук.
http://www.spiiras.nw.ru/dissovet/

Автореферат разослан "____"_____2018 г.

Ученый секретарь совета Д 002.199.01
кандидат технических наук Зайцева А.А.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Методы встраивания информации в контейнеры всех видов (изображения, аудио, видео и другие) находят широкое применение в современном мире. Встраивание информации применяется как в легальных целях, так и в ходе противоправной деятельности. К способам легального применения встраивания информации относятся цифровые водяные знаки в документах и объектах интеллектуальной собственности, организация каналов скрытной передачи информации спецслужбами в рамках их деятельности, организация каналов скрытной передачи информации в прочих целях (сохранение коммерческой тайны, тайны переписки и т. д.). К наиболее распространённым способам противоправного применения технологии встраивания информации относится организация каналов скрытной передачи информации в целях обеспечения противоправной деятельности, в частности, террористических актов. Кроме того, к противоправному применению встраивания информации следует относить действия по сокрытию факта выведения информации за охраняемый периметр в рамках, например, промышленного шпионажа, также осуществляемые на основе каналов скрытной передачи информации.

По этим причинам разработка и реализация методов и алгоритмов выявления встроенных сообщений в контейнерах различной природы необходимы как для противодействия противоправным способам применения встраивания, так и для контроля и проверки случаев легального применения. В частности, методы выявления встроенных сообщений целесообразно применять в задачах выявления цифровых водяных знаков с целью доказательства модификации контейнера в ситуации, когда извлечение сообщения штатными средствами невозможно (например, в случае недоступности ключевой информации встраивания). Другими областями возможного применения этих методов являются пассивное противодействие каналам скрытной передачи информации, используемых в нелегальных целях, в части обнаружения и классификации таких каналов, а также противодействие промышленному шпионажу в части обнаружения попыток выведения охраняемой информации за периметр охраняемой зоны.

Неподвижные цифровые изображения являются одним из наиболее распространённых видов информации, передаваемой в сети Интернет. Изображения могут выступать контейнером для встраивания сообщений. На сегодняшний день разработано множество способов встраивания в неподвижные цифровые изображения.

В условиях постоянно возрастающей разрешающей способности изображений, отношение объёма встраиваемой полезной нагрузки к объёму изображения как стеганоконтейнеру, или отношение «нагрузка-контейнер» (ОНК), постоянно снижается. Этим обуславливается тот факт, что задача выявления факта встраивания информации в неподвижные цифровые изображения при малых значениях ОНК, или при «малой полезной нагрузке» - МПН, и оценки эффективности методов такого выявления (методов

стеганодетектирования) в настоящее время имеет повышенный научный интерес.

Методы встраивания подразделяются в зависимости от пространства (домена) изображения. Пространственный домен изображения в наименьших значащих битах (НЗБ) является одним из наиболее распространённых доменов, используемых для стеганографического встраивания. Несмотря на то, что встраивание в плоскость НЗБ не лишено недостатков, высокая скрытность и большой объём контейнера обеспечивают в настоящее время его большую популярность. Существует множество методов выявления в пространственном домене изображения в плоскости НЗБ. Однако в условиях МПН эти методы показывают низкую эффективность. Это, в свою очередь, не позволяет эффективно противодействовать скрытным каналам передачи информации, основанным на встраивании в пространственный домен изображения

Таким образом, тенденция постоянного роста размеров стеганоконтейнеров совместно со снижением эффективности стеганодетектирования, обусловленным непрерывным уменьшением значения ОНК, определяет требование к методам выявления встроенных сообщений, заключающееся в дальнейшем повышении эффективности этих методов при малой полезной нагрузке, как общей необходимой реакции на изменение условий противодействия каналам передачи данных, основанных на встраивании информации. Эти обусловливается достаточно высокая актуальность выбранной темы диссертационной работы и полученных в ней результатов.

Степень разработанности темы. В работах В.И. Коржика обоснован общий подход по применению методов статистического стеганоанализа в задаче выявления встроенных сообщений в неподвижных изображениях. Основой для диссертационного исследования послужили работы Д. Фридрич, А. Д. Кера и М. Голяна по разработке методов стеганодетектирования в пространственной области неподвижных цифровых изображений. В ходе работы выполнен анализ существующих методов повышения эффективности стеганодетектирования, предложенных в работах Р. Бёме, Ю. Сяо, Б. Нобору, П. Шоттле и других. Также заделом для диссертационного исследования послужили работы В. Г. Грибунина, Р. М. Юсупова, А. В. Аграновского, И. В. Туринцева, А.А. Молдовяна, Н.А. Молдовяна, И. Н. Окова, Г. Ф. Кохановича и других исследователей.

Целью работы является повышение эффективности выявления встроенных сообщений в НЗБ пикселей неподвижных изображений при малой полезной нагрузке в интересах обеспечения защищённости информации путём предотвращения её утечек по каналам передачи информации на основе стеганографии.

Для достижения поставленной цели в диссертационной работе решалась **научная задача** по разработке модели, алгоритмов и метода выявления встроенных сообщений в плоскости наименьших значащих бит

пространственной области неподвижных изображений при малой полезной нагрузке.

Поставленная научная задача декомпозируется на следующие **частные задачи**:

1. Анализ существующих методов выявления встроенных сообщений в НЗБ неподвижных цифровых изображений при малой полезной нагрузке и выявление среди них наиболее эффективного на сегодняшний день.
2. Разработка модели выявления встроенных сообщений методом, определённым в задаче 1 в условиях малой полезной нагрузки, анализ модели и определение направлений по усовершенствованию метода.
3. Разработка алгоритмов, обеспечивающих повышение эффективности выявления встроенных сообщений в НЗБ неподвижных изображений при малой полезной нагрузке.
4. Разработка метода выявления встроенных сообщений в НЗБ фоновых зон неподвижных цифровых изображений с повышенной точностью при малой полезной нагрузке.
5. Экспериментальное подтверждение повышенной эффективности стеганодетектирования при применении метода выявления встроенных сообщений, разработанного в задаче 4.

