DOI: 10.55648/1998-6920-2024-18-2-113-126 УДК 003.26

Построение стегосистемы RDH на основе статистических свойств областей изображения *

Е. Ю. Мерзлякова

Сибирский гос. унив. телекоммуникаций и информатики (СибГУТИ)

Аннотация: В статье рассмотрен метод обратимого скрытия данных (RDH) для растровых изображений. Статистические свойства контейнера учитываются путем разделения изображения на связные области методом «лесного пожара» и сбора статистики младших бит для формирования встраиваемой последовательности с заданным распределением. Для разделения изображения на часть для сбора статистических свойств контейнера и часть для встраивания информации применяется метод интерполяции INP. Получение последовательности бит с заданным распределением обеспечивается арифметическим декодером. Построенная стегосистема имеет ёмкость встраивания 0.6 бит/пиксель. Проводится RS-стегоанализ на базе изображений BOSS_v1.01 и оценка свойств полученных контейнеров по показателям визуального искажения.

Ключевые слова: стеганография, цифровые изображения, встраивание информации, интерполяция, стегоанализ.

Для цитирования: Мерзлякова Е. Ю. Построение стегосистемы RDH на основе статистических свойств областей изображения // Вестник СибГУТИ. 2024. Т. 18, № 2. С. 113–126. https://doi.org/10.55648/1998-6920-2024-18-2-113-126.

Контент доступен под лицензией Creative Commons Attribution 4.0 License

© Мерзлякова Е. Ю., 2024

Статья поступила в редакцию 28.04.2024;
переработанный вариант – 02.05.2024;
принята к публикации 19.05.2024.

1. Введение

На сегодняшний день задача обеспечения безопасности передачи информации путем цифровой коммуникации имеет высокую степень значимости, и ее решение основывается на методах криптографии и стеганографии [1]. Классическая задача криптографии состоит в том, чтобы скрыть от третьих лиц содержание сообщения, в то время как классическая задача стеганографии – скрыть сам факт передачи сообщения. Данные области взаимосвязаны, например, стеганографические схемы могут использоваться для надежной передачи ключей в безусловно стойких криптосистемах [2], а криптографические методы, в свою очередь, повышают безопасность стеганографических схем [3–6].

Стеганография как наука возникла достаточно давно, развиваясь в направлении повышения надежности и безопасности методов передачи скрытой информации в безобидных на вид контейнерах. В качестве контейнеров в настоящее время могут использоваться текстовые файлы, исполняемые файлы, графические и звуковые файлы, а также графы [7]. Заполненный скрытой информацией контейнер называют стегоконтейнером. В стегосистеме,

* Работа выполнена в рамках госзадания СибГУТИ № 071-03-2024-008 от 19.01.2024.

предназначенной для скрытой передачи информации, противоборствующая сторона представлена пассивным нарушителем [1], который пытается установить факт применения стегосистемы, не воздействуя на стегоконтейнеры.

Отдельную значимую область занимает стеганография цифровых изображений. Внесение искажений в определенные данные изображений остаются незаметными для человека, в особенности при наличии некоторого шума квантования, свойственного оцифрованным объектам. Чаще всего информация встраивается в наименее значащие биты (LSB) элементов контейнера [8–10]. Особенный интерес представляют методы обратимого скрытия данных (RDH) в изображениях. Один из общих подходов к разработке стеганографических систем RDH состоит в применении методов интерполяции [11–16], позволяющих увеличить полезную нагрузку контейнера, сохраняя визуальное качество. Изображение интерполируется без изменения оригинальных пикселей, затем интерполированные пиксели используются для встраивания информации. Применяемые алгоритмы интерполяции в целях сокрытия данных исследуют с помощью показателей визуального искажения [17], выбирая наиболее оптимальный метод. Так, если значение показателя меры искажения изображения (PSNR) более 40 дБ, то качество обработки изображения считается высоким [18]. В [13] A. Tripathi и J. Prakash предложили новый подход к применению интерполяции изображений и достигли лучших значений индекса PSNR для полученного LSB-метода по сравнению с известными стегосистемами данного вида. Методы RDH особенно полезны при скрытии данных пациента в медицинских изображениях, а также в обеспечении конфиденциальности и целостности обрабатываемых данных интернета вещей (IoT), когда в качестве контейнера чаще всего используются цифровые изображения [19, 20].

Наряду с методами встраивания данных активно развиваются и методы раскрытия стеганографических систем [21], называемые стегоанализом. Под раскрытием стеганографической системы понимается нахождение ее уязвимости, которая позволяет определить факт скрытия информации в контейнере и возможность доказать данное утверждение третьей стороне с высокой степенью достоверности [1]. Один из простых и доступных методов обнаружения наличия встроенных данных для стеганографических систем LSB-замены – это метод визульной атаки, в котором отдельно рассматриваются младшие биты изображения. Более современные подходы к стегоанализу включают в себя модели оценки корреляций в изображениях [22], методы машинного обучения [23, 24] и сверточные нейронные сети [25, 26]. При разработке новых алгоритмов встраивания информации немаловажную роль играет применение универсальных методов стегоанализа к полученным контейнерам, но чаще всего исследования ограничиваются подсчетом показателя PSNR.

Цель данной работы заключается в построении RDH-стегосистемы для скрытия информации в младших битах растрового изображения, использующей кодирование встраиваемой информации в соответствии с оценками вероятностей появления значений младшего бита, что приблизит стегосистему к совершенной [27]. Для оценивания статистических свойств контейнера используется исходное растровое изображение, матрица пикселей которого разбивается на множество связных областей методом «лесного пожара» [28]. Затем исходное изображение интерполируется и добавленные пиксели используются для встраивания скрытой информации с учетом статистики младших бит по областям. Для получения последовательности заданного распределения целесообразно использовать арифметический декодер, а для восстановления встроенной информации – соответствующий арифметический кодер [29]. Разработанный метод может быть легко реализован на практике для большинства графических форматов данных, использующих неискажающие методы сжатия.

