УДК 519.6
doi:10.31799/1684-8853-2023-2-27-38
EDN: DXURBZ

Научные статьи
Articles

Стеганоанализ изображений, модифицированных алгоритмом Bit Plane Complexity Segmentation

Р. А. Солодуха[a], *канд. техн. наук, доцент, orcid.org/0000-0002-3878-4221, standartal@list.ru*
[a]*Воронежский государственный университет инженерных технологий, Революции пр., 19, Воронеж, 394036, РФ*

Введение: *стеганографический алгоритм Bit Plane Complexity Segmentation (BPCS) позволяет осуществить вложение до 50 % от размера контейнера. Вследствие этого программное обеспечение на основе BPCS может быть выбрано внутренним нарушителем для передачи информации из закрытой корпоративной или ведомственной компьютерной сети. При этом современные системы предотвращения утечки данных не имеют функционала, связанного с обнаружением цифровой стеганографии, в том числе по причине отсутствия соответствующего методического, алгоритмического и программного обеспечения.* ***Цель:*** *адаптировать для анализа BPCS существующие стеганоаналитические алгоритмы, сформировать векторы признаков. Экспериментально проверить эффективность векторов признаков и получить их парето-оптимальные комбинации.* ***Результаты:*** *выполнен трасологический анализ алгоритма BPCS, разработан вектор признаков на основе гистограмм сложности битовых плоскостей, его эффективность подтверждена численным экспериментом с использованием регрессионной модели машинного обучения в среде MatLab. Для обеспечения воспроизводимости эксперимента датасеты и программный код представлены в Kaggle. На основе экспериментальных данных рассчитаны базовые метрики результативности машинного обучения комбинаций векторов признаков для BPCS-стеганоанализа. Получены оптимальные по Парето комбинации векторов признаков.* ***Практическая значимость:*** *показана зависимость ошибки регрессии для комбинаций векторов признаков различной размерности для BPCS-стеганоанализа. С помощью полученных оценок аналитик может варьировать достоверность/размерность векторов признаков в зависимости от доступных вычислительных мощностей и размера обучающего множества.*

Ключевые слова *— стеганоанализ, вектор признаков, BPCS-стеганография, предотвращение утечки данных, стеганографический канал, машинное обучение, машина опорных векторов, регрессия.*

Для цитирования: Солодуха Р. А. Стеганоанализ изображений, модифицированных алгоритмом Bit Plane Complexity Segmentation. *Информационно-управляющие системы*, 2023, № 2, с. 27–38. doi:10.31799/1684-8853-2023-2-27-38, EDN: DXURBZ
For citation: Solodukha R. A. Steganalysis of Bit Plane Complexity Segmentation algorithm. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 2, pp. 27–38 (In Russian). doi:10.31799/1684-8853-2023-2-27-38, EDN: DXURBZ

Введение

Доступность стеганографических программных продуктов (обзор приведен в работе [1]) позволяет без труда осуществлять скрытый обмен информацией контрагентам с минимальными познаниями в области информационных технологий. Одним из противоправных направлений использования стеганографии является передача информации ограниченного доступа из ведомственной/корпоративной компьютерной сети.

Скрытый канал передачи данных может быть организован различными способами [2]. Методы, основанные на сетевых протоколах [3, 4], достаточно сложны для применения, требуют знаний относительно организации сети, имеют низкую пропускную способность [5]. При этом использование файловой стеганографии под силу любому пользователю компьютера. Наиболее популярными и простыми в использовании контейнерами для цифровой стеганографии являются изображения [6, 7], аудио- [8] и видеофайлы [9].

Передача файлов «наружу» не представляет труда, так как традиционный и пока незаменимый в большинстве бизнес-процессов сервис электронной почты доступен даже в ведомственных компьютерных сетях [10]. По данным, приведенным в «Исследовании уровня информационной безопасности в компаниях России и СНГ за 2020 год» компании «СёрчИнформ» (https://static.searchinform.ru/uploads/sites/1/2022/05/issledovaniya-2021.pdf), именно электронная почта является «самым популярным каналом для слива данных в компаниях — на них приходится 45 % утечек в России и 41 % в СНГ».

Следует отметить, что исторически основным направлением защиты компьютерных сетей было противодействие внешним угрозам при том, что внутренние утечки информации труднее предотвратить [11]. Актуальность противодействия внутренним угрозам, связанным с организацией скрытых каналов, нашла отражение в ГОСТ Р 53113.1-2008 «Защита информационных технологий и автоматизированных систем от угроз информационной безопасности, реализуемых с использованием скрытых каналов». В практическом аспекте это привело к появлению на рынке информационной безопасности

систем предотвращения утечек данных (DLP — Data Leakage Prevention), одной из функций которых является анализ трафика, выходящего за пределы сети.