Научная новизна положений, выносимых на защиту, состоит в следующем:

1. Разработанная модель выявления встроенных сообщений в наименьших значащих битах фоновых зон пространственной области неподвижных изображений отличается от существующих ориентацией на учет особых семантических областей анализируемого изображения – фоновых зон. Построение модели стеганодетектирования в фоновых зонах, а также анализ зависимости эффективности детектирования от особенностей работы метода в фоновых зонах изображения произведены впервые.
2. Алгоритмы выявления встроенных сообщений в НЗБ фоновых зон неподвижных изображений отличаются от известных учетом крупных структур анализируемых пикселей, специфичных для фоновых зон естественных изображений. В силу этого алгоритмы оперируют способами выделения соседства пикселей в специфичных областях изображения (в отличие от DIH, WS, SPAM и др.), сочетая это с использованием в процессе анализа контейнера накопленной статистики (в отличие от RS, SPA и др.).
3. Разработанный метод повышения эффективности выявления встроенных сообщений, в отличие от известных, фокусируется на прогнозе пикселей анализируемого изображения с точностью, критичной для стеганодетектирования в условиях малой полезной нагрузки, и применяет алгоритм выделения фоновой зоны изображения, специфичный для задачи стеганодетектирования методом WS.

Теоретическая и практическая значимость результатов заключается в использовании предложенного метода выявления встроенных сообщений в системах защиты информации, в частности, в компонентах пассивного

противодействия каналам передачи данных, основанных на стеганографии в плоскости НЗБ неподвижных цифровых изображений, что позволит повысить уровень защищённости информации за счёт снижения вероятности реализации риска её несанкционированной утечки по таким каналам.

Методология исследования заключается в постановке и формализации задач, связанных с оценкой эффективности методов и алгоритмов выявления встроенных сообщений, описании модели сущностей, используемых для проведения оценок, разработке модели, методов и алгоритмов выявления встроенных сообщений в неподвижных изображениях, апробации полученных теоретических результатов посредством сравнительного анализа их реализаций с существующими решениями с получением количественных и качественных сравнительных оценок.

Методы исследований. Поставленные задачи решены на основе применения теории защиты информации, теории вероятности и математической статистики, методов дискретной математики.

В соответствии с заявленными целью и задачами работы, **объектом исследования** являются контейнеры для встраивания, являющиеся неподвижными изображениями с информацией, встроенной в наименьшие значащие биты пространственной области. **Предметом исследования** являются методы и алгоритмы выявления встроенных сообщений в неподвижных изображениях при малой полезной нагрузке.

На защиту выносятся следующие основные положения:

1. Модель выявления встроенных сообщений в наименьших значащих битах фоновых зон пространственной области неподвижных изображений при малой полезной нагрузке обеспечивает оптимальный подход к выявлению встроенных сообщений в фоновых зонах.
2. Алгоритмы выявления встроенных сообщений в наименьших значащих битах фоновых зон пространственной области неподвижных изображений при малой полезной нагрузке обеспечивают повышенную точность прогноза пикселей анализируемого изображения в фоновых зонах.
3. Метод выявления встроенных сообщений в наименьших значащих битах пространственной области неподвижных изображений обеспечивает повышенную эффективность выявления встроенных сообщений при малой полезной нагрузке.

Достоверность полученных результатов определяется использованием апробированного математического аппарата, системным подходом при описании объекта исследования, проведением сравнительного анализа полученных результатов с существующими показателями, использованием проверенных методик в оценке эффективности методов стеганодетектирования, а также результатами практических экспериментов.

Апробация результатов. Основные результаты работы представлялись на следующих конференциях:

- Всероссийская научно-практическая конференция с международным участием «Информационные технологии в профессиональной деятельности и научной работе», 2014 г.
- III Всероссийский конгресс молодых учёных, 2014 г.
- Всероссийский студенческий форум «Инженерные кадры - будущее инновационной экономики России», 2015 г.
- V Всероссийский конгресс молодых учёных, 2016 г.
- VI Всероссийский конгресс молодых учёных, 2017 г.

Результаты работы **внедрены** в качестве составного компонента системы защиты информации Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова Российской академии наук ИЗМИРАН (северо-западного филиала в Санкт-Петербурге), в систему электронного документооборота АО «ОКБ Электроавтоматика», а также в образовательный процесс на кафедре Проектирования и безопасности компьютерных систем Университета ИТМО.

Публикации по теме диссертации. По результатам диссертационного исследования опубликовано 9 работ, из них 5 статей в изданиях из перечня ВАК и 1 статья в журнале, индексируемом в международной базе цитирования Scopus.

Структура диссертации. Диссертация состоит из введения, четырёх глав, заключения и списка литературы, состоящего из 104 пунктов, включающих труды автора. Материал изложен на 150 страницах машинописного текста, содержит 39 рисунков и 10 таблиц. Диссертация включает в себя 2 приложения.

СОДЕРЖАНИЕ

Во введении обоснована актуальность темы диссертационной работы; сформулированы цели и задачи исследования; сформулированы положения, выносимые на защиту, показана их научная новизна, показана практическая значимость работы, приведены сведения об апробации результатов проведённого исследования и об опубликованных статьях по теме работы.

В первой главе анализируются требования, предъявляемые к методу выявления встроенных сообщений (ВВС), приводится методика оценки эффективности метода ВВС. Под эффективностью метода ВВС понимается способность достигать определённого процента корректных классификаций в задаче различения стеганопосылки и чистого сообщения.

Независимость от семантики анализируемого изображения подразумевает под собой способность показывать близкие значения эффективности при анализе изображений, содержащих определённые специфические области, например, однородный фон, или, напротив, сильно зашумлённые области. Под семантикой изображения понимается наличие и распределение в пространственной области изображения определённых зон, удовлетворяющих

требованиям по характеру распределения значений яркости пикселей в них (семантических зон).

