ОЦЕНКА ПАРАМЕТРОВ НЕОБНАРУЖИВАЕМОСТИ РАЗРАБОТАННОГО ПОДХОДА К МАРКИРОВАНИЮ ТЕКСТОВЫХ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Козачок А. В.[1], Копылов С. А.[2], Бочков М. В.[3]

Цель статьи: *оценка основных параметров необнаруживаемости подхода к маркированию текстовых электронных документов и определение наличия зависимости стойкости встроенных данных к стеганографическому анализу от используемых параметров встраивания.*

Метод: *экспериментальная оценка и сравнительный анализ результатов исследования с существующими критериями посредством применения методов стеганографического анализа и средств теории распознавания образов.*

Полученный результат: *получены количественные значения параметров необнаруживаемости разработанного подхода, проведен сравнительный анализ подходов к стегоанализу изображений, обоснован выбор потенциально наилучших методов стегоанализа, которые могут быть применены к разработанному подходу, определены границы стойкости подхода маркирования к стегоанализу, оценены качественные показатели разработанного подхода, установлены граничные значения надежности разработанного подхода, определена зависимость невидимости и необнаруживаемости встроенных данных от используемых параметров встраивания.*

Ключевые слова: *маркирование текстовой информации, цифровой водяной знак, стегоанализ, классификация данных, ROC-анализ.*

DOI:10.21681/2311-3456-2020-01-62-73

Введение

Проблема обеспечения защищенности текстовой информации от утечки является одним из актуальных направлений исследований. По данным аналитического центра InfoWatch доля утечек информации, представленной в напечатанном виде, в 2018 г. выросла с 8,2% до 11%. При этом в распределении утечек доля инцидентов, связанных с утечкой бумажных текстовых документов, по сравнению с 2017 г. увеличилась с 13,5% до 17% [1]. Высокий процент утечек напечатанных текстовых документов обусловлен тем, что существенная часть документооборота совершается в неэлектронной форме. Кроме того, во многих компаниях и государственных организациях далеко не всегда соблюдаются правила обращения с бумажными носителями. Помимо утечки физических копий текстовых документов за пределы контролируемого периметра, отмечается утечка электронных версий напечатанных на бумаге текстов. Данная утечка реализуется посредством сканирования напечатанного на бумаге документа с последующей отправкой полученного изображения. При этом в современных средствах защиты текстовой информации каналу утечки отсканированных изображений, содержащих исходный текстовый документ, не уделяется должного внимания. В рамках решения данной задачи разработан подход к маркированию текстовой информации, основанный на внедрении робастного водяного знака (РВЗ) в текстовые документы в процессе вывода их на печать. Внедрение РВЗ позволяет обеспечить стойкость внедряемых данных к возможным преобразованиям и искажениям, которым может быть подвержен напечатанный на бумаге документ. Кроме того, наличие маркера позволяет осуществлять идентификацию владельца данных, а также отслеживать источник либо направление утечки.

Особенности разработанного подхода к маркированию текстовых документов

В разработанном подходе к маркированию [2–5] использован алгоритм стеганографического внедрения информации, основанный на изменении величины межстрочного интервала в процессе внедрения РВЗ в текстовые данные. Использование данного алгоритма внедрения обусловлено требованием по формированию перцептивно-невидимого (для человеческого глаза) РВЗ, который в последующем может быть извлечен из изображений, содержащих исходный текст, полученных посредством применения операции сканирования. Для извлечения встроенного РВЗ из изображений в разработан-

1 Козачок Александр Васильевич, кандидат технических наук, сотрудник Академия ФСО России, г. Орёл, Россия. E-mail: a.kozachok@academ.msk.rsnet.ru

2 Копылов Сергей Александрович, сотрудник Академия ФСО России, г. Орёл, Россия. E-mail: gremlin.kop@mail.ru

3 Бочков Максим Вадимович, доктор технических наук, профессор, ЧОУ ДПО «Центр предпринимательских рисков», г. Санкт-Петербург, Россия. E-mail: mvboch@yandex.ru

ном подходе к маркированию используется нормальное преобразование Радона, позволяющее осуществлять извлечение величин межстрочных интервалов, а также модель разделения смеси нормальных распределений в процессе бинаризации полученных значений.

Результаты проведенной авторами экспериментальной оценки разработанного подхода позволяют оценить основные характеристики невидимых РВЗ: емкость встраивания, робастность, извлекаемость, невидимость (перцептивная прозрачность) и необнаруживаемость (сложность обнаружения) [6–9].

В ходе экспериментальной оценки емкости встраивания установлено, что предельно достижимая емкость встраивания ограничена величиной в 60 бит и зависит от используемого кегля шрифта, величины межстрочного интервала и количества строк текста в исходном документе. При этом для встраивания маркера, содержащего информацию о конфиденциальности текстового документа, достаточно пяти строк текста.

Робастность водяного знака характеризуется способностью встроенных данных сохранять свойство инвариантности после осуществления различных преобразований над изображением. Разработанный РВЗ позволяет обеспечить стойкость встроенных данных к таким преобразованиям как преобразование формата электронного текстового документа в изображение посредством применения операции «печать-сканирование», поворот изображения на любой угол, фильтрация, масштабирование, сжатие изображения (в том числе с потерями) и преобразование в любой формат растрового изображения.

В процессе извлечения встроенного РВЗ из изображений необходимо оценить возможность правильного извлечения и точность извлечения данных. В ходе экспериментальной оценки возможности извлечения невидимого РВЗ из изображений осуществлен анализ существующих систем РВЗ. Согласно классификации, представленной в работе Петитколаса [10], системы РВЗ могут быть разделены на следующие категории (рис. 1):

Рис. 1. Классификация систем РВЗ

- закрытые системы РВЗ (I и II типа). Закрытые системы I типа позволяют осуществлять извлечение встроенных данных при наличии исходного (неизмененного) изображения. Закрытые системы II типа, в свою очередь, способны только обнаружить факт наличия/отсутствия встроенного водяного знака. Таким системам помимо исходного изображения требуется информация о встраиваемом водяном знаке.
- полузакрытые системы РВЗ. В отличие от закрытых систем II типа способны определять факт наличия/отсутствия встроенного водяного знака при отсутствии исходного изображения;
- открытые (слепые) системы РВЗ. Данные системы позволяют осуществлять извлечение встроенных данных только из подписанного изображения, не накладывая дополнительных требований.

Разработанная система РВЗ относится к открытым системам ввиду возможности извлечения встроенных данных только из подписанных изображений и отсутствия требований по наличию исходного (неподписанного) изображения в процессе извлечения. В то же время существующие системы РВЗ текстовых данных, в том числе и базовый аналог, описанный в работах Брассила [11-13], относятся к закрытым системам I или II типа, которым необходимо наличие исходного изображения или встраиваемой информации в процессе извлечения. Указанная особенность позволяет обеспечить высокие значения точности извлечения закрытыми системами РВЗ по сравнению с открытыми. При этом в процессе извлечения зачастую отсутствует доступ к исходной информации, что, в свою очередь, характеризует закрытые системы РВЗ как системы, не позволяющие осуществить извлечение встроенных данных.