Теоретически DLP-системы способны выявить структурную файловую стеганографию и вложения, совершенные программным обеспечением, оставляющим сигнатуру [10]. Слепой (универсальный) стеганоанализ [12] цифровой стеганографии несравнимо сложнее. Невзирая на значительное количество методов стеганоанализа [12], данный функционал в DLP-системах не заявлен [13]. На наш взгляд, это связано как с отсутствием спроса (непонимание заказчиками серьезности угрозы), так и с отсутствием методического и алгоритмического обеспечения, сложностью технической реализации проверки и принятия решения в онлайн-режиме. Следует отметить, что ряд производителей DLP, например Forcepoint (https://www.forcepoint.com/blog/insights/stop-pictures-hiding-malicious-content) и Forta (https://www.clearswift.com/resources/datasheets/anti-steganography-combating-external-threats-data-loss-images), решают проблему стеганографических каналов путем уничтожения вложения.

Следует отметить, что работы в направлении обнаружения стеганоконтента ведет компания McAfee. Веб-приложение Steganography Analysis Tool (https://www.mcafee.com/enterprise/ru-ru/downloads/free-tools/steganography.html) позволяет проанализировать графический файл на наличие стеганографии. Размер предполагаемого вложения отображается графически, также приводится степень достоверности анализа. Загрузка файлов доступна только в ручном режиме, сведения о включении данного функционала в продукты McAfee отсутствуют.

Можно предположить, что для инсайдера, пользующегося контейнерами-изображениями, наиболее важной характеристикой стеганографического канала является его пропускная способность, так как большой поток исходящих изображений более подозрителен и труднообъясним, нежели разовая передача изображения низкого качества. Таким образом, из средств цифровой стеганографии наиболее вероятно использование BPCS- (Bit Plane Complexity Segmentation) или LSB-стеганографии (Least Significant Bits) (последней с глубиной в несколько бит, приводящей к визуальным искажениям в изображении).

На момент написания статьи в свободном доступе находится несколько программных реализаций BPCS-стеганографии: opensource-решения на GitHub под Python (https://github.com/mobeets/bpcs) или Java (https://github.com/inmank/BPCS-Stegnography) и Windows-приложение (http://datahide.org/BPCSe/QtechHV-download-e.html), что делает ее вполне доступной для использования людьми с небольшими познаниями в IT-сфере.

При этом одним из основных критериев функционирования DLP-систем является быстродействие, т. е. алгоритмы обнаружения должны работать в потоковом режиме и быть относительно несложными. Другими словами, необходимо найти баланс между достоверностью и ресурсоемкостью обнаружения.

Целью данной работы является анализ результативности стеганоаналитических векторов признаков и выявление их парето-оптимальных комбинаций при атаке на основании известной стеганопрограммы применительно к задаче обнаружения BPCS-стеганографии в ведомственных компьютерных сетях. Также предлагается собственный вектор признаков.

BPCS-стеганография

Суть алгоритма Bit Plane Complexity Segmentation [14], предложенного в 1998 г. группой из Технологического института Кюсю под руководством Еийи Кавагучи (Eiji Kawaguchi), в том, что контейнер разбивается на информативные и шумоподобные блоки в каждой битовой плоскости, затем шумоподобные блоки заменяются стегановложением. Если стеганографические методы семейства LSB позволяют внедрить 10–15 % от размеров контейнера, то BPCS поднимает эту планку до 50 % при сопоставимом уровне искажений контейнера. Эта техника использует свойство человеческого зрения не различать графические объекты в сложных бинарных шаблонах.

Для BPSC-стеганографии битовые плоскости представляются кодом Грея. Авторы называют это переходом от Pure Binary Coded (PBC) bit planes к Canonical Gray Coded (CGC) bit planes. Это связано с тем, что хотя PBC-плоскости обеспечивают больше места для стегановложения, CGC-плоскости менее подвержены эффекту Hamming cliff, когда небольшое изменение в значении цвета пикселя приводит к существенным изменениям в его битовом представлении. Например, $127_{10} \rightarrow 0111111_2$, а $128_{10} \rightarrow 10000000_2$.

Для оценки шумоподобности блока используется несколько метрик. Базовая метрика основывается на black-and-white border (BWB) complexity — сложности черно-белых границ и представляет собой суммарное количество переходов между 0 и 1 по строкам и столбцам блока. Например, один ноль в окружении единиц дает BWB–сложность 4. Сложность блока оценивается приведенной величиной

$$a = \frac{b}{2 \cdot n \cdot (n-1)},$$

$$b = \sum_{i=1}^{n-1} \sum_{j=1}^{n} \left| I_{i,j} - I_{i+1,j} \right| + \sum_{i=1}^{n} \sum_{j=1}^{n-1} \left| I_{i,j} - I_{i,j+1} \right|,$$

где n — размер блока; $I_{i,j}$ — значение элемента блока.