Показана связь эффективности метода ВВС с риском утечки информации по каналу несанкционированной передачи данных, основанном на стеганографическом встраивании через формулу:

$$R = (P_{\text{реал}} - P_{\text{прот}}) * U = \left(P_{\text{реал}} - \sum P_{\text{сп}}\right) * U$$

где R – риск, $P_{\text{реал}}$ – вероятность реализации угрозы, $P_{\text{прот}}$ – вероятность противодействия угрозе с использованием средств защиты информации, U – ущерб от реализации угрозы, $P_{\text{сп}}$ – вероятность противодействия с использованием конкретного средства защиты информации.

Рассматриваемое средство пассивного противодействия каналу передачи информации является одним из средств защиты информации. Вероятность корректного обнаружения встроенного сообщения определяет эффективность средства пассивного противодействия каналу передачи информации. Поскольку эффективность ВВС напрямую влияет на вероятность обнаружения угрозы, она является основополагающей характеристикой при выборе метода стеганодетектирования.

Обоснован способ оценки эффективности метода ВВС. Метод статистического ВВС рассматривается с двух позиций: как инструмент оценки длины сообщения, встроенного в анализируемое изображение, и как инструмент бинарной классификации, позволяющий отнести анализируемое изображение к стеганопосылкам или к чистым изображениям.

Предложен следующий сценарий эксперимента, позволяющего оценить эффективность метода стеганодетектирования с обеих позиций:

1. Выборка изображений. В качестве выборок использованы наборы изображений коллекций BOSS, BOWS2, eTrim, Places, UCID.
2. Имитация для части изображений стеганографического встраивания в LSB.
3. Анализ изображения тестовой выборки с помощью метода стеганодетектирования. Сравнительная оценка длины встроенного сообщения либо факта наличия встраивания и действительной длины сообщения и факта его наличия.
4. Построение статистики эффективности анализируемого метода стеганодетектирования на основе результатов сравнительной оценки. В качестве показателя эффективности метода как средства оценки длины встроенного сообщения предложено использовать среднюю ошибку прогноза длины сообщения в пикселях, получаемую при анализе набора изображений:

$$E_L = \frac{\sum_1^N |L_a - L_p|}{N},$$

где N – число анализируемых последовательно изображений, L_a – прогнозируемая длина встроенного сообщения, L_p – действительная длина встроенного сообщения.

В качестве показателя эффективности метода как бинарного классификатора предложено использовать показатели, применяемые в существующем инструментарии оценки эффективности бинарной классификации (методе Монте-Карло), оперирующем понятиями ложной положительной и ложной отрицательной классификации. Предложен способ оценки собственной эффективности метода стеганодетектирования – кривые доверительных интервалов, выражающие зависимость между вероятностью корректной классификации и вероятностью ложноположительной классификации при изменяющемся значении пороговой величины. В качестве численной оценки эффективности при этом используется значение вероятности ложной положительной классификации при определённом значении вероятности корректной классификации. Используется исходное значение корректной классификации 95%, что, в соответствии с рядом работ отечественных и зарубежных учёных, считается оптимальной Байесовой стратегией для оценки бинарной классификации в задаче стеганодетектирования.

На основании анализа требований к алгоритму статистического стеганодетектирования отобраны алгоритмы для дальнейшего исследования и оценки эффективности: Triples analysis, Sample pairs analysis, RS-analysis, Weighted Stego Image, Difference Image Histogram.

Во второй главе проводится анализ характера распределения значений яркости пикселей в пространственной области распределения неподвижного цифрового изображения и искажений, вносимых стеганографическим встраиванием. Анализируются семантические зоны изображений, имеющие особое значение в задаче стеганодетектирования.

Описывается модель стеганографического встраивания в LSB пикселей пространственного распределения изображения. Описывается влияние LSB встраивания на характер распределения значений пикселей в пространственной области изображения. LSB-встраивание детектируется по аномальному изменению значения яркости пикселя на «1» по сравнению с ожидаемым. Ожидаемое значение при этом прогнозируется методом стеганодетектирования, исходя из данных, доступных из самого анализируемого изображения.

Вводится понятие естественного изображения, полученного методом фотографии и изображающего предметы реального мира, в противовес искусственному изображению – рисункам и графике. В пространственной области затруднительно выделить общие характеристики распределения значений уровней яркости пикселей для последовательного набора случайных естественных изображений. Тем не менее, в пространственной области можно выделить крупные семантические структуры, повторяющиеся на большом наборе естественных изображений – однородные фоновые зоны.

Под однородными фоновыми зонами подразумеваются такие зоны изображения, в которых значения соседних пикселей изменяются не более чем на определённую величину в среднем для зоны и не более чем на

определённую величину от среднего значения пикселей для той же зоны. Математически однородная фоновая зона определяется следующим образом. Пусть I – множество пикселей изображения. Тогда однородная фоновая зона B – это такая наибольшая зона изображения, что

$$\forall p \in B: |p - p_N| < T_N, |p - p_A| < T_A,$$

где p – значение яркости пикселя, p_N – значение яркости соседнего пикселя, наиболее отличающегося по значению среди всех соседних пикселей данного, p_A – среднее значение всех пикселей, принадлежащих множеству B, T_N, T_A – пороги, вводимые индивидуально для каждого изображения.

Характер искажений, вносимых встраиванием в LSB, определяет важность наличия и доли фоновых зон в анализируемом изображении. В фоновых зонах, где разница между значениями соседних пикселей невелика, естественный характер распределения значений пикселей мало отличим от следов встраивания в LSB. Таким образом, можно ожидать падения эффективности стеганодетектирования в изображениях с большой долей однородного фона.