Для оценки разработанной системы, помимо показателей качества контейнеров, использующихся в большинстве работ по схожей тематике, применяется метод оценки визуального искажения, а также метод двойной статистики, полученной из пространственных корреляций в изображениях [22]. Это позволит более полно оценить результаты работы стегосистемы, выявить ее недостатки и определить дальнейшие пути улучшения стойкости к стегоанализу.

2. Материалы и методы

2.1. Общая схема предложенной стегосистемы

Рассмотрим алгоритм работы предлагаемой стегосистемы в общем виде. Исходными данными отправителя являются сообщение $\boldsymbol{M}$ и оригинальный контейнер $\boldsymbol{A}$. Сообщение представляет собой псевдослучайную битовую последовательность, которая имитирует зашифрованную информацию, предназначенную для хранения в контейнере. Оригинальный контейнер – это 8-битное bmp-изображение в оттенках серого (в данном случае размером 256 на 256 точек), представляющее собой информационную ценность при передаче. В данных исследованиях для формирования тестового набора контейнеров использовались 1000 изображений из базы данных BOSSBase vl.01 [30], содержащей файлы формата pgm размером 512 на 512 точек. Стандартными средствами изображения из базы были преобразованы в формат bmp для формирования набора **изображений-источников** и затем уменьшены до размера 256 на 256 точек для набора оригинальных контейнеров.

На первом шаге алгоритма в контейнере $\boldsymbol{A}$ определяются связные области, для каждой из которых происходит сбор статистики младших бит данного контейнера. Под связной областью изображения здесь понимается такая его область, в которой все точки имеют одинаковое значение признака и в которой между любыми двумя точками существует непрерывный путь точек данной области [28]. В представленной в данной работе стегосистеме области выделяются методом «лесного пожара» по признаку попадания в определенные диапазоны значений яркости. Таким образом, формируется большое количество наборов областей изображения со своим распределением.

На втором шаге применяется алгоритм интерполяции для оригинального контейнера $\boldsymbol{A}$, в результате которого создается контейнер-обложка $\boldsymbol{C}$ размером 512 на 512 точек, содержащий точки оригинального контейнера и интерполированные значения.

На третьем шаге происходит аналогичное выделение областей для интерполированных значений контейнера-обложки $\boldsymbol{C}$. Для каждой такой области используется собранная статистика младших бит соответствующей области оригинального контейнера $\boldsymbol{A}$, чтобы преобразовать сообщение $\boldsymbol{M}$ в код с заданным распределением вероятностей. Соответствующий код записывается в младшие биты каждой из выделенных областей контейнера-обложки $\boldsymbol{C}$ со своей статистикой, в результате чего создается стегоконтейнер $\boldsymbol{S}$ со встроенным сообщением (рис. 1).

Рис. 1. Схема встраивания сообщения $\boldsymbol{M}$ в контейнер $\boldsymbol{C}$

Для извлечения сообщения из $\boldsymbol{S}$ восстанавливается оригинальный контейнер $\boldsymbol{A}$, по которому определяется статистика младших бит по выделенным областям тем же способом, что и при встраивании. Затем $\boldsymbol{A}$ снова интерполируется, чтобы получить контейнер $\boldsymbol{C}$ без

встроенного сообщения для безошибочного определения соответствующих областей в $\boldsymbol{S}$. Далее с помощью полученного распределения вероятностей младших бит по областям контейнера $\boldsymbol{A}$ восстанавливается сообщение $\boldsymbol{M}$ из кода, находящегося в областях $\boldsymbol{S}$ (рис. 2).

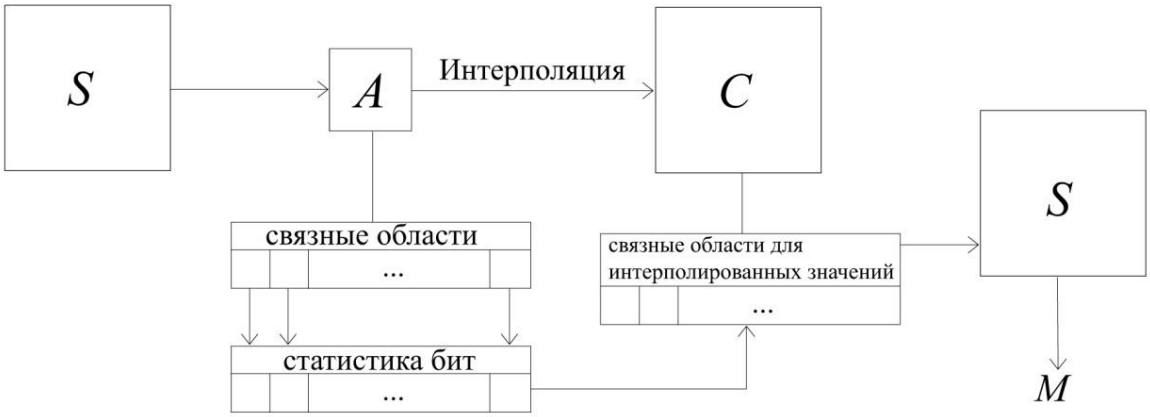


Рис. 2. Схема извлечения сообщения $\boldsymbol{M}$ из контейнера $\boldsymbol{S}$

Таким образом, контейнер $\boldsymbol{S}$ всегда содержит в себе пиксели оригинального контейнера $\boldsymbol{A}$, который сам по себе может иметь информационную нагрузку при передаче, являясь, например, произведением искусства, медицинским снимком или изображением с камеры умного дома.