Если ноли и единицы в блоке расположены с определенной периодичностью в строках или столбцах, то такие блоки для внедрения не подходят, но BWB их принимает за сложные. Поэтому в модификации [15] используют еще две метрики: иррегулярность последовательностей (Run-Length Irregularity — RLI) и шумность границ (Border Noisiness — BN).

Run-Length Irregularity основывается на гистограмме $\mathbf{H} = \{h_i\}$ длин непрерывных серий нолей и единиц по строкам или по столбцам, где i — размер серии. RLI вычисляется по формуле Шеннона и показывает неравномерность распределения длин серий:

$$h = -\sum_{i=1}^{n} h_i \cdot \log_2 p_i,\ p_i = \frac{h_i}{\sum_{j=1}^{n} h_j},$$

где n — максимальная длина серии. Для квадратных блоков размером $N \times N$, где r_i — i-я строка, а c_j — j-й столбец, вводится метрика β:

$$\beta = \min\left\{\overline{\mathbf{H}(r)},\ \overline{\mathbf{H}(c)}\right\},\ \mathbf{H}(r) = \left\{h(r_0),\ \ldots,\ h(r_{N-1})\right\},$$
$$\mathbf{H}(c) = \left\{h(c_0),\ \ldots,\ h(c_{N-1})\right\},$$

где $\overline{\mathbf{X}}$ — среднее значение $\mathbf{X} = \{x_0, \ldots, x_{N-1}\}$.

Run-Length Irregularity может работать некорректно, принимая простые блоки за сложные, так как не способен оценить сходство смежных строк или столбцов (рис. 1).

Если информация будет скрываться на границе шумовых и информационных областей, то после внедрения общая шумность увеличится, и искажения могут стать существенными для восприятия. Шумность границ — мера сложности, показывающая, как распределены пиксели по блоку. Вычисляется на основе различий между смежными пиксельными последовательностями. Для квадратных блоков размером $N \times N$, где r_i — i-я строка, а c_j — j-й столбец, вводится метрика γ:

$$\gamma = \frac{\min\left\{E\left(\mathbf{P}(r)\right),\ E\left(\mathbf{P}(c)\right)\right\}}{N},$$

$$\mathbf{P}(r) = \left\{p\left(r_0 \oplus r_1\right),\ \ldots,\ p\left(r_{N-2} \oplus r_{N-1}\right)\right\},$$

$$\mathbf{P}(c) = \left\{p\left(c_0 \oplus c_1\right),\ \ldots,\ p\left(c_{N-2} \oplus c_{N-1}\right)\right\},$$

$$E(\mathbf{X}) = \overline{\mathbf{X}}\left(\frac{1 - V(\mathbf{X})}{\max\left\{V(\mathbf{X})\right\}}\right),$$

где $\overline{\mathbf{X}}$ — среднее значение $\mathbf{X} = \{x_0, \ldots, x_{N-1}\}$; $V(\mathbf{X})$ — дисперсия; $\oplus$ — XOR; $p(\mathbf{X})$ — количество единиц в бинарной последовательности.

Битовая последовательность для встраивания сообщения организуется в блоки, которыми заменяются блоки изображения. Однако блоки замены могут не обладать должной сложностью и вызвать значительные искажения в контейнере. В этом случае над блоком производят операцию конъюгации. Конъюгация представляет собой побитовый XOR блока изображения **P** с шаблоном **Wc**. Блок до — **P** и после конъюгации — $\mathbf{P}^*$ соотносятся как один к одному; справедливы следующие соотношения: $\mathbf{P}^* = \mathbf{P} \oplus \mathbf{Wc}$, $(\mathbf{P}^*)^* = \mathbf{P}$ (рис. 2).

При этом самое важное то, что сложность изображения до и после конъюгации в сумме дают 1: $\alpha(\mathbf{P}^*) = 1 - \alpha(\mathbf{P})$.

Таким образом, если сложность внедряемого блока меньше порогового, то после конъюгации она симметрично отразится от 0,5, например: $\alpha(\mathbf{P}) = 0{,}2 \Rightarrow \alpha(\mathbf{P}^*) = 0{,}8$.

Алгоритм BPCS:

1. Преобразование изображения-контейнера из простого двоичного кодирования (PBC) в код Грея (CGC).

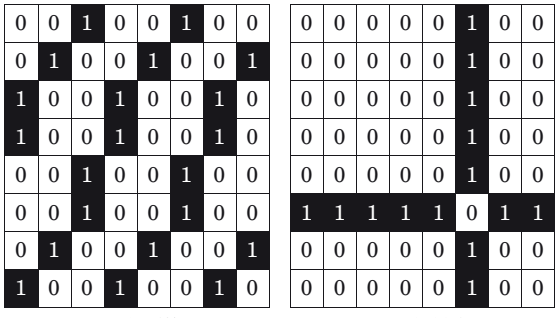
2. Разбиение изображения на битовые плоскости (изображение с глубиной цвета будет дезагрегировано в 24 плоскости).