Проводится исследование возможностей современных методов статистического стеганодетектирования в неподвижных цифровых изображениях. Исследуются методы, отобранные в главе 1. ROC-кривые, полученные для наглядной демонстрации эффективности исследуемых методов в задаче обнаружения факта стеганографического встраивания, представлены на рисунке 1. Из графиков на рисунке 1 видно, что эффективность современных методов ВВС на малых значениях ОНК характеризуется большой долей ложно положительной классификации при доле корректного срабатывания, равной 95%. Сделан вывод о необходимости усовершенствования существующих методов стеганодетектирования. Также из графиков видно, что вне зависимости от значения ОНК, наибольшую эффективность демонстрирует метод ВВС Weighted Stego (WS).

Значимость эффективности ВВС именно на малых значениях ОНК обусловлена тем, что в современном мире разрешение изображений, и, следовательно, объём стеганоконтейнеров на их основе постоянно растёт.

Для подтверждения предположения о зависимости эффективности ВВС от доли однородного фона проведён эксперимент. Тестовая выборка изображений разделена на подвыборку изображений с высокой долей однородного фона, превышающей 40%, (подвыборку HB) и подвыборку изображений с малой долей однородного фона, не достигающей 5%, (подвыборку LB). Кривые на рисунке 2 показывают зависимость эффективности стеганодетектирования методом WS от доли однородного фона. Видно, что на подвыборке HB метод обеспечивает меньшую эффективность ВВС.

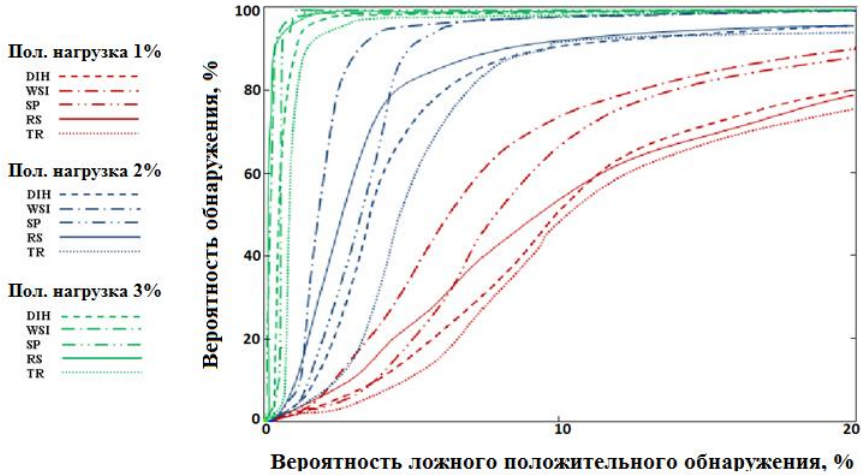

Рисунок 1 – Возможности современных методов выявления встроенных сообщений в неподвижных цифровых изображениях

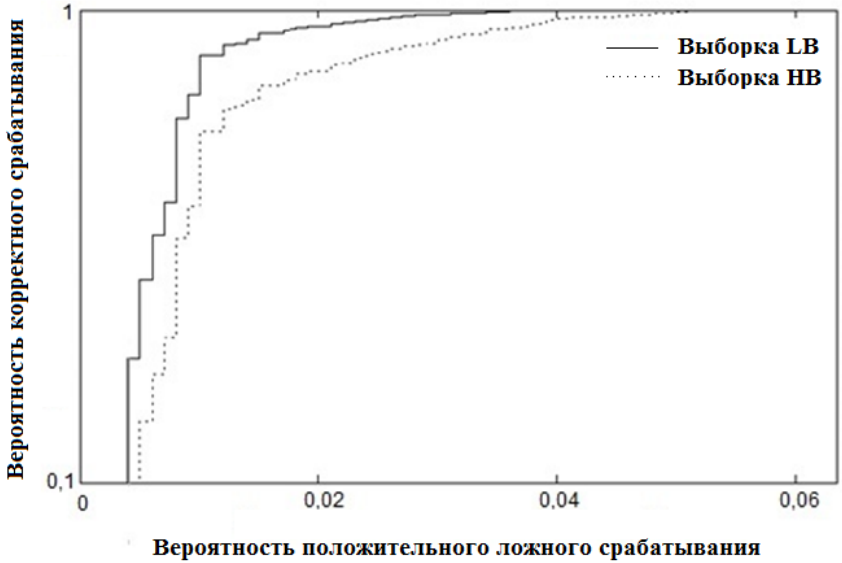

Рисунок 2 – Падение эффективности ВВС методом WS в условиях большой доли однородного фона в изображении

В третьей главе исследуются причины существования зависимости эффективности ВВС методом WS от доли однородного фона в анализируемом изображении. Разрабатывается и анализируется модель ВВС встроенных сообщений в фоновых зонах неподвижных изображений при малых значениях

ОНК. Разрабатываются алгоритмы ВВС через прогноз пикселей анализируемого изображения в фоновой зоне.

Низкая эффективность метода WS при анализе изображений с большой долей однородного фона объясняется высокой долей ложно положительных классификаций, что показывают данные таблицы 1. Полученная доля ложных классификаций соответствует величине ОНК 3%.

Таблица 1

Доля ложных классификаций в зависимости от доли фоновой зоны изображения

Выборка / Подвыборка	HB	LB
Чистые изображения (Доля ПЛК)	22,4%	13,1%
Стеганограммы (Доля ОЛК)	10,0%	9.5%

При выявлении методом WS оценивается длина встроенного сообщения. Поскольку длина есть единственное выходное данное метода, от точности её оценки напрямую зависит эффективность стеганодетектирования методом WS.