2.2. Метод интерполяции

В представленной стегосистеме за основу может быть взят любой алгоритм интерполяции изображений. Критерием выбора алгоритма является показатель PSNR. Для того, чтобы его оценить, необходимо иметь источник оригинального контейнера того же размера, что и контейнер-обложка, полученный путем интерполяции оригинального изображения. Поэтому для оценки метода интерполяции используется тестовый набор изображений- источников, о которых шла речь в разд. 2.1. Заметим, что непосредственно в работе стегосистемы изображения-источники не принимают участия и используются только для оценки качества метода.

Исследования [31, 32] показали, что алгоритм INP [16] является наиболее оптимальным с точки зрения качества получаемого контейнера в стегосистемах RDH. В представленной в данной работе стегосистеме также применяется алгоритм интерполяции INP. Значения точек с координатами i и j контейнера-обложки $\boldsymbol{C}$ из оригинального изображения $\boldsymbol{A}$ размером m на n пикселей вычисляются по формуле:

$$C(i,j)=\begin{cases} A(i,j),\ i=2m,\ j=2n, \\ \dfrac{(A(i,j-1)+\left(\dfrac{A(i,j-1)+A(i,j+1)}{2}\right)}{2},\ i=2m,\ j=2n+1, \\ \dfrac{(A(i-1,j)+\left(\dfrac{A(i-1,j)+A(i+1,j)}{2}\right)}{2},\ i=2m+1,\ j=2n, \\ \dfrac{C(i-1,j)+C(i,j-1)}{2},\ \text{иначе.} \end{cases}. \quad (1)$$

Пример вычисления интерполированных значений на фрагменте некоторого изображения $\boldsymbol{A}$ показан на рис. 3. Серым цветом выделены значения контейнера $\boldsymbol{C}$, полученные по формуле (1).

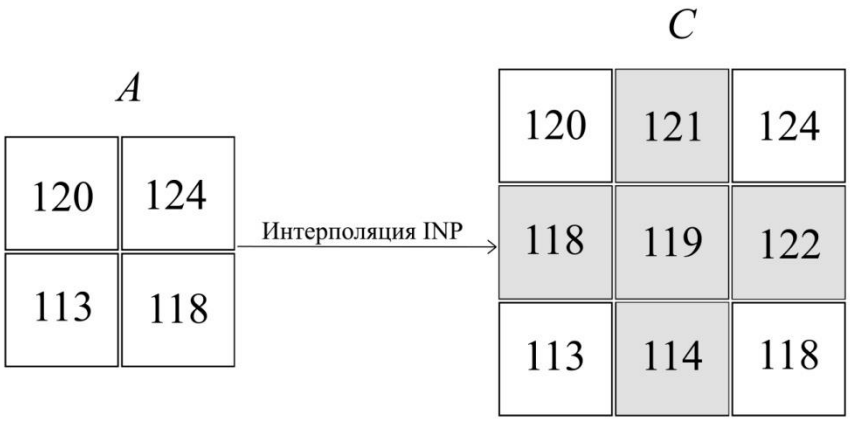


Рис. 3. Пример вычисления значений интерполяции

2.3. Оценка статистики младших бит

Известно, что существует статистическая зависимость младших бит от остальных элементов контейнера. Эта зависимость нарушается, если независимые данные будут записаны в младшие биты. В работе [33] ранее была предложена адаптивная статистическая модель, позволяющая оценить вероятности распределения младших бит в контейнере. С помощью оценки частотной энтропии были выявлены наиболее существенные элементы контекста, в наибольшей степени связанные с младшими битами. В качестве обучающей части была использована половина изображения-контейнера, выделенная в шахматном порядке. Но в настоящей работе мы используем метод интерполяции для отделения обучающей части изображения, что накладывает определенные условия на формирование контекста, так как каждая вторая строка не будет содержать обучающих данных для сбора статистики. В качестве обучающей матрицы для формирования статистики здесь используется оригинальное изображение $\boldsymbol{A}$, которое затем интерполируется.

Вероятности оцениваются по пикселям оригинального изображения, причем адаптивно к каждой конкретной области. Пространство состояний для оценки вероятностей определяется с учетом частотной энтропии: чем меньше величина энтропии, тем ближе статистическая модель к реальному источнику. Оптимальный контекст состоит из самых младших бит предыдущего и последующего пикселей, а также второго, третьего и четвертого младших бит из текущего пикселя. Заметим, что предыдущий и последующий пиксели в контейнере $\boldsymbol{C}$ относятся к матрице оригинального изображения $\boldsymbol{A}$, поэтому они остаются неизменными при встраивании информации, и их наименее значащие биты можно использовать при восстановлении статистики. Используемые биты текущего пикселя также остаются неизменными и используются при восстановлении статистики, так как информация будет записываться в наименее значащий бит интерполированных значений матрицы пикселей $\boldsymbol{C}$.

2.4. Метод получения заданного распределения

Наиболее эффективным приближенным решением задачи кодирования, преобразующего зашифрованное сообщение в поток бит с заданным распределением, является арифметическое декодирование [34]. Арифметическому декодеру указывается требуемое распределение вероятностей нуля и единицы для каждого следующего выходного символа, в соответствии с которым он воспроизводит этот символ, рассматривая зашифрованное сообщение как код, построенный ранее соответствующим арифметическим кодером. Чтобы восстановить

зашифрованное сообщение из его кода, применяется арифметический кодер, на вход которого подаются те же самые распределения вероятностей, что и декодеру.

Для оценки статистики младших бит и внедрения информации с учетом этой статистики, следуя схемам на рис. 1, 2, используется контейнер $\boldsymbol{A}$, пиксели которого являются «обучающей» частью контейнера $\boldsymbol{C}$ при внедрении и контейнера $\boldsymbol{S}$ при извлечении сообщения $\boldsymbol{M}$.