3. Сегментирование каждой плоскости на блоки 8×8 (**P**).

4. Классификация (с помощью α, β, γ) блоков **P** на информационные и шумоподобные.

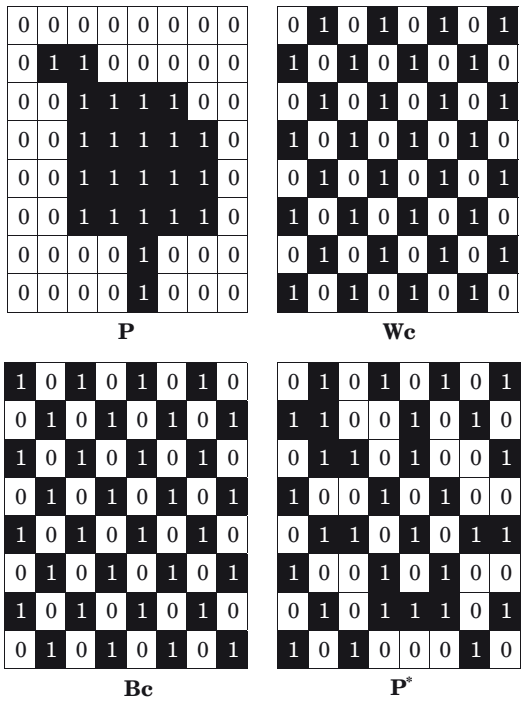
5. Представление стегановложения в виде бинарных блоков (**S**) 8×8 (возможны предварительные операции шифрования и архивации).

6. Классификация блоков **S** по сложности. В случае неудовлетворительной сложности — конъюгация блока **S**. Данный факт необходимо



■ ***Рис. 1.*** Пример RLI-оценки блоков

■ ***Fig. 1.*** Example of blocks RLI-estimations



■ ***Рис. 2.*** Пример сопоставления **P**, **Wc**, **Bc**, **P***
■ ***Fig. 2.*** Example of **P**, **Wc**, **Bc**, **P*** matching

сохранить в «карте конъюгаций», чтобы провести еще одну конъюгацию при извлечении $\mathbf{S}^*$.

7. Замена шумоподобных блоков **P** блоками вложения **S** и $\mathbf{S}^*$, блоками «карты конъюгаций».

8. Преобразование контейнера из CGC в PBC.

В качестве варьируемых параметров алгоритма выступают пороговые значения α, β, γ, размер блока, пароль шифра, последовательность замены блоков.

Трасологический анализ BPCS

Под трасологическим анализом понимается анализ изменений контейнера, модифицированного стеганографическим алгоритмом. Пример для файла JPEG приведен в работе [16].

В качестве стеганопрограммы выбрана бесплатная реализация BPCS-алгоритма Qtech-HV02 с фиксированным порогом сложности, равным 40. Максимальный размер вложения рассчитывается для каждого файла и доступен для считывания, что позволяет автоматизировать процесс заполнения стеганоконтейнеров.

Обозначим пустой контейнер размером $M \times N$ как **C**, а заполненный — как **S**. Оценивается модуль максимального отклонения max_diff = $= \max|\mathbf{C}_i - \mathbf{S}_i|$ и усредненное искажение в виде суммы абсолютных значений искажений, приведенных к количеству пикселей:

$$\text{mean} = \frac{1}{M \cdot N} \sum_{i=1}^{M \cdot N} |\mathbf{S}_i - \mathbf{C}_i|.$$