Точность извлечения встроенных данных может быть описана посредством методов оценки классификации используемых в задачах теории распознавания образов [14–16]. Для количественной оценки точности извлечения данных рассмотрены следующие метрики, характеризующие точность классификатора: точность (выражение 1) и F-мера (выражение 2):

$$Точность = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$F - мера = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \quad (2)$$

где TP – истинно-положительный, TN – истинно-отрицательный, FP – ложно-положительный и FN – ложно-отрицательный результат.

В процессе экспериментальной оценки точности извлечения было извлечено более 10000 бит (более 250 страниц текстовой информации), что позволяет утверждать о том, что доверительный интервал равен 0,95 при точности 0,01. Полученные результаты оценки точности извлечения данных на примере текстового документа с размером кегля 14 пт. представлены в таблице 1.

Анализ полученных результатов позволяет сделать вывод о возможности гарантированного извлечения (с точностью более 95%) встроенных данных при внедрении РВЗ посредством изменения величины межстрочного интервала на 0,04 и выше от исходного значения.

Результаты, полученные в ходе анализа зависимости точности извлечения от изменения величины разрешения изображения (DPI), позволяют сделать вывод о том, что точность извлечения встроенных данных из отсканированных изображений, содержащих встроенный РВЗ, с показателем DPI в 200 точек на дюйм и выше превышает значение в 93%. График зависимости точности извлечения от величины DPI для текста с различными параметрами встраивания представлен на рисунке 2.

DOI:10.21681/2311-3456-2020-01-62-73

Таблица 1

Оценка результата точности извлечения данных

Величина изменения межстрочного интервала	Показатель истинно-положительных значений (TPR)	Показатель истинно-отрицательных значений (TNR)	Точность (Accuracy)	F-мера (F-measure)
0,01	0,50	0,54	0,68	0,68
0,02	0,57	0,65	0,60	0,61
0,03	0,66	0,80	0,73	0,73
0,04	0,93	0,97	0,95	0,95
0,05	0,97	0,95	0,96	0,95
0,06	0,97	0,95	0,96	0,96
0,07	0,97	1	0,98	0,98
0,08	1	0,97	0,98	0,98
0,09	1	0,97	0,98	0,98
0,10	0,97	1	0,98	0,98

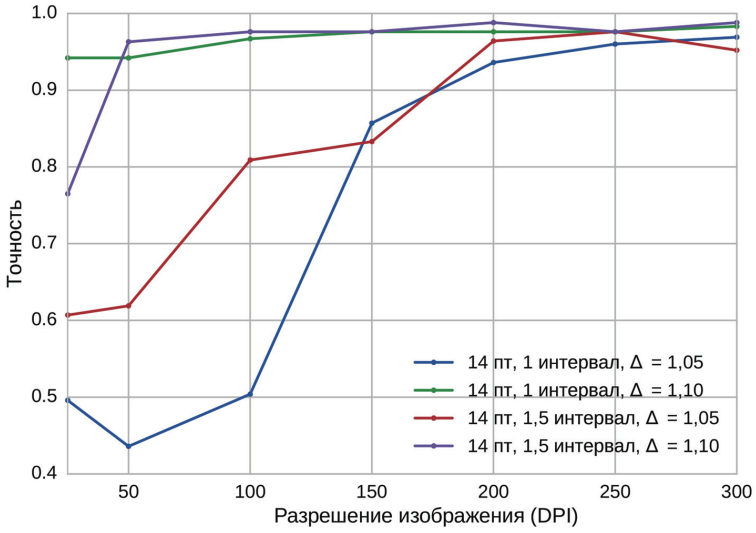

Рис. 2. Зависимость точности извлечения от разрешения изображения

Оценка перцептивной невидимости встроенных данных проведена посредством проведения наблюдателем визуального анализа текстов как содержащих, так и не содержащих встроенный невидимый водяной знак. Исходя из полученных результатов определен порог (граница) невидимости водяного знака, который равен величине изменения межстрочного интервала на значение $\pm 0,15$ от исходного. При этом использование значений в диапазоне $0,07 \ldots 0,15$ приводит к переносу последней строки на новый лист (в случае полного заполнения страницы текстом).

Для устранения указанного недостатка необходимо использовать РВЗ меньшего размера, чем величина предельно достижимой емкости встраивания. В случае выполнения данного требования перцептивная невидимость разработанного невидимого РВЗ может быть ограничена значением изменения величины межстрочного интервала в $\pm 0,10$.

Необнаруживаемость водяного знака (разработанного подхода к маркированию) может быть описана несколькими критериями [17–19]:

- стойкость невидимого ЦВЗ (подхода маркирования) к стеганографическому анализу;
- надежность извлечения встроенной информации, которая характеризуется показателем ложноположительных срабатываний (величиной ошибочного извлечения бит информации) (FPR);
- качество изображения после внедрения информации;
- скорость вложения информации (R_w) – отношение количества бит встроенной информации к предельно достижимой емкости встраивания данного контейнера;
- секретность ЦВЗ (подхода к маркированию), характеризующаяся видом и параметрами ROC-кривой.

Необнаруживаемость разработанного подхода может быть оценена как по одному конкретному показателю, так и по всем представленным. Использование только одного критерия нерационально с позиций комплексного анализа разработанного подхода. Для

оценки первого критерия – стойкости разработанного невидимого РВЗ к стеганографическому анализу – необходимо провести анализ существующих подходов в данной области.

Сравнительный анализ научных исследований в области стеганографического анализа изображений

Стеганографический анализ (стегоанализ) – наука об обнаружении присутствия скрываемой информации в анализируемых файлах-контейнерах (данных мультимедиа) [20,21]. Помимо обнаружения скрываемых данных дополнительной целью стегоанализа является извлечение встроенной информации. Ввиду отсутствия информации о характеристиках файла-контейнера, а также используемого способа внедрения данных стегоанализ является одной из наиболее сложных областей исследования. При этом подходы к стегоанализу, в большинстве случаев, зависят от используемого стеганографического алгоритма внедрения данных. По аналогии со стеганографическими методами подходы стегоанализа можно разделить на: стегоанализ изображений, а также стегоанализ аудио- и видеоданных (данных мультимедиа).

Классификация методов стеганографического анализа изображений представлена на рисунке 3 [20,22–24]. На основании данной классификации все методы стегоанализа изображений могут быть разделены на три категории: сигнатурные, визуальные и статистические. Второй ступенью классификации является разделение всех методов стегоанализа на методы анализа только подписанного изображения и методы анализа одновременно исходного и подписанного изображения. Однако в большинстве случаев исходное изображения может быть недоступно для стегоанализа.