2.5. Метод выделения связных областей изображения

Ранее в подобных экспериментах был использован метод разделения изображения на связные области по яркостям. Всё изображение делилось на 16 связных областей по заданным диапазонам яркости. Но такой подход позволял собирать в одну область пиксели, которые могли находиться в разных частях изображения, что в определенных случаях искажало статистические свойства контейнера.

В данной работе использован стековый алгоритм прослеживания связных областей методом «лесного пожара» [28]. Идея данного метода состоит в том, что область «поджигается» в одной точке, а затем каждая «подожжённая» точка далее «поджигает» все соседние точки, имеющие ту же яркость. Сначала «поджог» начинается с левой верхней точки изображения. Каждая «подожжённая» точка имеет свою метку номера области и затем для формирования последующих областей точка нового «поджога» берется как следующая точка, не имеющая метки. В итоге мы получаем некоторое количество непересекающихся связных областей, для каждого изображения разное. Объем выжженной области может варьироваться от 1 до максимально возможного количества пикселей в картинке заданного размера. Важно учесть, что каждая полученная область будет использоваться для сбора статистики младших бит в определенном контексте, размер которого в данной работе составляет 5 элементов. Поэтому необходимо адаптировать метод выделения областей под данную задачу сбора статистики.

Типичное изображение-контейнер для встраивания данных представляет собой достаточно шумную оцифрованную фотографию. Если в методе лесного пожара каждая область будет выделяться по точному значению яркости, то мы получим огромное количество областей, большинство из которых будет содержать недостаточное количество пикселей для сбора представительной статистики. Поэтому в данной работе предлагается определять пиксели в ту или иную область по принципу попадания в общий диапазон значений для каждой области. Чем меньше разброс значений в диапазоне, тем больше количество областей и меньше их объем. Таким образом, необходимо установить наиболее эффективное значение диапазонов, при котором будет получено достаточное количество представительных областей для сбора статистики, позволяющих достигнуть наибольшей ёмкости метода встраивания.

Пусть R – это максимальная разность между пикселями в каждой выделяемой области. Тогда, например, при $R = 16$ пиксели изображения будут распределены на области из диапазонов: 0 – 16, 17 – 33, 34 – 50, 51 – 67, 68 – 84 и так далее. При $R = 8$ диапазоны будут: 0 – 8, 9 – 17, 18 – 26, 27 – 35, 36 – 44 и так далее. Заметим, что при любом значении R неизбежно будет встречаться некоторое количество мелких областей, которые не позволят сформировать статистику и при встраивании будут иметь скорее случайное распределение младших бит. В то же время большие области при встраивании информации будут иметь наиболее приближенное к истинному распределение вероятностей младших бит.

2.6. Используемые методы стегоанализа

Первое условие, которое выполняют все стеганографические программы, состоит в том, что искажения, вносимые в контейнер при внедрении сообщения, должны быть незаметны для человека. Так, даже при стопроцентном заполнении, т.е. при замене всех младших бит матрицы изображения, мы обычно не видим искажений, особенно если под рукой нет исходного файла для сравнения. Однако при помощи простого визуального метода стегоанализа легко

обнаружить замену младших бит. Для этого значения пикселей контейнера $\boldsymbol{S}$ преобразуются в соответствии с правилом:

$$S'(i,j)=\begin{cases}255, \text{ если } S(i,j)=1,\\ 0, \text{ если } S(i,j)=0,\end{cases} \tag{2}$$

где i, j – индексы текущего пикселя.

Далее методом визуальной оценки происходит сравнение общих контуров на изображении до и после преобразования. Так, при замене младших бит мы увидим случайный шум, а при сохранении корреляций в младших битах будут определяться общие контуры изображения.

В данной работе также применяется метод RS-анализа (Regular-Singular) для оценки стегостойкости полученного метода. Суть анализа заключается в поиске пространственных корреляций в контейнере: изображение разбивается на связные группы пикселей, для которых затем определяется вещественная функция гладкости:

$$f(G)=\sum_{i=1}^{n}\left|x_{i+1}-x_i\right|, \tag{3}$$

где G – это набор из n групп $x_1 \ldots x_n$ связных пикселей контейнера. Чем больше значение данной функции, тем более зашумленной является группа G. Затем моделируется обратимое добавление шума в группы, что увеличивает функцию гладкости в случае пустых контейнеров и усиливает всплески значений в заполненных контейнерах [26]. На основании результатов группы делятся на регулярные (regular) и сингулярные (singular), вычисляются их количественные характеристики, по которым строится график. На выходе получают предполагаемую длину встроенной в контейнер информации (L). Метод RS может указать небольшую ненулевую длину сообщения из-за случайных отклонений даже для пустого контейнера [22]. В работе [35] установлено, что при $L \geq 5$ RS-анализ классифицирует контейнер как заполненный.

В стеганографии задача классификации контейнеров сводится к выбору одной из двух гипотез: H_0 – контейнер является пустым и H_1 – контейнер имеет встроенную информацию. Но любые стегоалгоритмы могут ошибочно классифицировать контейнер. В связи с этим принято разделять ошибки на два рода: ошибка первого рода возникает, когда стегоаналитиком выбрана гипотеза H_1 при пустом контейнере (ложное срабатывание), и ошибка второго рода, когда выбрана гипотеза H_0 при заполненном контейнере (пропуск). Поэтому, прежде чем оценивать метод на устойчивость к обнаружению, необходимо определить долю ложных срабатываний для тестового набора пустых контейнеров.

Для оценки качества заполненных контейнеров в стеганографии также используется показатель максимального соотношения сигнал/шум (PSNR):

$$\text{PSNR} = 10 * \log \frac{255^2}{\varepsilon}, \tag{4}$$

где $\varepsilon = \frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N}\left(S(i,j)-I(i,j)\right)^2$ – искажение, M, N – высота и ширина изображения соответственно. Чем больше значение PSNR, тем меньше расхождений между сравниваемыми изображениями.