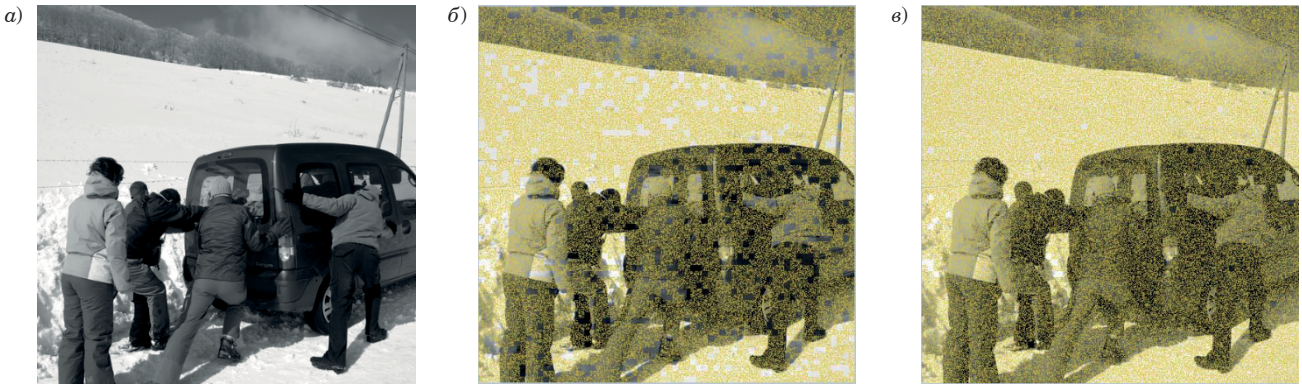
Искажения, вносимые в файл, в зависимости от размера вложения в процентах от максимально возможного приведены в табл. 1 (файл № 7 из коллекции BOSSbase 1.01 (http://dde.binghamton.edu/download/ImageDB/BOSSbase_1.01.zip)), откуда видно, что модификация выполняется «послойно-поцветно», т. е. сначала заполняется LSB_1 красного, затем зеленого, затем синего. Далее переход на LSB_2 и так до LSB_5. Под LSB_n понимается номер битового среза: $LSB_1 - 2^0$, $LSB_2 - 2^1$, $LSB_3 - 2^2$, $LSB_4 - 2^3$, $LSB_5 - 2^4$.

Изменения в плоскости красного цвета при уровнях заполнения 0, 9, 29 % от максимально возможного иллюстрирует рис. 3 (изображения получены с помощью программы WinMerge (https://winmerge.org/downloads)). Можно наблюдать, как алгоритм стремится сначала заполнить плоскость младшего бита во всех цветах, затем переходит к плоскости следующего по старшинству бита. Обобщенно можно сказать, что BPCS осуществляет псевдослучайное изменение битов в каждой битовой плоскости, независимо.

Визуальный анализ гистограмм изображения до и после BPCS-преобразования показывает достаточно сильные искажения (рис. 4, файл № 9 из коллекции BOSSbase 1.01), что предопределяет попытку гистограммной атаки по аналогии с методом Pair of Values.

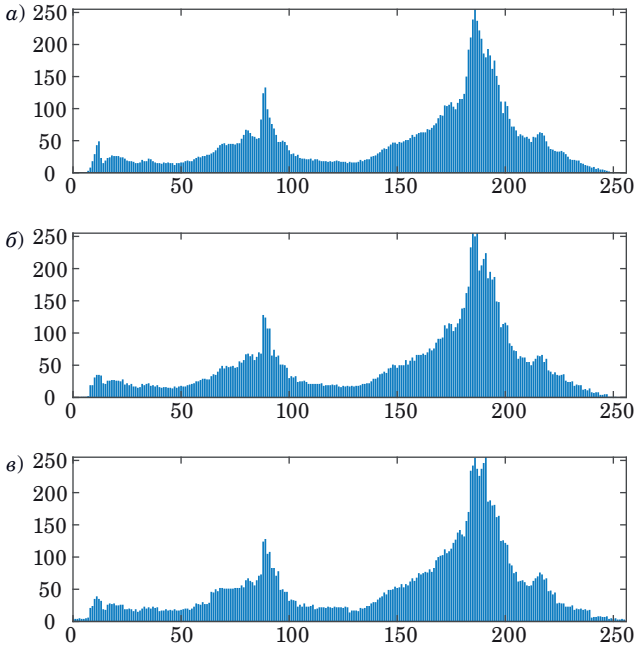
■ ***Таблица 1.*** Метрики искажения контейнера при реализации стегановложения
■ ***Table 1.*** Metrics of image distortion after payload

%	color	max_diff	mean	%	color	max_diff	mean
9	R	1	0,4	59	R	7	1,2
	G	0	0		G	3	1,2
	B	0	0		B	3	1,2
19	R	1	0,5	69	R	7	2
	G	1	0,4		G	7	1,7
	B	0	0		B	3	1,2
29	R	1	0,5	79	R	15	2,1
	G	1	0,5		G	7	2
	B	1	0,4		B	7	2
39	R	3	1	89	R	15	3,3
	G	1	0,5		G	15	3,3
	B	1	0,5		B	15	2,3
49	R	3	1,2	99	R	31	4,4
	G	3	1,1		G	31	4,4
	B	1	0,5		B	31	4



■ ***Рис. 3.*** Визуализация изменений исходного изображения (*а*) при вложении 9 % (*б*) и 29 % (*в*) от максимально возможного при модификации программой Qtech-HV02

■ ***Fig. 3.*** Visualization of original image (*а*) changes after payload 9% (*б*), 29% (*в*) of maximum capacity using Qtech-HV02



■ ***Рис. 4.*** Гистограммы исходного изображения (*а*) с вложением 49 % (*б*) и 99 % (*в*)

■ ***Fig. 4.*** Histogram of original image (*а*) with 49% (*б*) and 99% (*в*) payloaded image

В работе [17], одним из авторов которой является сам Еийи Кавагучи, описана уязвимость BPCS. При построении гистограммы сложности блоков (Complexity Histogram — HC) у заполненного контейнера наблюдается провал в области выбранного порога сложности. К настоящему времени данная уязвимость осталась на уровне идеи, работы по ее реализации в свободном доступе отсутствуют.