Длина встроенного сообщения оценивается по следующей формуле:

$$M = 2\sum_{i=1}^{N}(s_i - \hat{c}_i)(s_i - \overline{s_i}),$$

где M – оцененная длина сообщения, N – размер изображения в пикселях, s_i – фактическое значение пикселя анализируемого изображения, $\overline{s_i}$ – фактическое значение пикселя анализируемого изображения с инвертированным НЗБ, $\hat{c_i}$ – значение пикселя, спрогнозированное методом WS. Каждый элемент суммы есть взвешенное представление пикселя.

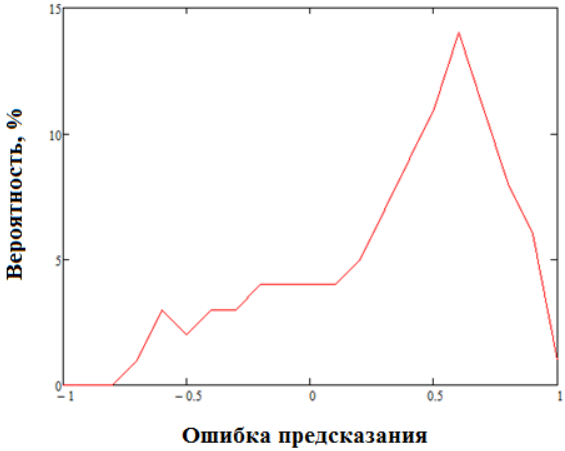

Рисунок 3 – Ошибка прогноза значения пикселя методом WS

Итоговая длина сообщения складывается из разности между спрогнозированным и действительным пикселем изображения-стеганограммы. Множитель $(s_i - \overline{s_i})$ не зависит ни от чего, кроме одного отдельно взятого пикселя изображения для каждого слагаемого итоговой суммы, и, следовательно, распределение его значений не зависит от доли фона в анализируемом изображении. Учитывая, что растёт именно доля положительных ложных срабатываний, то есть эффект наблюдается в том числе при анализе только чистых изображений, полагается, что в фоновых зонах систематически завышается спрогнозированное значение пикселя по отношению к реальному в тех случаях, когда $s_i - \overline{s_i} = 1$.

График на рисунке 3 показывает характер распределения разности спрогнозированного и реального значения пикселя $s_i - \hat{c}_i$ для $s_i - \overline{s_i} = 1$ для подвыборки HB. Видно, что прогноз значения пикселя стабильно завышен по сравнению с реальным значением.

Метод WS оперирует четырьмя или восемью соседними пикселями при прогнозе данного. Поскольку задействование всех восьми соседних пикселей не позволяет достичь высокой точности прогноза значения пикселя, усовершенствование алгоритма возможно через задействование большего количества пикселей, окружающих прогнозируемый.

Специальная модель BBC в фоновых зонах неподвижного изображения базируется на идее прогноза значений пикселей фоновой зоны изображения исходя из более широкой области соседства пикселей. Существующее представление о связности значений соседних пикселей естественных изображений расширено для фоновой зоны с учётом более сильной связи значений и утверждения о большем, чем в нефоновых зонах, эффективном диаметре соседства пикселей в задаче прогноза их значений.

Исходя из специфики выявленной связи между точностью прогноза значения пикселя и эффективностью BBC, метод повышения эффективности BBC базируется на трёх алгоритмах прогноза значения пикселей в фоновых зонах.

Алгоритм *прогноза пикселей по кортежам (алгоритм кортежей)* в фоновой зоне задействует более дальние соседние пиксели для прогнозирования данного. Алгоритм основан на идее поиска стабильно повторяющихся последовательностей пикселей в фоновой зоне – кортежей. Используется только ряд пикселей в определённом направлении от данного – кортеж соседних пикселей.

Пусть S – анализируемое изображение, элемент пространственного распределения которого определён как $S(i,j)$, где i,j – координаты пикселя. Тогда n-кортеж для пикселя $S(i,j)$ – кортеж $S(i \pm 1, j) \dots S(i \pm n, j)$. Таким образом, n-кортеж для пикселя представляет собой упорядоченное множество n соседних пикселей в определённом направлении. В зависимости от направления, такие кортежи обозначаются nL, nR, nT и nB (соответственно, для направлений влево, вправо, вверх и вниз).

Для каждого уникального кортежа определяется количество пикселей, для которого он был построен. Самый часто встречающийся пиксель для данного кортежа определяется как спрогнозированный пиксель. Набор пар «кортеж – спрогнозированный пиксель» составляет матрицу кортежей пикселей. При прогнозе значения пикселя анализируемого изображения для него строится n-кортеж. Если такой n-кортеж присутствует в матрице кортежей пикселей, значение пикселя берётся из матрицы для данного кортежа. Если кортеж отсутствует, значение прогнозируется как среднее четырёх соседних пикселей.

Эксперимент с применением прогноза по кортежам пикселей в фоновых зонах показал значительное снижение средней ошибки прогноза пикселей в фоновых зонах. Численные оценки средней ошибки прогноза, приведённые в таблице 2, подтверждают эффективность метода прогноза по кортежам.

Таблица 2

Средняя ошибка прогноза пикселей для выборок HB и LB

Метод / Коллекция	**HB**	**LB**
По среднему	1.22	1.74
По кортежам	0.74	0.96

Алгоритм *адаптивного прогноза в градиентах (алгоритм градиентов)* так же, как и алгоритм кортежей, ставит целью снижение ошибки прогноза пикселей. Снижение ошибки прогноза происходит за счёт адаптации прогнозируемого значения пикселя под выделенные статистические особенности градиентной области, которой принадлежит анализируемый пиксель. Под градиентом подразумевается область изображения, в которой наблюдается устойчивый и равномерный переход от более низких значений яркости пикселей к более высоким. Поскольку в градиенте, исходя из его определения, существует закономерность распределения значений пикселей, нерационально рассматривать задачу прогноза в градиенте как задачу прогноза с заранее неизвестными характеристиками распределения.