Сигнатурный метод стегоанализа основан на поиске в анализируемом изображении сигнатуры (шаблона), характеризующего конкретный алгоритм или программное средство, осуществляющее стеганографическое внедрение информации. Методы сигнатурного стегоанализа позволяют однозначно определить факт встраивания и идентифицировать используемый стеганографический алгоритм (программное средство). При этом количество программных средств внедрения информации, имеющих собственные сигнатуры, исчисляется несколькими десятками, что не позволяет использовать данный подход для обнаружения новых (неизвестных) алгоритмов (программных средств), осуществляющих стеганографическое внедрение информации.

В процессе стеганографического внедрения информации осуществляется изменение не только статистических свойств изображения, но и качественных характеристик, которые влияют на представление изображения. Для поиска визуальных отклонений используются методы визуального стегоанализа. Методы визуального стегоанализа основаны на перцептивном (относящемся к зрительной системе человека) восприятии анализируемого изображения. Визуальный стегоанализ может быть реализован посредством визуального сравнения исходного и подписанного изображения, анализа только подписанного изображения, а также посредством анализа результатов компьютерной обработки анализируемого изображения.

Сравнение анализируемого изображения с оригиналом позволяет обнаружить наличие изменений в изображении, а, следовательно, установить факт присутствия встроенной информации. Основная сложность реализации данного подхода заключается в том, что не всегда удается получить доступ к оригинальной версии изображения. Визуальный анализ только подписанного изображения характеризуются меньшей точностью обнаружения, чем метод сравнения с оригиналом. Указанные методы относятся к вероятностным и характеризуются большим числом ложных срабатываний ввиду наличия легитимных искажений, вызванных операциями сжатия, фильтрации и преобразования формата изображения, которые могут быть приняты за результат стеганографического внедрения. В тоже время применение данных подходов стегоанализа накладывает дополнительное требование по перцептивной невидимости встроенных данных на используемые в процессе встраивания стеганографические подходы.

Наиболее эффективным среди визуальных методов стегоанализа является подход, основанный на компьютерной обработке анализируемого изображения. Результаты обработки позволяют не только обнаруживать аномалии и отклонения в качественных характеристиках изображения, но и определять позиции бит информации, в которых осуществлены изменения. Данная особенность, помимо установления факта наличия/от-

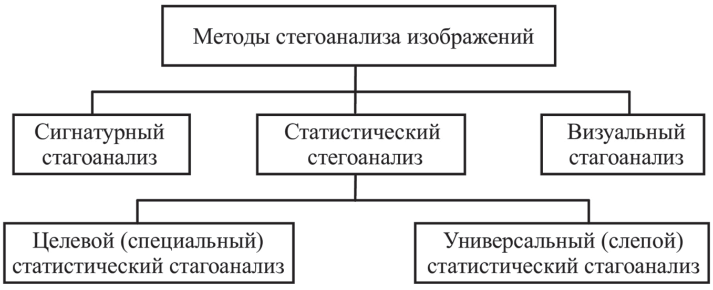

Рис. 3. Классификация методов стегоанализа изображений

DOI:10.21681/2311-3456-2020-01-62-73

сутствия встроенных данных, позволяет осуществлять их извлечение. Основным недостатком, присущим данному подходу, является невозможность обнаружения искажений в изображениях, сжатых посредством применения алгоритмов, основанных на дискретном вейвлет и дискретном косинусном преобразованиях. При этом изображения данных форматов являются наиболее распространенным типом данных, используемых в цифровой обработке изображений.

Статистические методы стегоанализа основаны на оценке статистики распределения информации, а также статистических свойствах изображения. Данные методы можно разделить на две группы: анализирующие исходное и подписанное изображение и осуществляющие анализ только подписанного изображения. Первая группа методов осуществляет сравнение статистики исследуемого изображения со статистикой изображения-шаблона, который характеризует исходное изображение. В случае обнаружения отклонений делается вывод о наличии встроенной информации. Основным недостатком данного подхода является невозможность обнаружения (построения) для исследуемого изображения изображения-шаблона.

Вторая группа методов статистического стегоанализа изображений делится на две категории:

- целевой (специальный) стегоанализ – направлен на обнаружение (извлечение) информации из изображений, встраивание которой осуществлено заранее известным стеганографическим методом. Достоинством методов данной группы является возможность извлечения встроенной информации из подписанного изображения. В то же время методы данной группы малоэффективны против новых или неизвестных алгоритмов стеганографического внедрения информации.
- универсальный (слепой) стегоанализ – осуществляет обнаружение встроенной информации при отсутствии информации об используемом методе стеганографического внедрения информации. Обнаружение встроенной информации основано на поиске изменений, а также аномалии в структуре, статистических характеристиках и уникальных признаках формата изображения. К достоинствам относится возможность обнаружения встроенных данных, внедренных посредством как известных, так и новых алгоритмов стеганографического встраивания.

В разработанном подходе к маркированию текстовых данных внедрение РВЗ осуществляется за счет изменения форматирования пространства текстового документа и не вносит дополнительной статистики в напечатанный документ, а, следовательно, и в отсканированное изображение. Кроме того, встраивание РВЗ осуществляет в процессе печати текстового документа, что делает невозможным сравнительный стегоанализ подписанного и исходного изображения ввиду отсутствия последнего.

Указанные особенности накладывают ограничение на использование сигнатурных методов стегоанализа, а также большинства слепых статистических методов стегоанализа. При этом разработанный подход может быть уязвим для методов визуального и целевого стегоанализа. Так визуальный анализ подписанного документа может выявить отклонения некоторых строк текста в случае использования больших величин изменения межстрочных интервалов. Целевой стегоанализ, направленный на обнаружение статистических отклонений в значениях величин межстрочных интервалов, также может позволить обнаружить наличие встроенных данных.

Для оценки необнаруживаемости (стойкости) разработанного подхода к маркированию текстовых данных необходимо произвести количественно оценку рассмотренных критериев необнаруживаемости.

Экспериментальная оценка необнаруживаемости разработанного подхода маркирования текстовых данных

Для оценки стойкости разработанного подхода к стеганографическому анализу определены потенциально наилучшие методы стегоанализа: визуальный и целевой статистический, основанный на применении нормального преобразования Радона в процессе обнаружения (извлечения) данных.

Стойкость разработанного подхода к потенциально наилучшим методам стегоанализа

В ходе предварительной оценки стойкости к визуальному стегоанализу разработанного подхода маркирования определен порог перцептивной невидимости встроенных данных и установлены ограничения по минимальной емкости контейнера, необходимой для встраивания разработанного РВЗ.