Исходя из вышеизложенного можно определить следующие направления анализа BPCS.

1. Использование признаков, полученных стеганоалгоритмами, не имеющими возможности адаптации к глубине искажения. В эту группу можно отнести Gradient Paths (GP) [18, 19], так как он основан на сравнении градиентных путей в исходном и LSB_n-обнуленном изображениях, и SPAM (Subtractive Pixel Adjacency Matrix) [20], прогнозирующий значение пикселя по соседям.

2. Формирование признаков с помощью одномерных детекторов LSB, применяя их к каждому искажаемому битовому срезу, обнуляя биты более младших срезов. Сюда можно отнести Sample Pairs Analysis (SP) [21], Asymptotically Uniformly Most Powerful Test (AUMP) [22], Weighted Stego-Image Method (WS) [23], Triples Analysis (T) [24]. Коррелированность данных методов проверена в работе [25].

3. Формирование признаков с помощью гистограммного детектора Pairs of Values (PoVs) [26], адаптированного к анализу изменений в нескольких битовых плоскостях.

4. Формирование признаков на основе гистограммы сложности блоков (доведение описанной в [17] уязвимости до практической реализации).

Формирование векторов признаков

Модификация количественных детекторов LSB

Поскольку BPCS вносит искажения в LSB_{1-5}, то целесообразно применить одномерные количественные детекторы для каждой из пяти битовых плоскостей. Значение каждого пикселя изображения сдвигается в сторону младшего разряда с обнулением старших (функция *bitshift*(**X**, *n*) в MatLab, **X** — матрица изображения, *n* — размер сдвига).

Модификация PoVs

Pair of Values является одним из первых стеганоаналитических методов, выявляющих LSB-стеганографию на основе гистограммной атаки.

При этом в исходном виде PoVs отслеживает переходы между младшим четным и старшим нечетным значениями гистограммы, что характерно для LSB replacement. В случае с BPCS изменения затрагивают до пяти битовых плоскостей. Механизм вычисления элементов гистограммы, подлежащих попарному сравнению при вложении в разные плоскости, и диаграмма переходов представлены на рис. 5.

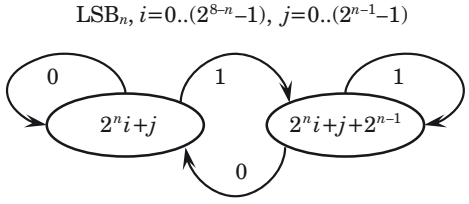
Также PoVs предполагается использовать в модификации без применения «функции активации по хи-квадрат», так как это преобразование практически бинаризирует результаты, делая их нулевыми или близкими к единице, что приводит к нецелесообразности применения регрессии для их анализа. В итоге с помощью модифицированного PoVs формируется 5D вектор признаков.

Формирование вектора признаков на основе гистограммы сложности

Пусть $\mathbf{H} = \{h(x)\}$ — гистограмма сложности изображения. Визуальный анализ гистограмм сложности показывает, что изменения затрагивают область (0,325…0,45), характерный пример — рис. 6. Очевидно, что для распознавания вложения наиболее важна область перегиба, т. е. значения $h(0{,}35)$, $h(0{,}375)$, $h(0{,}4)$. Однако области справа и слева от перегиба также оказывают влияние на распознавание. Предварительные эксперименты показали, что это влияние целесообразно учитывать в агрегированном виде, как площадей $S1$, $S2$ под кривой на участках (0,25…0,325) и (0,4…0,475) соответственно. Таким образом, получаем 25D вектор признаков:

$$HC = \{HC_1, \ldots, HC_5\}, HC_i = \{S1_i, h_i(0{,}35),$$

$$h_i(0{,}375), h_i(0{,}4), S2_i\}, i = 1\ldots5.$$



LSB$_1$: 0↔1, 2↔3, 4↔5, … , 252↔253, 254↔255;

LSB$_2$: 0↔2, 1↔3, 4↔6, 5↔7, … , 252↔254, 253↔255;