За направление градиента принято направление любого из соседних пикселей данного. Таким образом, получено восемь направлений, обозначенных U, D, L, R, UL, UR, DL, DR (Up, Down, Up-Right и так далее). Для выделения градиентов на изображении рассмотрен каждый пиксель фоновой зоны. Выделены градиентные области по следующему правилу: пиксель считается принадлежащим градиентной области направления D, если для любых P его соседей в данном направлении разница между двумя соседними пикселями не отличается от её среднеквадратичного значения на всём рассматриваемом наборе более, чем в k раз, и для N его соседей в направлениях, перпендикулярных данному, для данного направления выполняется то же условие. Таким образом, принадлежность пикселя с множеством перпендикулярных соседей B_N размером N множеству пикселей градиентных областей A_{gr} изображения A определяется следующим условием:

$$\nexists a_i: |a_i - a_{i+1}| > k\sqrt{\sum_{1}^{P}(a_i - a_{i+1})^2} \cap \forall b \in B_n: |b_i - b_{i+1}|$$

$$< k\sqrt{\sum_{1}^{P}(b_i - b_{i+1})^2},$$

$$a \in A_{gr}, i \in [0, P],$$

где a – значение пикселя, k – сила градиента, P – длина градиентного кортежа, b – значение перпендикулярного соседа пикселя a.

P соседей пикселя в выбранном направлении обозначены как его градиентный кортеж. За силу градиента принято среднеквадратичное значение величины изменения значения пикселя по сравнению с предыдущим для его градиентного кортежа. Если по условию пиксель принадлежит нескольким градиентам, то выбирается тот, для которого N больше. Введены условия соседства по N, которые позволяют выделить системные, обширные градиенты и не ухудшать характеристики модели прогноза случайными совпадениями в распределении яркости фоновых пикселей.

Результатом работы алгоритма выделения является множество A_{gr} с указанием для каждого силы и направления градиента, а также связи «пиксель – градиентный кортеж», вместе составляющие матрицу градиентов. Прогнозирование значения пикселя, принадлежащего кортежу, производится подбором значения, наименьшим образом меняющего силу кортежа.

Вводится множество целых значений пикселей U, из которого ведётся подбор данного. Множество значений определяется как среднее значение пикселей, окружающих данный прогнозируемый пиксель, $\pm$ определённое значение допуска R. Несмотря на то, что алгоритм позволяет прогнозировать пиксели, использовать её в отрыве от прогнозирования по кортежам нерационально, так как последний хорошо справляется в фоновых зонах в целом. Для прогноза предложено использовать взвешенное среднее прогнозов по кортежам и в градиентах. Предложенная формула итогового прогноза пикселя имеет следующий вид:

$$a_F = wa_{fm} + (w - 1)a_{fg},$$

где a_{fm} – значение, спрогнозированное по кортежам, a_{gf} – значение, спрогнозированное по градиентному кортежу, w – экспериментально подбираемый вес, $w \in [0; 1]$.

Алгоритм прогноза значений пикселей *за счёт накопления статистики (накопительный алгоритм)* развивает алгоритм кортежей. При рассмотрении задачи стеганодетектирования множества изображений подряд можно использовать статистику кортежей, вычисленную для уже проанализированных изображений. Матрица кортежей, построенная для изображения, дополняется информацией о том, для скольких случаев в

изображении встречается такая пара «пиксель – кортеж». Таким способом образуется гистограмма кортежей. Схожесть гистограмм кортежей двух изображений используется в качестве метрики схожести этих изображений. Это позволяет не использовать гистограммы для анализа изображений, обладающих сильно отличающейся статистикой кортежей.

Факт схожести гистограмм определяется следующим образом:

1. Для двух гистограмм G_1 и G_2 определяются пары «кортеж – пиксель», входящие в обе гистограммы. Множество таких пар обозначается как G_M.

2. Для каждого элемента g множества G_M рассчитывается отношение схожести по формуле:

$$S = \begin{cases} \frac{m(G_1,g)}{m(G_2,g)}, m(G_1,g) < m(G_2,g) \\ \frac{m(G_2,g)}{m(G_1,g)}, m(G_2,g) < m(G_1,g) \end{cases},$$

где $m(G,g)$ – количество вхождений элемента g в гистограмме G.

Определяется количество элементов, для которых отношение схожести S меньше определённого порога S_T. Если доля таких элементов в множестве G_M больше определённого порога T_G, считается, что гистограммы G_1 и G_2 схожи.

Для схожих гистограмм матрицы кортежей объединяются, и при анализе изображения используется объединённая матрица, включающая информацию обеих матриц. При анализе потока изображений образуется множество накопленных гистограмм кортежей $\mathbb{G}$. При анализе очередного изображения из множества определяется гистограмма, обладающая наибольшей степенью схожести с гистограммой данного изображения. Эти гистограммы объединяются. Результат объединения используется для анализа текущего изображения и замещает гистограмму во множестве.

Алгоритмы прогноза по кортежам обладает линейной вычислительной сложностью $O(n)$, где n – длина анализируемого изображения в пикселях. Сложность предложенных алгоритмов не превышает сложность оригинального алгоритма прогноза пикселей, предложенного авторами метода WS.

В четвёртой главе рассматривается вопрос выделения однородного фона изображения, приводится метод ВВС с учётом предложенных алгоритмов прогноза в фоновых зонах и разработанного алгоритма выделения однородного фона, приводятся данные экспериментальных подтверждений повышенной эффективности ВВС методом WS с использованием разработанного метода.

Важность задачи выделения фона изображения обосновывается тем фактором, что методы кортежей и градиентов целесообразно применять именно на фоновых зонах изображений. При этом определение фона, данное выше, имеет недостаток, вызванный необходимостью определять пороги T_N и T_A. Задача определения универсальных порогов нетривиальна. При применении к изображениям такой метод даёт неравномерную картину выделения фоновой зоны, при этом значительная доля пикселей фона не попадает в зону. При анализе открытых источников был найден метод,

позволяющий корректно выделять фоновые зоны изображений – сегментационные (полносвёрточные) нейронные сети.