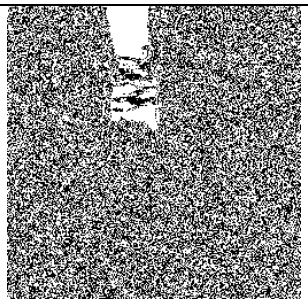
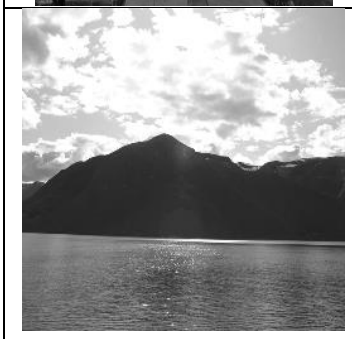
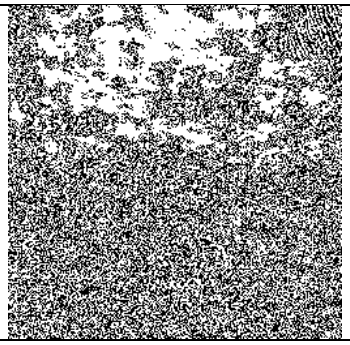
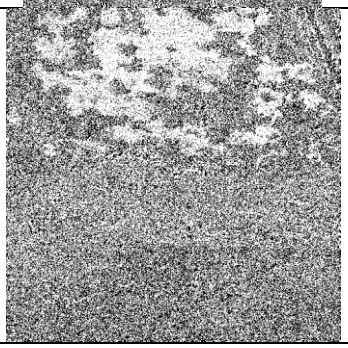
3. Результаты

Максимальная средняя ёмкость встраивания по предложенному в данной работе стеганографическому алгоритму составила 0.6 бит/пиксель при $R = 8$.

Визуальный метод анализа показал, что некоторые корреляции младших бит заметно сохраняются в заполненном контейнере. В табл. 1 показаны значения младших бит, следуя

формуле (2) для изображения-источника, для пустого интерполированного контейнера и для заполненного контейнера. Размеры представленных контейнеров: 512 на 512.

Таблица 1. Визуальный анализ младших бит

Источник	Источник $'$	Заполненный контейнер S'
		
		

Метод RS-анализа на наборе из 1000 пустых контейнеров-источников размером 512 на 512 показал ошибку первого рода 27 %. Для соответствующих заполненных контейнеров процент файлов, классифицированных как заполненные, составил 35 %, что говорит о 8 % раскрытых стегоконтейнеров.

Таблица 2. RS-анализ пустых и заполненных контейнеров

L	0	1–4	5 и более
Изображения-источники	16 %	57 %	27 %
Контейнеры S	9 %	56 %	35 %

Значения PSNR для нескольких тестовых контейнеров приведены в табл. 3. Данный показатель в разработанном методе свидетельствует о достаточно хорошем качестве полученных стегоконтейнеров и превосходит результат методов, основанных на интерполяции без использования статистики [11, 16, 12], в которых ёмкость встраивания составляла 0.4–0.9 бит/пиксель.

Таблица 3. PSNR контейнеров

Изображения-источники	Контейнеры S	PSNR	Бит/пиксель
		59	0.55
		65	0.57
		57	0.62
		61	0.59
		58	0.60

4. Обсуждение и заключения

Предложенная стеганографическая схема имеет более высокий показатель PSNR по сравнению с другими стеганографическими схемами, использующими методы интерполяции в RDH и при этом сохраняет корреляцию в младших битах контейнеров. Ёмкость встраивания для большинства контейнеров составляет 0.6 бит/пиксель, за исключением тех контейнеров, которые содержат большие однотонные области, в которых значение яркости идентично. Однотонные непрерывные области растягивают последовательность бит на выходе декодера, тем самым сокращая полезную нагрузку. Такие файлы легко исключить из набора контейнеров, так как даже визуально их несложно определить.

Анализ методом RS показал, что часть контейнеров уязвимы к стегоанализу, что указывает на необходимость более качественного учёта зависимостей в значениях младших бит изображения. В то же время полученный результат оказался лучше, чем предыдущий с разделением областей, что свидетельствует об эффективности применения метода «лесного пожара» для выделения связных областей. Некоторое искажение статистики младших бит в стегоконтейнере ожидаемо по причине неизбежного наличия мелких связных областей, не подходящих для сбора статистики. Таким образом, в дальнейшем планируется адаптировать данный алгоритм, используя мелкие связные области для встраивания с применением других стегометодов.

Литература

1. *Грибунин В. Г., Костюков В. Е., Мартынов А. П., Николаев Д. Б., Фомченко В. Н.* Прикладная стеганография: для студентов, аспирантов, научных работников, изучающих вопросы обеспечения безопасности информации. Саров: Российский Федеральный ядерный центр – Всероссийский научно-исследовательский институт экспериментальной физики, 2021. 484 с.
2. *Ryabko B.* Unconditionally secure short key ciphers based on data compression and randomization // Designs, Codes, and Cryptography. 2023. V. 91, № 6. P. 2201–2212.
3. *Монарёв В. А., Фионов А. Н., Шокин Ю. И.* Обзор современных теоретико-информационных подходов к решению основных задач криптографии и стеганографии // Вычислительные технологии. 2010. Т. 15, № 2. С. 69–86.
4. *Tutuncu K., Çataltaş Ö.* Compensation of degradation, security, and capacity of LSB substitution methods by a new proposed hybrid n-LSB approach // Computer Science and Information Systems. 2021. V. 18, № 4. P. 1311–1332.
5. *Talasila S., Vijaya Kumar G., Vijaya Babu E., Nainika K., Veda Sahithi M., Mohan P.* The Hybrid Model of LSB-Technique in Image Steganography Using AES and RSA Algorithms // Soft Computing and Signal Processing. 2023. V. 840. P. 403–413.
6. *Rajabi-Ghaleh S., Olyaeefar B., Kheradmand R., Ahmadi-Kandjani S.* Image security using steganography and cryptography with sweeping computational ghost imaging // Frontiers in Physics. 2024. V. 12. № 1336795.
7. *Нечта И. В.* Цифровая стеганография в программах и текстовых файлах. М.: Горячая линия – Телеком, 2023. 112 с.
8. *Novanto F., Nugraha A., Kurniawan J., Prayogo A.* Optimizing Digital Image Steganography through Hybridization of LSB and Zstandard Compression // Sinkron. 2024. V. 9, № 1. P. 75–82.
9. *Wu W., Li H.* A novel scheme for random sequential high-capacity data hiding based on PVD and LSB // Signal, Image and Video Processing. 2023. V. 18. P. 2277–2287.
10. *Mali A., Dongre M.* Block Based Self-Secured LSB Embedding Scheme for Reversible Steganography // IOT with Smart Systems. 2023. V. 2. P. 625–635.