Стегоанализ проводился экспертами путем зрительного анализа подписанных текстовых документов, представленных как в напечатанном, так и в электронном виде (в формате отсканированных изображений). В ходе экспертной оценки подписанных изображений проведено три группы исследований.

Первая группа исследований направлена на определение факта наличия встроенной информации, содержащейся в анализируемых подписанных изображениях. При этом стегоаналитик не обладает информацией о содержании в анализируемых изображениях внедренного РВЗ. В качестве анализируемых изображений выступали изображения в которых осуществлено внедрение РВЗ посредством изменения величины межстрочного интервала на значения $\pm 0{,}10$ от исходного с шагом 0,01. В результате анализа подписанных изображений не удалось определить факт наличия встроенных данных.

При проведении второй группы исследований стегоаналитик обладал информацией о факте наличия в анализируемых изображениях встроенных данных. При этом алгоритм стеганографического внедрения информации остается неизвестным. В результате проведенного визуального анализа обнаружены аномалии в структуре текста подписанных изображений, характеризующихся использованием величины встраивания в границах изменения межстрочного интервала на $0{,}07 \ldots 0{,}10$ от исходного. Данный интервал встраи-

вания характеризует разработанный подход маркирования как уязвимый к визуальному стегоанализу.

В ходе проведения третьей группы экспериментов стегоаналитик обладал информацией об используемом алгоритме стеганографического внедрения информации. В результате визуального анализа установлены значения необнаруживаемости встроенных данных:

- текст с межстрочным интервалом 1 – изменение величины межстрочного интервала на $\pm(0,01\ldots0,05)$ от исходного;
- текст с межстрочным интервалом 1,25 – $\pm(0,01\ldots0,06)$;
- текст с межстрочным интервалом 1,5 – $\pm(0,01\ldots0,07)$.

Результаты визуального анализа позволяют сделать вывод о стойкости разработанного подхода маркирования текстовых документов к данному методу стегоанализа в случае использования в процессе встраивания величины изменения межстрочного интервала в диапазоне значений $\pm(0,01\ldots0,05)$. Принимая во внимание результаты извлекаемости и точности встроенных данных, полученный диапазон встраивания может быть сужен до значений в диапазоне $\pm(0,04\ldots0,05)$.

Помимо стойкости к визуальному стегоанализу проведена оценка стойкости разработанного подхода к целевому статистическому стегоанализу. В качестве подхода целевого статистического стегоанализа использован подход к извлечению встроенных данных, описанный авторами. Извлечение встроенных данных в предложенном подходе основано на применении к изображению нормального преобразования Радона, позволяющего установить значения величин межстрочных интервалов.

В ходе экспериментальной оценки осуществлено извлечение информации как из чистого (неподписанного) изображения, так и из изображения, содержащего встроенные данные. Результаты извлечения величин межстрочных интервалов из неподписанного изображения и изображений с различными параметрами встраивания представлены в таблице 2.

Результаты извлечения величин межстрочных интервалов изображения, не содержащего встроенных данных, помимо ошибок извлечения, характеризуются отклонением в распределении величин межстрочных интервалов на величину 1 пиксель. Обозначим данную величину как допустимую погрешность извлечения или «шум» изображения. Наличие данной погрешности позволяет осуществлять внедрение информации, которое не может быть обнаружено методом целевого статистического стегоанализа. Для решения данной задачи в процессе встраивания РВЗ необходимо выбирать параметр встраивания Δ (величину изменения межстрочного интервала), который при извлечении величин межстрочных интервалов будет находится в границах «шума» изображения.

Анализ результатов извлечения величин межстрочных интервалов показал, что при использовании параметра встраивания Δ равного изменению величины межстрочного интервала на 0,04 или 0,05 от исходного, результат извлечения встроенных данных находится в границах допустимой погрешности. Использование $\Delta \geq 0,06$ характеризуется величиной статистического отклонения в 2 пикселя, что не соответствует «шуму» изображения и может быть обнаружено в процессе целевого статистического стегоанализа.

Дальнейшая оценка стойкости разработанного подхода маркирования к целевому статистическому стегоанализу осуществлялась посредством сравнительного анализа гистограмм распределения значений межстрочных интервалов изображения, не содержащего встроенных данных, с изображением, содержащим встроенный РВЗ (параметр встраивания Δ равный 0,04 (рис. 4) и 0,05 (рис. 5)).

Наиболее близкой (идентичной) к гистограмме распределения величин межстрочных интервалов изображения, не содержащего встроенные данные, является гистограмма распределения изображения с параметром встраивания $\Delta = 0,05$. При этом отдельные всплески (скачки значений) в полученных значениях могут быть объяснены. Так, значения межстрочных интервалов, не превышающие величину $0,6 \cdot mode$ (где $mode$ – статистическая мода в полученном массиве

Таблица 2

Результат извлечения величин межстрочных интервалов

Параметр встраивания	Извлеченные межстрочные интервалы
Без встраивания	[53, 52, 53, 53, 52, 53, 52, 53, 52, 53, 53, 6, 5, 53, 53, 52, 53, 52, 1, 54, 53, 52, 53, 52, 53, 3, 53, 53, 53, 52, 53, 14, 53]
$\Delta = 0,04$	[54, 53, 54, 53, 54, 52, 54, 53, 54, 53, 8, 45, 52, 54, 53, 54, 53, 54, 52, 55, 8, 43, 54, 53, 54, 52, 8, 46, 52, 55]
$\Delta = 0,05$	[54, 53, 54, 53, 54, 53, 55, 52, 55, 52, 8, 46, 52, 55, 52, 55, 52, 55, 52, 55, 8, 44, 54, 53, 54, 53, 8, 45, 53, 55]
$\Delta = 0,06$	[55, 53, 55, 53, 55, 53, 55, 2, 2, 53, 55, 53, 55, 52, 55, 8, 44, 55, 52, 55, 53, 54, 126, 52, 55, 54, 55, 52, 55]

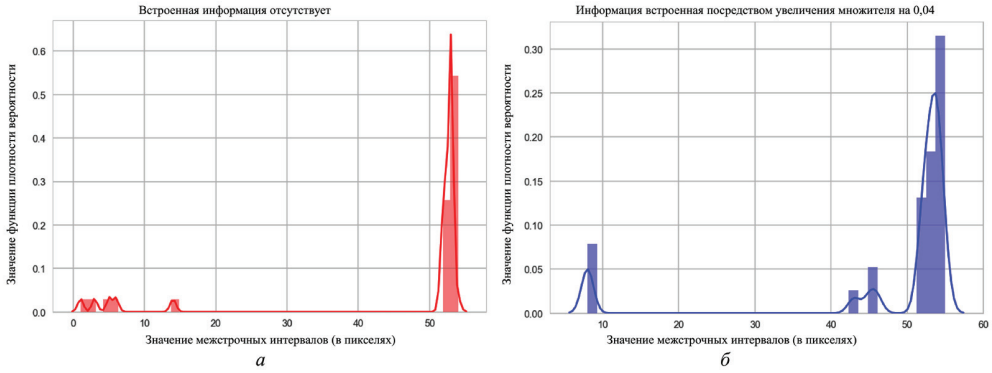

Рис. 4. Гистограмма распределений значений межстрочных интервалов для изображения: а) не содержащего встроенный РВЗ; б) содержащего встроенный РВЗ ($\Delta = 0,04$)