Приведена итоговая блок-схема реализации метода стеганодетектирования WS для одного изображения с учётом предложенных улучшений, представленная на рисунке 4. Предлагаемый метод повышения эффективности стеганодетектирования включает в себя алгоритм выделения фоновой зоны изображения и алгоритмы прогноза пикселей в фоновых зонах изображения.

Приводятся данные экспериментального подтверждения факта повышения эффективности стеганодетектирования при применении предложенного метода повышения эффективности. ROC-кривые на графике, приведённом на рисунке 5, показывают эффективность ВВС разработанным методом (WSM) по сравнению с оригинальным методом WS. Полученные численные оценки показывают, что применение предложенных усовершенствований позволяет добиться увеличения эффективности ВВС методом WS на 5-7%.

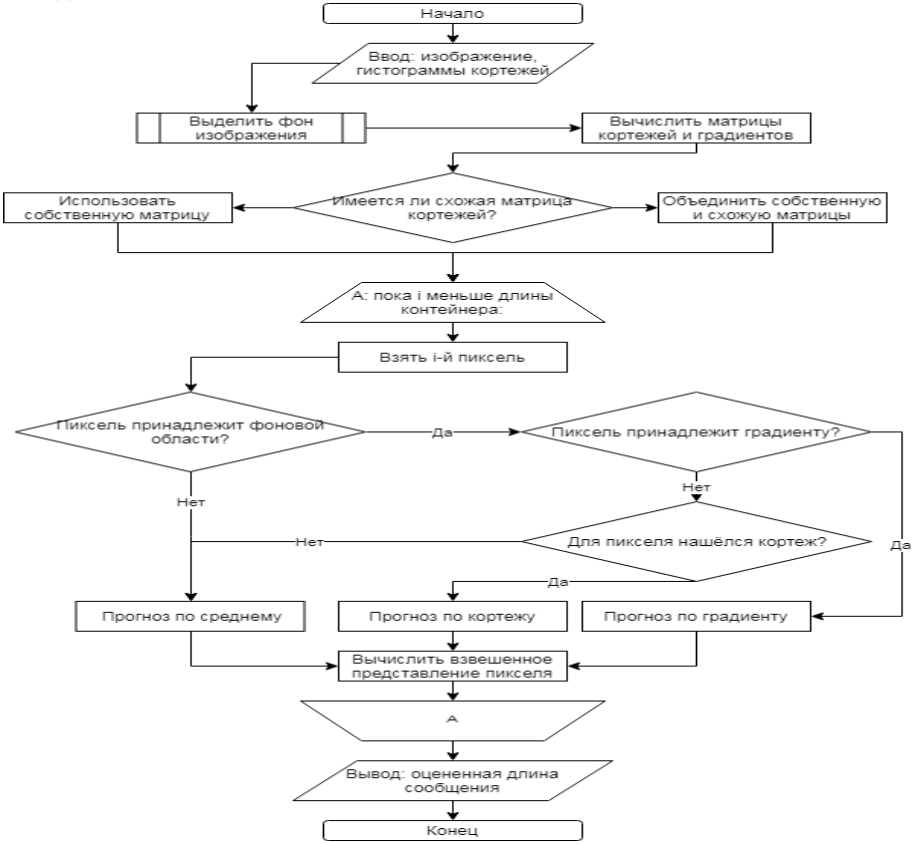

Рисунок 4 – Схема алгоритма ВВС с учётом предложенных усовершенствований

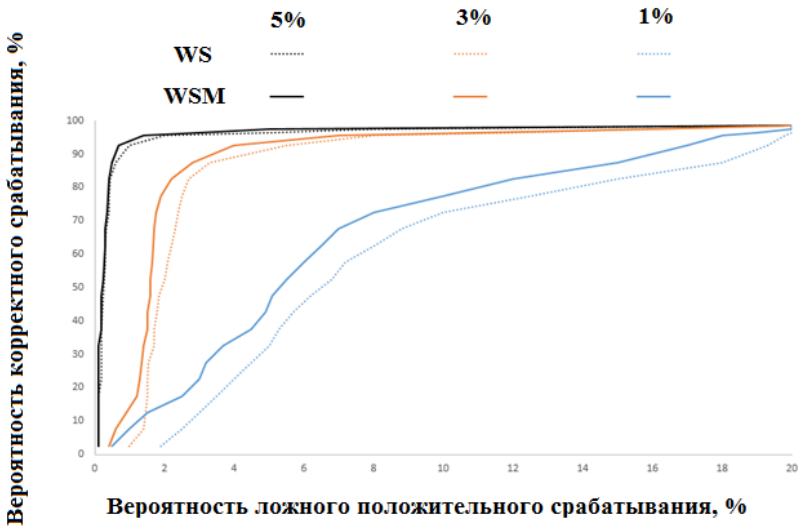

Рисунок 5 – Эффективность ВВС методом WS с учётом предложенных усовершенствований и без них

ЗАКЛЮЧЕНИЕ

В диссертационной работе решена научная задача по разработке модели, алгоритмов и метода ВВС в плоскости наименьших значащих бит пространственной области неподвижных изображений при малой полезной нагрузке за счёт использования специальной модели, алгоритмов и метода ВВС в фоновых зонах естественных изображений. Были получены следующие научные результаты, составляющие **итоги** исследования:

1. Проведён анализ существующих методов ВВС в неподвижных цифровых изображениях в НЗБ. Отобраны методы, обладающие наибольшей эффективностью из доступных.
2. Проведён анализ эффективности отобранных методов ВВС, сделан вывод о недостаточной эффективности методов ВВС при малых значениях отношения нагрузка-контейнер. Сделан вывод о том, что метод Weighted Stego Image (WS) обладает наивысшей эффективностью в задаче ВВС в неподвижных цифровых изображениях из рассматриваемых.
3. Разработана и проанализирована модель ВВС методом WS в условиях малых значений ОНК. Сделан вывод о зависимости эффективности метода ВВС от доли однородного фона в анализируемом изображении.
4. Разработаны алгоритмы ВВС в НЗБ фоновых зон неподвижных изображений при малой полезной нагрузке.
5. Разработан метод ВВС в НЗБ фоновых зон неподвижных цифровых изображений с повышенной точностью при малой полезной нагрузке.