11. *Mohammad A., Al-Haj A., Farfoura M.* An improved capacity data hiding technique based on image interpolation // Multimedia Tools and Applications. 2018. V. 78, № 6. P. 7181–7205.
12. *Jung K., Yoo K.* Data hiding method using image interpolation // Computer Standards & Interfaces. 2009. V. 31, № 2. P. 465–470.
13. *Tripathi A., Prakash J.* Interpolation Based Reversible Data Hiding using Pixel Intensity Classes // International Journal of Next-Generation Computing. 2023. V. 14, № 4. P. 621–643.
14. *Punia R., Malik A., Singh S.* An interpolation-based reversible data hiding scheme for internet of things applications // Discover Internet of Things. 2023. V. 3, № 18.
15. *Mohammad A. A.* high quality interpolation-based reversible data hiding technique using dual images // Multimedia Tools and Applications. 2023. V. 82. P. 36713–36737.
16. *Lee C-F., Huang Y-L.* An efficient image interpolation increasing payload in reversible data hiding // Expert Systems with Applications. 2012. V. 39, № 8. P. 6712–6719.
17. *Конахович Г. Ф., Пузыренко А. Ю.* Компьютерная стеганография: теория и практика. Киев: МК–Пресс, 2006. 283 с.
18. *Kumari L., Ramanathan P., Rani J., Vinothkumar D., Sneha A., Amalarani V., Joe B.* Selection of optimum compression algorithms based on the characterization on feasibility for medical image // Biomedical Research. 2017. V. 28, № 13. P. 5633–5637.
19. *Евсютин О. О., Кокурина А. С., Мещеряков Р. В.* Обзор методов встраивания информации в цифровые объекты для обеспечения безопасности в «интернете вещей» // Компьютерная оптика. 2019. Т. 43, № 1. P. 137–154.
20. *Kuri J., Rafi M.* Securing Data in Internet of Things (IoT) using Cryptography and Steganography Techniques // International Journal for Research in Applied Science & Engineering Technology. 2020. V. 8, № 7. P. 1933–1939.
21. *Шелухин О. И.* Стеганография. Алгоритмы и программная реализация. М.: Горячая линия – Телеком, 2017. 592 с.
22. *Fridrich J., Goljan M., Du R.* Reliable Detection of LSB Steganography in Grayscale and Color Images // Special Session on Multimedia Security and Watermarking. Ottawa, Canada, 2001. P. 27–30.
23. *Pevny T., Bas P., Fridrich J.* Steganalysis by subtractive pixel adjacency matrix // IEEE Transactions on Information Forensics and Security. 2010. V. 5, № 2. P. 215–224.
24. *Holub V., Fridrich J.* Random projections of residuals for digital image steganalysis // IEEE Transactions on Information Forensics and Security. 2013. V. 8, № 12. P. 1996–2006.
25. *Сирота А. А., Дрюченко М. А., Иванков А. Ю.* Стегоанализ цифровых изображений с использованием методов поверхностного и глубокого машинного обучения: известные подходы и новые решения // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2021. № 1. С. 33–52.
26. *Полунин А. А., Яндашевская Э. А.* Использование аппарата свёрточных нейронных сетей для стегоанализа цифровых изображений // Труды Института системного программирования РАН. 2020. Т. 32, № 4. С. 155–164.
27. *Ryabko B., Ryabko D.* Information-theoretic approach to steganographic systems // Proc. IEEE International Symposium on Information Theory, Nice, France, 2007. P. 2461–2464.
28. *Визильтер Ю. В., Желтов С. Ю., Бондаренко А. В., Ососков М. В., Моржин А. В.* Обработка и анализ изображений в задачах машинного зрения: курс лекций и практических занятий. М.: Физматкнига, 2010. 672 с.
29. *Said A.* Introduction to Arithmetic Coding – Theory and Practice. Hewlett-Packard Laboratories Report. USA, 2004. 64 p.
30. База BOSS изображений для исследований методов стеганографии [сайт]. URL: `http://agents.fel.cvut.cz/boss/index.php?mode=VIEW& tmpl=about` (дата обращения: 26.03.2024).
31. *Mahasree M.* Improved Reversible Data Hiding in Medical images using Interpolation and Threshold based Embedding Strategy // International Journal of Emerging Trends in Engineering Research. 2020. V. 8. P. 3495–3501.

32. *Lu T.-C., Lin M.-C., Huang C.-C., Deng K.-M.* Reversible Data Hiding Based on Image Interpolation with a Secret Message Reduction Strategy // International Journal of Computer Software Engineering. 2016. V. 1. P. 124–130.
33. *Eltysheva K., Fionov A.* Stegosystem construction on the basis of statistical structure of covertext // Proc. XII International Symposium on Problems of Redundancy, St.-Petersburg, May 26–30, 2009. P. 180–185.
34. *Рябко Б. Я., Фионов А. Н.* Эффективный метод адаптивного арифметического кодирования для источников с большими алфавитами // Проблемы передачи информации. 1999. Т. 35, № 4. С. 1–14.
35. *Жилкин М. Ю.* Теоретико-информационные методы стегоанализа графических данных: дис. … канд. тех. наук: 05.12.13. Сиб. гос. ун-т телекоммуникаций и информатики, Новосибирск, 2009. 153 с.