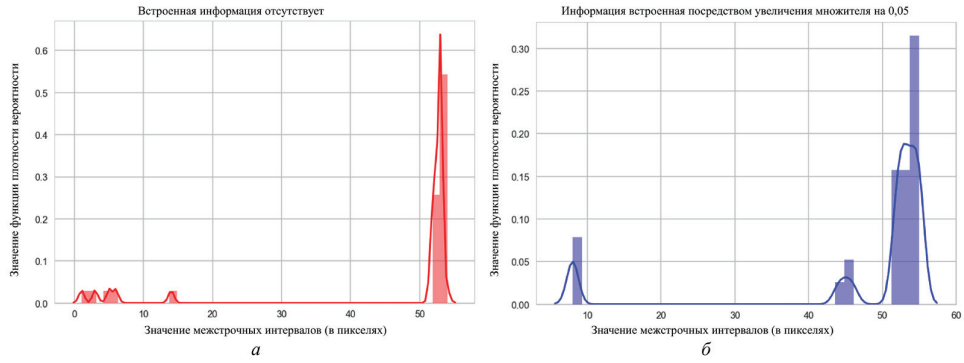

Рис. 5. Гистограмма распределений значений межстрочных интервалов для изображения: а) не содержащего встроенный РВЗ; б) содержащего встроенный РВЗ ($\Delta = 0,05$)

значений), соответствуют ошибкам первого рода процесса извлечения. Второй вид ошибок – значения из диапазона $\{0,6 \cdot mode \ldots mode\} \cup \{mode \ldots 1,6 \cdot mode\}$, характеризуют ошибки, полученные в процессе обработки синограммы, которые могут быть исправлены посредством применения процедуры корректирования.

Таким образом, полученные результаты стойкости разработанного подхода маркирования текстовых данных к потенциально наилучшим методам стегоанализа позволяют сделать вывод о устойчивости разработанного РВЗ в случае выбора параметра встраивания, находящего в диапазоне изменения величины межстрочного интервала относительно исходного на значение, не превышающее 0,05.

Надежность извлечения встроенной информации

Надежность извлечения встроенной информации является величиной противоположной по значению показателю ложно-положительных срабатываний (величине ошибочного извлечения бит информации) и определяется выражением:

$$Надежность = 1 - FPR = 1 - \frac{FP}{FP + TN}. \quad (3)$$

Зависимость FPR от используемых параметров встраивания и разрешения отсканированного изображения для текста набранного гарнитурой Times New Roman с кеглем шрифта 14 пт представлена в таблице 3.

Полученные значения FPR позволяют сделать вывод о том, что надежность извлечения встроенной информации превышает показатель в 0,92 при использовании параметра встраивания (величины изменения межстрочного интервала) $\Delta \geq 0,05$ и разрешении отсканированного изображения в 150 точек на дюйм и выше. При этом использование значения $\Delta \geq 0,06$ в процессе внедрения РВЗ в текстовые данные недопустимо ввиду отсутствия стойкости к потенциально наилучшим методам стегоанализа.

Качество изображения после внедрения информации

Оценка качества изображения после внедрения в него информации может быть осуществлена посредством сравнительного анализа изображения, содержащего встроенный РВЗ, и его прототипа – не содержащего встроенные данные. Для практической реализации сравнительного анализа подготовлены два текстовых документа. Первый текстовый документ выводится на

Таблица 2.

Зависимость величины ошибочного извлечения бит информации от используемых параметров встраивания

Разрешение изображения	Межстрочный интервал – 1		Межстрочный интервал – 1,25		Межстрочный интервал – 1,5	
	$\Delta = 0,05$	$\Delta = 0,10$	$\Delta = 0,05$	$\Delta = 0,10$	$\Delta = 0,05$	$\Delta = 0,10$
25	0,55	0,09	0,52	0,26	0,19	0,21
50	0,26	0,07	0,40	0,09	0,49	0,06
100	0,57	0,03	0,35	0,11	0,27	0,02
150	0,08	0,02	0,05	0,02	0,05	0
200	0,08	0,02	0,06	0,04	0,07	0
250	0,08	0,02	0,06	0,02	0,05	0
300	0,06	0,02	0,04	0,02	0,05	0

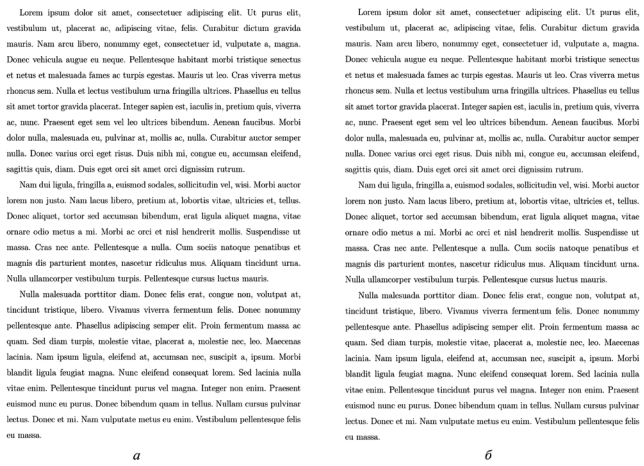

Рис. 6. Отсканированное изображение: а) не содержащее встроенные данные; б) содержащее встроенный РВЗ ($\Delta = 0,05$)

печать в неизмененном виде (без внедрения РВЗ), во второй производится внедрение РВЗ с заданным параметром встраивания Δ. В результате сканирования полученных текстовых документов формируются соответствующие изображения. На рисунке 6 представлен пример «чистого» изображения (не содержащего встроенных данный) и «подписанного» изображения с параметром встраивания $\Delta = 0,05$. Разрешение полученных изображений составляет 300 точек на дюйм. Параметры исходного текстового документа: гарнитура – Times New Roman, кегль – 14 пт., величина межстрочного интервала – 1,5.

Визуальный анализ как напечатанных на бумаге текстов, так и соответствующих им изображений показал, что «подписанный» документ не отличается от «чистого». Данная особенность обусловлена используемым алгоритмом внедрения информации, который не оказывает влияние на качественные характеристики формируемого документа. Наличие указанной особенности позволяет осуществлять внедрение РВЗ в текстовые документы без ухудшения качества их восприятия.