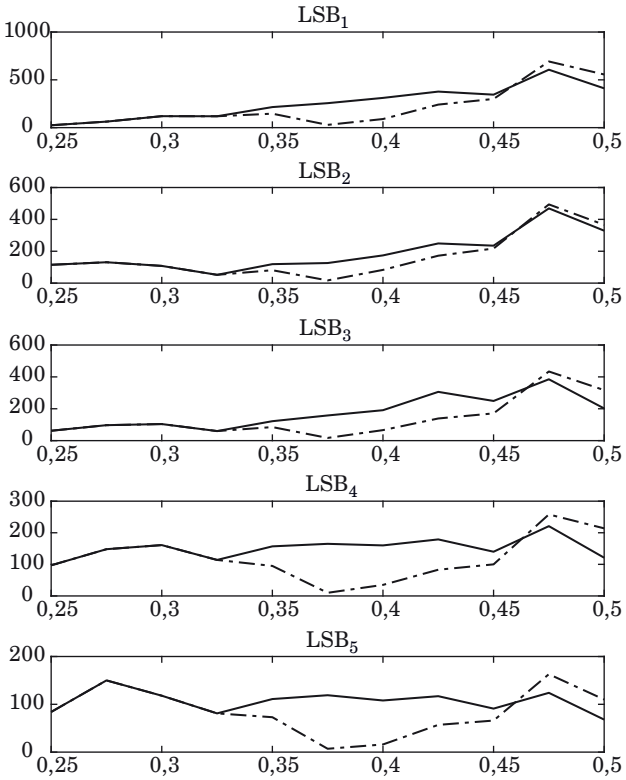
LSB$_3$: 0↔4, 1↔5, 2↔6, 3↔7, 8↔12, … , 250↔254, 251↔255;

LSB$_4$: 0↔8, 1↔9, 2↔10, 3↔11, 4↔12, 5↔13, … , 246↔254, 247↔255;

LSB$_5$: 0↔16, 1↔17, 2↔18, 3↔19, 4↔20, 5↔21, 6↔22, … , 238↔254, 239↔255.

■ ***Рис. 5.*** Диаграмма переходов в первой–четвертой младших битовых плоскостях

■ ***Fig. 5.*** Transition diagram of first–fourth least significant bit plains



■ ***Рис. 6.*** Гистограммы сложности блоков по битовым плоскостям (сплошная линия — без вложения, штрихпунктирная — с вложением 99 %)

■ ***Fig. 6.*** Block complexity histograms by bit plains (solod line — original image, dot-dash line — 99% payloaded image)

Экспериментальная часть

Количественные LSB-детекторы взяты на сайте Digital Data Embedding Laboratory, Binghampton University (http://dde.binghamton.edu/download/structural_lsb_detectors/), остальные алгоритмы запрограммированы самостоятельно.

В качестве источника контейнеров взяты первые 2000 файлов коллекции BOSSbase 1.01 (переконвертированы из PGM в BMP). Для анализа использовалась плоскость красного цвета. Это незначительно смещает оценку размера вложения в большую сторону, так как BPCS-алгоритм, реализованный в программе Qtech-HV02, заполняет битовые слои от младшего (LSB$_1$) к старшему (LSB$_5$) и от красного к синему.

Поскольку эксперимент посвящен атаке на основании известной стеганопрограммы, а не стеганоалгоритма [27, 28] или его имитации [29], то необходимы контейнеры, полученные с помощью конкретного стеганографического программного обеспечения. Контейнеры заполнялись автоматически (Свидетельство о регистрации программы для ЭВМ № 2022682838 от 28.11.2022) с помощью скриптов AutoIt с шагом 10 % от максимального размера вложения от 9 до 99 %, выборка составила 22 000 контейнеров (https://www.kaggle.com/datasets/romansolodukha/bpcs-qtech).

В качестве прогнозной модели выбрана регрессионная, де-факто являющаяся стандартом в количественном стеганоанализе [30]. Авторы разделяют мнение, изложенное в работе [31], что с «точки зрения успеха в решении задач машинного обучения качество данных, как правило, намного важнее качества алгоритма обучения», в связи с чем использован стандартный регрессор на базе машины опорных векторов (SVM) [32] из среды машинного обучения MatLab Regression Learner, с настройками по умолчанию. В зависимости от количества признаков (D) изменялся только масштаб гауссова ядра машины опорных векторов:

$$\sqrt{D} \mid D \geq 20;\ \frac{1}{4}\sqrt{D} \mid D < 20.$$

Выборка делилась на обучающую/тестирующую в соотношении 50/50. В качестве метрик результативности машинного обучения [33] использованы коэффициент детерминации (R-Squared, R^2) и среднеквадратическая ошибка (RMSE).

Результаты экспериментов с модифицированными детекторами LSB приведены в табл. 2. Видно, что с ростом количества признаков распознавание улучшается. LSB_5 практически не дает прироста результативности, так как модифицируется лишь в 1/11 части выборки. Модифицированный PoVs показал наихудшие результаты, использовать его далее нецелесообразно.