Мерзлякова Екатерина Юрьевна
к.т.н., доцент кафедры прикладной математики и кибернетики, Сибирский государственный университет телекоммуникаций и информатики (СибГУТИ, 630102, Новосибирск, ул. Кирова, д. 86), тел. +7 383 2698 272, e-mail: katerina.artist@yandex.ru, ORCID ID: 0000-0001-6847-5588.

Автор прочитал и одобрил окончательный вариант рукописи.
Автор заявляет об отсутствии конфликта интересов.

Construction of the RDH Stegosystem Based on the Statistical Properties of Image Areas

Ekaterina Yu. Merzlyakova

Siberian State University of Telecommunications and Information Science (SibSUTIS)

Abstract: This article discusses the reversible data hiding (RDH) method for raster images. The statistical properties of the container are taken into account by dividing the image into coherent regions using a wildfire method and collecting statistics of the least significant bits to form an embedded sequence with a given distribution. The INP interpolation method is used to divide the image into a part for collecting statistical properties of the container and a part for embedding information. Obtaining a sequence of bits with a given distribution is provided by an arithmetic decoder. The constructed stegosystem has an embedding capacity of 0.6 bits/pixel. RS steganalysis is carried out on the basis of BOSS_v1.01 images and the properties of the resulting containers are assessed based on visual distortion indicators.

Keywords: steganography, digital images, information embedding, interpolation, steganalysis.

For citation: Merzlyakova E. Yu. Construction of the RDH stegosystem based on the statistical properties of image areas. 2024 (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 2, pp. 113-126. https://doi.org/10.55648/1998-6920-2024-18-2-113-126.

Content is available under the license Creative Commons Attribution 4.0 License

©. Merzlyakova E. Yu., 2024

The article was submitted: 28.04.2024;
revised version: 02.05.2024;
accepted for publication 19.05.2024.

References

1. Gribunin V. G., Kostyukov V. E., Martynov A. P., Nikolaev D. B., Fomchenko V. N. *Prikladnaya steganografiya: dlya studentov, aspirantov, nauchnykh rabotnikov, izuchayushchikh voprosy obespecheniya bezopasnosti informatsii* [Applied steganography: for students, graduate students, researchers studying information security issues]. Sarov, Rossiiskii Federal'nyi yadernyi tsentr - Vserossiiskii nauchno-issledovatel'skii institut eksperimental'noi fiziki, 2021, 484 p.
2. Ryabko B. Unconditionally secure short key ciphers based on data compression and randomization. *Designs, Codes, and Cryptography*, 2023, vol. 91, no. 6, pp. 2201-2212.
3. Monarev V. A., Fionov A. N., Shokin Yu. I. Obzor sovremennykh teoretiko-informatsionnykh podkhodov k resheniyu osnovnykh zadach kriptografii i steganografii [Review of modern information-theoretic approaches to solving basic problems of cryptography and steganography]. *Vychislitel'nye tekhnologii*, 2010, vol. 15, no. 2, pp. 69-86.
4. Tutuncu K., Çataltaş Ö. Compensation of degradation, security, and capacity of LSB substitution methods by a new proposed hybrid n-LSB approach. *Computer Science and Information Systems*, 2021, vol. 18, no 4, pp. 1311-1332.
5. Talasila S., Vijaya Kumar G., Vijaya Babu E., Nainika K., Veda Sahithi M., Mohan P. The Hybrid Model of LSB-Technique in Image Steganography Using AES and RSA Algorithms. *Soft Computing and Signal Processing*, 2023, vol. 840, pp. 403–413.
6. Rajabi-Ghaleh S., Olyaeefar B., Kheradmand R., Ahmadi-Kandjani S. Image security using steganography and cryptography with sweeping computational ghost imaging. *Frontiers in Physics*, 2024, vol.12, no. 1336795.
7. Nechta I. V. *Tsifrovaya steganografiya v programmakh i tekstovykh failakh* [Digital steganography in programs and text files]. Moscow, Goryachaya liniya, Telekom, 2023, 112 p.
8. Novanto F., Nugraha A., Kurniawan J., Prayogo A. Optimizing Digital Image Steganography through Hybridization of LSB and Zstandard Compression. *Sinkron*, 2024, vol. 9, no. 1. pp. 75-82.
9. Wu W., Li H., A novel scheme for random sequential high-capacity data hiding based on PVD and LSB. *Signal, Image and Video Processing*, 2023, vol. 18, pp. 2277-2287.
10. Mali A., Dongre M. Block Based Self-Secured LSB Embedding Scheme for Reversible Steganography. *IOT with Smart Systems*, 2023, vol. 2, pp. 625-635.
11. Mohammad A., Al-Haj A., Farfoura M. An improved capacity data hiding technique based on image interpolation. *Multimedia Tools and Applications*, 2018, vol. 78, no. 6, pp. 7181-7205.
12. K. Jung, K. Yoo. Data hiding method using image interpolation. *Computer Standards & Interfaces*, 2009, vol. 31, no. 2, pp. 465-470.
13. Tripathi A., Prakash J. Interpolation Based Reversible Data Hiding using Pixel Intensity Classes. *International Journal of Next-Generation Computing*, 2023, vol. 14, no. 4, pp. 621-643.
14. Punia R., Malik A., Singh S. An interpolation-based reversible data hiding scheme for internet of things applications. *Discover Internet of Things*, 2023, vol. 3, no. 18.
15. Mohammad A. A. High quality interpolation-based reversible data hiding technique using dual images. *Multimedia Tools and Applications*, 2023, vol. 82, pp. 36713-36737.
16. Lee C-F., Huang Y-L. An efficient image interpolation increasing payload in reversible data hiding. *Expert Systems with Applications*, 2012, vol. 39, no. 8, pp. 6712-6719.
17. Konakhovich G. F., Puzyrenko A. Yu. *Komp'yuternaya steganografiya: teoriya i praktika* [Computer steganography: theory and practice]. Kyiv: MK-Press, 2006, 283 p.
18. Kumari L., Ramanathan Pandian., Rani J., Vinothkumar D., Sneha A., Amalarani V., Joe B. Selection of optimum compression algorithms based on the characterization on feasibility formedical image. *Biomedical Research*, 2017, vol. 28, no. 13, pp. 5633-5637.
19. Evsyutin O. O., Kokurina A. S., Meshcheryakov R. V., Obzor metodov vstraivaniya informatsii v tsifrovye ob"ekty dlya obespecheniya bezopasnosti v «internete veshchei» [Review of methods for embedding information into digital objects to ensure security in the Internet of Things]. *Komp'yuternaya optika*, 2019, vol. 43, no. 1, pp.137-154.
20. Kuri J., Rafi M. Securing Data in Internet of Things (IoT) using Cryptography and Steganography Techniques. *International Journal for Research in Applied Science & Engineering Technology*, 2020, vol. 8, no. 7, pp. 1933-1939.
21. Shelukhin O. I. *Steganografiya. Algoritmy i programmnaya realizatsiya* [Steganography. Algorithms and software implementation]. Moscow, Goryachaya liniya - Telekom, 2017, 592 p.