Скорость вложения информации

Скорость вложения информации (R_W) определяется как отношение количества бит встроенной информации к предельно достижимой емкости встраивания. В ходе оценки емкости встраивания разработанного подхода установлено, что предельно достижимая емкость встраивания варьируется в пределах от 28 до 60 бит на страницу текста в зависимости от величины кегля шрифта и размера межстрочного интервала. В случае $R_W = 1$ разработанная система РВЗ является максимально эффективной, т.е. позволяет осуществлять полное заполнение контейнера встраиваемой информацией. Однако в случае внедрения РВЗ, сопоставимого по длине с предельно достижимой емкостью встраивания, осуществляется перенос строки текста на новую строку при полном заполнении станицы текста. Указанная особенность характерна при использовании параметра встраивания Δ в диапазоне значений $0,07 \ldots 0,10$. Ввиду указанной особенности целесообразно ограничить скорость вложения информации величиной $R_W \le 0,08$ в случае полного заполнения текстом страницы документа.

Секретность разработанного подхода

Секретность разработанного подхода (водяного знака) может быть количественна оценена через следующие показатели:

- TPR (True Positive Rate) – чувствительность или показатель истинно-положительных значений результата классификации извлеченных данных:

$$TPR = \frac{TP}{TP + FN}; \quad (4)$$

- FPR (False Positive Rate) – выпадение или показатель ложно-положительных значений результата классификации извлеченных данных:

$$FPR = \frac{FP}{FP + TN}. \quad (5)$$

Рассмотренные показатели позволяют количественно оценить число верно классифицированных объектов (TPR) и число ошибочно классифицированных объектов (FPR). Зависимость данных показателей может быть представлена посредством ROC-кривой (Receiver Operating Characteristic) [25,26]. ROC-кривая представляет собой двухмерный график, у которого по оси абсцисс расположены значения FPR, а по оси ординат – TPR. При этом стоит отметить, что результат классификации, проведенной бинарным классификатором, характеризуется парой значений (FPR, TPR). Каждая пара значений соответствует точке, задающей ROC-кривую [14]. Пример возможных ROC-кривых представлен на рисунке 7.

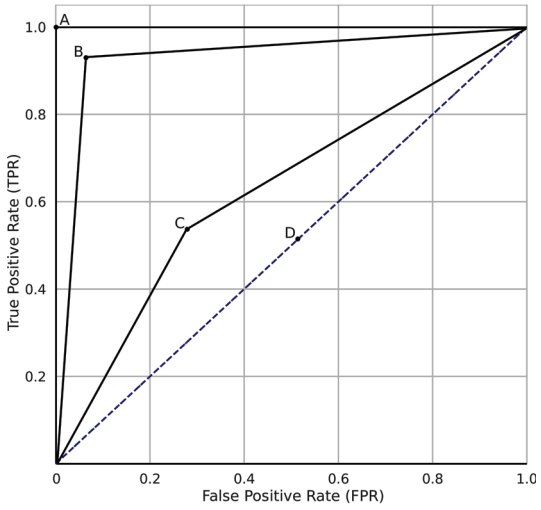

Рис. 7. Примеры задания ROC-кривой

ROC-кривая позволяет охарактеризовать обнаружитель (детектор встроенных данных). Так на рисунке 7 кривая A соответствует идеальному обнаружителю ($FPR = 0, TPR = 1$), кривая B – хорошему обнаружителю, кривая C – посредственному обнаружителю, а кривая D – обнаружителю со случайным угадыванием ($FPR = TPR = 0,5$). При этом критерий, задающий секретность системы ЦВЗ (стегосистемы), является величиной обратной по отношению к характеристике обнаружителя (детектора). В таком случае ROC-кривая D будет описывать наилучшую секретную систему ЦВЗ (стегосистему), а ROC-кривая A – несекретную систему ЦВЗ (стегосистему).

В ходе экспериментальный оценки секретности разработанного подхода осуществлена оценка секретности разработанного невидимого РВЗ посредством формирования ROC-кривых, построенных на основе полученных значений TPR и FPR. В процессе оценки осуществлен анализ зависимости секретности разработанного невидимого РВЗ от параметра встраивания (величины изменения межстрочного интервала). На рисунке 8 представлен результат построения ROC-кривых в зависимости от изменения параметра встраивания ($\Delta = 0,01 \ldots 0,08$ с шагом 0,01) для текстов, набранных гарнитурой Times New Roman, кеглем шрифта 14 пт. и величиной межстрочного интервала 1.

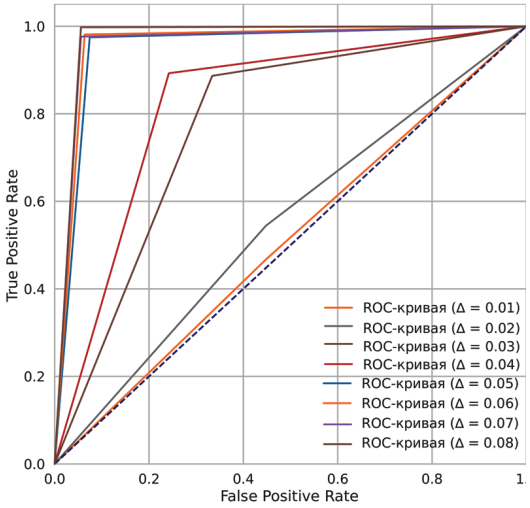

Рис. 8. ROC-кривые, характеризующие разработанный подход

Анализ полученных данных позволяет сделать вывод о зависимости секретности разработанной системы РВЗ от изменения величины межстрочного интервала Δ в процессе внедрения данных. В том случае, если $\Delta = 0,01$, то разработанная система РВЗ (разработанный подход) характеризуется как секретная система ЦВЗ, соответствующая обнаружителю со случайным угадыванием. Однако, стоит заметить, что при использовании $\Delta = 0,01$ в процессе извлечения встроенных данных возникает большое число ошибок, что может быть компенсировано введением избыточности в водяной знак. В свою очередь наличие избыточности снижает полезную нагрузку, содержащуюся в водяном знаке.

Выбор параметра встраивания $\Delta = 0,03 (0,04)$ позволяет повысить точность извлечения детектора и отнести его к хорошему обнаружителю, но, как и в случае с $\Delta = 0,01 (0,02)$, требуется внесение избыточности или применения помехоустойчивых кодов в процессе формирования РВЗ, что также приведет к снижению полезной нагрузки. В свою очередь, использование $\Delta = 0,05 \ldots 0,08$ позволяет сделать вывод о несекретной системе ЦВЗ близкой к системе с идеальным обнаружителем.