Сформулированы **рекомендации** по применению результатов работы в различных прикладных областях. Представленные результаты дают инструмент для построения систем противодействия каналам скрытной передачи данных на основе встраивания. Разработанные алгоритмы и метод рекомендуется применять при противодействии встраиванию в НЗБ неподвижных изображений в ситуациях, когда, исходя из доступных потенциальному нарушителю данных, ожидается использование естественных изображений в качестве контейнеров для встраивания либо объём доступных нарушителю данных позволяет организовать канал передачи данных на малых значениях ОНК.

В качестве **направлений дальнейших исследований** предлагается проведение дальнейшего анализа модели ВВС в фоновых зонах неподвижных изображений с выделением более глубоких особенностей распределения значений пикселей в таких зонах; анализ и выделение более сложных структур пикселей, прилегающих к анализируемому пикселю в фоновых зонах, позволяющих добиться большей точности прогноза пикселей; выделение и применение в задаче прогноза значений пикселей других структур в фоновых зонах.

Соответствие паспорту специальности. Положения, выносимые на защиту, соответствуют следующим пунктами паспорта специальности 05.13.19 – «Методы и системы защиты информации, информационная безопасность»: «5. Методы и средства (комплексы средств) информационного противодействия угрозам нарушения информационной безопасности в открытых компьютерных сетях, включая Интернет»; «6. Модели и методы формирования комплексов средств противодействия угрозам хищения (разрушения, модификации) информации и нарушения информационной безопасности для различного вида объектов защиты вне зависимости от области их функционирования».

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации из перечня ВАК:

1. Башмаков Д.А. Точность предсказания пикселей фоновых областей цифровых изображений в задаче стеганоанализа методом Weighted Stego // Кибернетика и программирование. – 2018. – № 2. – С. 38-47. DOI: 10.25136/2306-4196.2018.2.25706. URL: http://e-notabene.ru/kp/article_25706.html

2. Башмаков Д.А. Адаптивное предсказание пикселей в градиентных областях для улучшения точности стеганоанализа в неподвижных цифровых изображениях // Кибернетика и программирование. – 2018. – № 2. – С. 83-93. DOI: 10.25136/2306-4196.2018.2.25514. URL: http://e-notabene.ru/kp/article_25514.html

3. Башмаков Д.А., Прохожев Н.Н., Михайличенко О.В., Сивачев А.В. Применение матриц соседства пикселей для улучшения точности стеганоанализа неподвижных цифровых изображений с однородным фоном //

Кибернетика и программирование. — 2018. - № 1. - С.64-72. DOI: 10.25136/2306-4196.2018.1.24919. URL: http://e-notabene.ru/kp/article_24919.html

4. Прохожев Н.Н., Сивачев А.В., Михайличенко О.В., Башмаков Д.А. Повышение точности стеганоанализа в области ДВП путем использования взаимосвязи между областями двумерного и одномерного разложений // Кибернетика и программирование. – 2017. - №2. – С. 78 – 87. DOI: 10.7256/2306-4196.2017.2.22308. URL: https://e-notabene.ru/kp/article_22412.html

5. Сивачев А.В., Прохожев Н.Н., Михайличенко О.В., Башмаков Д.А. Эффективность стеганоанализа на основе методов машинного обучения // Научно-технический вестник информационных технологий, механики и оптики. - 2017. - Т. 17. - № 3(109). - С. 457-466. DOI: DOI: 10.17586/2226-1494-2017-17-3-457-466. URL: https://ntv.ifmo.ru/ru/article/16738/effektivnost_steganoanaliza_na_osnove_metodov_mashinnogo_obucheniya_.htm

Публикации из перечня Scopus:

6. Prokhozhev N., Mikhailichenko O., Sivachev A., Bashmakov D., Korobeynikov A.G. Passive Steganalysis Evaluation: Reliabilities of Modern Quantitative Steganalysis Algorithms // Advances in Intelligent Systems and Computing. - 2016. - Vol. 451. - Pp. 89-94. URL: https://link.springer.com/chapter/10.1007/978-3-319-33816-3_9

Прочие публикации:

7. Прохожев Н.Н., Михайличенко О.В., Башмаков Д.А., Сивачев А.В., Коробейников А.Г. Исследование эффективности применения статистических алгоритмов количественного стеганоанализа в задаче детектирования скрытых каналов передачи информации // Программные системы и вычислительные методы. - 2015. - № 3. - С. 281-292. DOI: 10.7256/2305-6061.2015.3.17233, URL: http://www.nbpublish.com/library_read_article.php?id=-35795

8. Башмаков Д.А., Сивачев А.В. Влияние параметров маски на практическую точность RS-анализа // Сборник трудов IV Всероссийского конгресса молодых ученых (Санкт-Петербург, 7-10 апреля 2015 г.). - 2015. - С. 49-53.

9. Сивачев А.В., Башмаков Д.А. Влияние предварительной обработки изображения - контейнера фильтрами на точность статистического стеганоанализа // Сборник трудов IV Всероссийского конгресса молодых ученых (Санкт-Петербург, 7-10 апреля 2015 г.). - 2015. - С. 361-365.

Подписано в печать "__" _____ 2018 г.
Формат 60x84 1/16. Бумага офсетная. Печать офсетная.
Усл.печ.л. 1,0. Тираж 100 экз.
Заказ №