22. Fridrich J., Goljan M., Du R. Reliable Detection of LSB Steganography in Grayscale and Color Images. *Special Session on Multimedia Security and Watermarking*. Ottawa, Canada, 2001, pp. 27-30.
23. Pevny T., Bas P., Fridrich J. Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on Information Forensics and Security*, 2010, vol. 5, no. 2, pp. 215-224.
24. Holub V., Fridrich J. Random projections of residuals for digital image steganalysis. *IEEE Transactions on Information Forensics and Security*, 2013, vol. 8, no. 12, pp. 1996-2006.
25. Sirota A. A., Dryuchenko M. A., Ivankov A. Yu. Stegoanaliz tsifrovykh izobrazhenii s is-pol'zovaniem metodov poverkhnostnogo i glubokogo mashinnogo obucheniya: izvestnye podkhody i novye resheniya [Steganalysis of digital images using shallow and deep machine learning methods: known approaches and new solutions]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyi analiz i informatsionnye tekhnologii*, 2021, no. 1, pp. 33-52.
26. Polunin, A. A., Yandashevskaya E. A. Ispol'zovanie apparata svertochnykh neironnykh setei dlya stegoanaliza tsifrovykh izobrazhenii [Using convolutional neural networks for steganalysis of digital images]. *Trudy Instituta sistemnogo programmirovaniya RAN*, 2020, vol. 32, no 4, pp. 155-164.
27. Ryabko B., Ryabko D. Information-theoretic approach to steganographic systems. *IEEE International Symposium on Information Theory.* Nice, France, 2007, pp. 2461-2464.
28. Vizil'ter Yu. V., Zheltov S. Yu., Bondarenko A. V., Ososkov M.V., Morzhin A.V. *Obrabotka i analiz izobrazhenii v zadachakh mashinnogo zreniya: kurs lektsii i prakticheskikh zanyatii* [Image processing and analysis in computer vision problems: a course of lectures and practical exercises]. Moscow, Fizmatkniga, 2010, 672 p.
29. Said A. Introduction to Arithmetic Coding – Theory and Practice. *Hewlett-Packard Laboratories Report*, USA, 2004, p. 64.
30. *Baza BOSS izobrazhenii dlya issledovanii metodov steganografii* [BOSS Image Database for Steganography Research], available at: `http://agents.fel.cvut.cz/boss/index.php?mode=VIEW&tmpl=about` (accessed 26.03.2024).
31. Mahasree M. Improved Reversible Data Hiding in Medical images using Interpolation and Threshold based Embedding Strategy. *International Journal of Emerging Trends in Engineering Research*, 2020, vol. 8, pp. 3495-3501.
32. Lu T.-C., Lin M.-C., Huang C.-C., Deng K.-M. Reversible Data Hiding Based on Image Interpolation with a Secret Message Reduction Strategy. *International Journal of Computer Software Engineering*, 2016, vol. 1, pp.124-130.
33. Eltysheva K., Fionov A. Stegosystem construction on the basis of statistical structure of covertext. *XII International Symposium on Problems of Redundancy*, St.-Petersburg, 26-30 May, 2009, pp. 180-185.
34. Ryabko B. Ya., Fionov A. N. Effektivnyi metod adaptivnogo arifmeticheskogo kodirovaniya dlya istochnikov s bol'shimi alfavitami [An Efficient Adaptive Arithmetic Coding Method for Sources with Large Alphabets]. *Problemy peredachi informatsii*, 1999, vol. 35, no. 4, pp. 1-14.
35. Zhilkin M. Yu. *Teoretiko-informatsionnye metody stegoanaliza graficheskikh dannykh* [Information-theoretic methods for steganalysis of graphical data]. Ph. D. thesis. Novosibirsk, 2009, 153 p.

Ekaterina Yu.Merzlyakova

PhD (Engineering), Docent of the Department of Applied Mathematics and Cybernetics, Siberian State University of Telecommunications and Informatition Science (SibSUTIS, 630102, Novosibirsk, Kirova str., 86), tel. +7 383 2698 272, e-mail: `katerina.artist@yandex.ru`, ORCID ID: 0000-0001-6847-5588.