Исходя из полученных результатов, можно сделать вывод о том, что выбор параметра встраивания определяет секретность разработанного подхода. Указанная особенность позволяет использовать разработанный

подход к маркированию для решения различных задач. В случае, если требуется обеспечить высокую степень необнаруживаемости и невидимости встраиваемого маркера, необходимо выбирать параметр встраивания $\Delta \le 0,02$, а также осуществлять введение избыточности в маркер посредством использования помехоустойчивых кодов. Если же требуется обеспечить гарантированное извлечение каждого бита встроенной информации, необходимо использовать $\Delta = 0,05(0,06)$. Оптимальным с позиции соотношения объема полезной нагрузки к точности извлечения является использование параметра $\Delta = 0,03(0,04)$, что соответствует пограничному положению между секретной и несекретной системами ЦВЗ.

Выводы

Разработанный подход к маркированию текстовых документов, помимо внедрения перцептивно невидимого маркера в текстовые документы, позволяет обеспечить стойкость встроенной информации к преобразованию напечатанного на бумаге текстового документа в изображение. Проведенный анализ стойкости разработанного подхода маркирования позволил определить границы необнаруживаемости встроенных данных. Так, разработанный подход характеризуется как стойкий к стеганографическому анализу, проводимому потенциально наилучшими методами при внедрении информации на величину изменения межстрочного интервала $\Delta = 0,04(0,05)$. При этом надежность извлечения встроенной информации из отсканированных изображений, содержащих встроенные данные, превышает показатель в 0,92 при разрешении изображения выше 200 точек на дюйм. Использование указанных параметров встраивания не оказывает влияние на качество формируемого документа, что позволяет говорить о высокой необнаруживаемости встроенного маркера и возможности применения разработанного подхода в системах скрытого внедрения идентификационной информации.

Рецензент: *Марков Алексей Сергеевич, доктор технических наук, старший научный сотрудник, профессор Московского государственного технического университета им. Н.Э. Баумана, г. Москва, Россия. E-mail: a.markov@bmstu.ru*

Литература

1. Через бумажные документы случается каждая девятая утечка конфиденциальных данных // Аналитический центр InfoWatch. 2019. URL: https: //www.infowatch.ru/ analytics/digest/15511 (дата обр. 03.06.2019).
2. Text marking approach for data leakage prevention / A. V. Kozachok [et al.] // Journal of Computer Virology and Hacking Techniques, 2019. URL: DOI:10.1007/s11416-019-00336-9.
3. Подход к извлечению робастного водяного знака из изображений, содержащих текст / А. Козачок [и др.] // Труды СПИИРАН, 2018. 5(60). С. 128-155.
4. Козачок А.В., Копылов С.А., Бочков М.В. Робастный водяной знак как способ защиты текстовых данных от утечки // Защита информации. INSIDE, 2018. Т. 82, № 4. С. 26-32.
5. Козачок А.В., Копылов С.А. Подход к внедрению робастного водяного знака в текстовые данные. 2018. URL: http ://www.ruscrypto.ru/resource/archive/rc2018/files/11_ Kozachok_Kopylov. pdf (дата обр. 13.06.2018).
6. Salomon D. Data privacy and security: encryption and information hiding. // Springer Science & Business Media, 2003. 469 p.
7. Woo C.-S. Digital image watermarking methods for copyright protection and authentication. // Queensland University of Technology, 2007. 197 p.
8. Phadikar A. Robust Watermarking Techniques for Color Images. 2009. URL: https://www.isical.ac.in/~scc/seminars/robust%20watermarking%20techniques.pdf (дата обр. 13.03.2019).
9. Алгоритм встраивания информации в сжатые цифровые изображения на основе операции замены с применением оптимизации / О. Евсютин [и др.] // Компьютерная оптика, 2017. Т. 41, № 3. С. 412-421.
10. Petitcolas F. A., Anderson R. J., Kuhn M. G. Information hiding-a survey // Proceedings of the IEEE, 1999. Vol. 87, no. 7. pp. 1062–1078.
11. Electronic marking and identification techniques to discourage document copying / J. T. Brassil [et al.] // IEEE Journal on Selected Areas in Communications, 1995. Vol. 13, no. 8. pp. 1495-1504.
12. Brassil J. T., Low S., Maxemchuk N. F. Copyright protection for the electronic distribution of text documents // Proceedings of the IEEE, 1999. Vol. 87, no. 7. pp. 1181–1196.
13. Marking text features of document images to deter illicit dissemination / J. T. Brassil [et al.] // Proceedingsof the 12th IARP International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5). Vol. 2. IEEE, 1994. pp. 315-319.
14. Fawcett T. An introduction to ROC analysis // Pattern recognition letters, 2006. Vol. 27, no. 8. pp. 861-874.
15. Powers D. M. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation // Journal of Machine Learning Technologies, 2011. Vol. 2. no. 1. pp 37-63.
16. Sammut C., Webb G. I. Encyclopedia of machine learning // Springer Science & Business Media, 2011. 1032 p.

17. Цифровая стеганография и цифровые водяные знаки. Часть 1. Цифровая стеганография / В. Коржик [и др.] // Санкт-Петербург: СПб ГУТ, 2016. 226 с.
18. Стеганографические системы. Цифровые водяные знаки / В.Г. Грибунин [и др.] // Учеб.-метод. пособие/под ред. д-ра тех. наук В.Г. Грибунина. Саров: ФГУП "РФЯЦ - ВНИИЭФ", 2016. 210 с.
19. Грибунин В., Оков И., Туринцев И. Цифровая стеганография // Москва: СОЛОН-Пресс, 2017. 262 с.
20. Meghanathan N., Nayak L. Steganalysis algorithms for detecting the hidden information in image, audio and video cover media // International journal of Network Security & Its application (IJNSA), 2010. Vol. 2, no. 1. pp. 43-55.
21. Vyas A. O., Dudul S. V. Study of Image Steganalysis Techniques. // International Journal of Advanced Research in Computer Science, 2015. Vol. 6, no. 8. pp. 7-11.
22. Швидченко И. Методы стеганоанализа для графических файлов // Искусственный интеллект, 2010. № 50. С. 697-705.
23. Bachrach M., Shih F. Y. Image steganography and steganalysis // Wiley Interdisciplinary Reviews: Computational Statistics. 2011. Vol. 3, no. 3. pp. 251-259.
24. Kaur M., Kaur G. Review of various steganalysis techniques // International Journal of Computer Science and Information Technologies, 2014. Vol. 5, no. 2. pp. 1744-1747.
25. ROC curve estimation: An overview / L. Gonçalves [et al.] // REVSTAT–Statistical Journal, 2014. Vol. 12, no. 1. pp. 1–20.
26. Cook J. A. ROC curves and nonrandom data // Pattern Recognition Letters. 2017. Vol. 85. pp. 35-41.

UNDETECTABILITY PARAMETERS ESTIMATION OF THE DEVELOPED APPROACH TO TEXT ELECTRON DOCUMENTS MARKING

Kozachok A.V.[4], Kopylov S.A. [5], Bochkov M.V. [6]

The article describes the analysis and evaluation of undetectability parameters to the developed text electronic documents marking approach. A comparative analysis of existing approaches to steganographic analysis of images is carry out. The potentially best steganalysis methods applicable to the developed approach were identified. The following parameters were quantitatively evaluated: stegoanalysis resistance, data extraction reliability, information embedding ratio, image quality after watermark embedding, and watermark secrecy. In the course of the steganalysis, the embedding parameters, which characterize the developed approach as steganalysis resistant, were determined. The perceptual invisibility boundaries of embedded data to visual analysis are established. The obtained results of quantitative and qualitative estimates made it possible to establish the dependence of embedded data invisibility on data embedding parameters used.

Keywords: *text information marking, digital watermarking, steganalysis, data classification, ROC-analysis.*

References:

1. Cherez bumazhnye dokumenty sluchaetsya kazhdaya devyataya utechka konfidencial'nyh dannyh // Analiticheskij centr InfoWatch. 2019. URL: https:// www.infowatch.ru/ analytics/digest/15511 (accessed 03.06.2019). [In Russ.]
2. Text marking approach for data leakage prevention / A. V. Kozachok [et al.] // Journal of Computer Virology and Hacking Techniques, 2019. URL: https://doi.org/10.1007/s11416-019-00336-9.
3. Podhod k izvlecheniyu robastnogo vodyanogo znaka iz izobrazhenij, soderzhashchih tekst / A. Kozachok [et al.] // Trudy SPIIRAN [SPIIRAS Proceedings], 2018. 5(60). pp. 128-155. [In Russ.]
4. Kozachok A.V., Kopylov S.A., Bochkov M.V. Robastnyj vodyanoj znak kak sposob zashchity tekstovyh dannyh ot utechki // Zashchita informacii. INSIDE, 2018, Vol. 82, no. 4. pp. 26-32. [In Russ.]
5. Kozachok A.V., Kopylov S.A. Podhod k vnedreniyu robastnogo vodyanogo znaka v tekstovye dannye. 2018. URL: http//www.ruscrypto.ru/resource/archive/rc2018/files/11_ Kozachok_Kopylov. pdf (accessed 13.06.2018). [In Russ.]
6. Salomon D. Data privacy and security: encryption and information hiding. // Springer Science & Business Media, 2003. 469 p.

4 Alexander Kozachok, Ph. D., Academy of Federal Guard Service, Oryol, Russia. E-mail: a.kozachok@academ.msk.rsnet.ru

5 Sergey Kopylov, Academy of Federal Guard Service, Oryol, Russia. E-mail: gremlin.kop@mail.ru

6 Maksim Bochkov, Dr. Sc., Professor, Private educational institution of additional professional education «Center of entrepreneurial risks», St. Petersburg, Russia. E-mail: mvboch@yandex.ru

7. Woo C.-S. Digital image watermarking methods for copyright protection and authentication. // Queensland University of Technology, 2007. 197 p.
8. Phadikar A. Robust Watermarking Techniques for Color Images. 2009. URL: https://www.isical.ac.in/~scc/seminars/ROBUST%20WATERMARKING%20TECHNIQUES.pdf (accessed. 13.03.2019).
9. Algoritm vstraivaniya informacii v szhatye cifrovye izobrazheniya na osnove operacii zameny s primeneniem optimizacii / O. Evsyutin [et al.] // Komp’yuternaya optika [Computer optics], 2017. Vol. 41, no. 3. pp. 412-421. [In Russ.]
10. Petitcolas F. A., Anderson R. J., Kuhn M. G. Information hiding-a survey // Proceedings of the IEEE, 1999. Vol. 87, no. 7. pp. 1062–1078.
11. Electronic marking and identification techniques to discourage document copying / J. T. Brassil [et al.] // IEEE Journal on Selected Areas in Communications, 1995. Vol. 13, no. 8. pp. 1495-1504.
12. Brassil J. T., Low S., Maxemchuk N. F. Copyright protection for the electronic distribution of text documents // Proceedings of the IEEE, 1999. Vol. 87, no. 7. pp. 1181–1196.
13. Marking text features of document images to deter illicit dissemination / J. T. Brassil [et al.] // Proceedingsof the 12th IARP International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5). Vol. 2. IEEE, 1994. pp. 315-319.
14. Fawcett T. An introduction to ROC analysis // Pattern recognition letters, 2006. Vol. 27, no. 8. pp. 861-874.
15. Powers D. M. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation // Journal of Machine Learning Technologies, 2011. Vol. 2, no. 1. pp 37-63.
16. Sammut C., Webb G. I. Encyclopedia of machine learning // Springer Science & Business Media, 2011. 1032 p.
17. Cifrovaya steganografiya i cifrovye vodyanye znaki. CHast’ 1. Cifrovaya steganografiya / V. Korzhik [et al.] // Sankt-Peterburg: SPb GUT, 2016. 226 p. [In Russ.]
18. Steganograficheskie sistemy. Cifrovye vodyanye znaki / V.G. Gribunin [et al.] // Ucheb.-metod. posobie/pod red. d-ra tekh. nauk V.G. Gribunina. Sarov: FGUP “RFYAC-VNIIEF”, 2016. 210 p. [In Russ.]
19. Gribunin V., Okov I., Turincev I. Cifrovaya steganografiya // Moskva: SOLON-Press, 2017. 262 p. [In Russ.]
20. Meghanathan N., Nayak L. Steganalysis algorithms for detecting the hidden information in image, audio and video cover media // International journal of Network Security & Its application (IJNSA), 2010. Vol. 2, no. 1. pp. 43-55.
21. Vyas A. O., Dudul S. V. Study of Image Steganalysis Techniques. // International Journal of Advanced Research in Computer Science, 2015. Vol. 6, no. 8. pp. 7-11.
22. Shvidchenko I. Metody steganoanaliza dlya graficheskih fajlov // Iskusstvennyj intellekt [Artificial intelligence], 2010, no. 50. pp. 697-705. [In Russ.]
23. Bachrach M., Shih F. Y. Image steganography and steganalysis // Wiley Interdisciplinary Reviews: Computational Statistics. 2011. Vol. 3, no. 3. pp. 251-259.
24. Kaur M., Kaur G. Review of various steganalysis techniques // International Journal of Computer Science and Information Technologies, 2014. Vol. 5, no. 2. pp. 1744-1747.
25. ROC curve estimation: An overview / L. Gonçalves [et al.] // REVSTAT–Statistical Journal, 2014. Vol. 12, no. 1. pp. 1–20.
26. Cook J. A. ROC curves and nonrandom data // Pattern Recognition Letters. 2017. Vol. 85. pp. 35-41.