Результаты экспериментов с векторами признаков без разделения на битовые плоскости приведены в табл. 3. SPAM показал лучшие результаты, GP не достиг уровня LSB-детекторов. При исключении из комбинации GP точность повышается, поэтому из дальнейших экспериментов GP исключен.

Для определения вклада каждого вектора признаков (HC, AUMP, SP, WS, T) в точность регрессии проверены комбинации из четырех и более векторов признаков (табл. 4) путем последовательного исключения каждого вектора. Анализ табл. 4 показывает, что наибольший вклад вносит HC, несущественно влияние AUMP

■ ***Таблица 2.*** Результаты применения векторов признаков с разделением изображения на битовые плоскости

■ ***Table 2.*** Results of applying feature vectors with image division into bit planes

Детектор LSB	Метрика	Плоскость LSB				
		LSB_1 (1D)	$LSB_{1,2}$ (2D)	LSB_{1-3} (3D)	LSB_{1-4} (4D)	LSB_{1-5} (5D)
PoVs	RMSE	31	29	28,5	27,4	26,5
	R-Squared	0,03	0,11	0,19	0,24	0,3
AUMP	RMSE	26,2	17,9	10,7	8	7,8
	R-Squared	0,3	0,67	0,88	0,93	0,94
SP	RMSE	27,4	17,88	12,3	10,3	9,91
	R-Squared	0,25	0,68	0,85	0,89	0,9
WS	RMSE	26,2	18	12,5	10,3	9,9
	R-Squared	0,31	0,67	0,84	0,89	0,9
T	RMSE	29,2	19,7	13	10,62	10,6
	R-Squared	0,15	0,61	0,83	0,89	0,89

■ ***Таблица 3.*** Результаты применения векторов признаков без разделения изображения на битовые плоскости

■ ***Table 3.*** Results of applying feature vectors without dividing the image into bit planes

Метрика	Вектор признаков		
	SPAM (686D)	HC (25D)	GP (16D)
RMSE	6,3	9,1	13
R-Squared	0,96	0,92	0,8

■ ***Таблица 4.*** Результаты применения комбинаций более трех векторов признаков

■ ***Table 4.*** Results of applying combinations of more than three feature vectors

Метрика	Комбинация векторов признаков						
	HC+AUMP+SP+WS+T+GP (61D)	HC+AUMP+SP+WS+T (45D)	HC+AUMP+SP+WS (40D)	HC+AUMP+SP+T (40D)	HC+AUMP+WS+T (40D)	HC+SP+WS+T (40D)	AUMP+SP+WS+T (20D)
RMSE	5,8	5,6	5,62	5,73	5,62	5,77	6,8
R-Squared	0,96	0,97	0,97	0,97	0,97	0,97	0,95

■ ***Таблица 5.*** Результаты применения комбинаций менее четырех векторов признаков

■ ***Table 5.*** Results of applying combinations of less than four feature vectors

Метрика	Комбинация векторов признаков			
	HC+AUMP+WS (35D)	HC+WS (30D)	AUMP+WS (10D)	HC+AUMP (30D)
RMSE	5,68	6,07	6,95	6,06
R-Squared	0,97	0,96	0,95	0,96

и WS, а влияние SP и T находится в рамках погрешности.

Далее, аналогичным образом, проверены комбинации из трех и менее векторов признаков (табл. 5). Подтверждается лучшая прогностическая способность HC по сравнению с AUMP и WS.

Для целей статьи важно, что ряд комбинаций векторов признаков табл. 4, 5 обеспечивают ошибку регрессии на уровне и менее, чем SPAM (686D).

Получение парето-оптимальных решений

Если принять в качестве допущения, что ресурсоемкость каждого элемента вектора признаков одинакова, то можно сформулировать задачу оптимизации в следующем виде [34].

Пусть $\mathbf{F} = (f_1, \ldots, f_N)$ — комбинации векторов признаков, $N = \sum_{k=1}^{M} \frac{M!}{k!(M-k)!}$, M — количество векторов признаков.

Введем векторный критерий $\mathbf{Q} = (q_1, q_2)$, где $q_1 \to \min$ имеет смысл ошибки обнаружения, $q_2 \to \min$ — ресурсоемкости. В качестве ограничения установим требуемый уровень достоверности.

Введем правило, позволяющее оценивать комбинации векторов признаков — безусловный критерий предпочтения [35, 36]. Вариант f_2 лучше варианта f_1 ($f_1 < f_2$) в смысле векторного критерия $\mathbf{Q}$, если $\forall i\ q_i(f_1) \geq q_i(f_2)$ и хотя бы одно неравенство строгое. Если $\forall i\ q_i(f_1) = q_i(f_2)$, то вариант f_2 эквивалентен варианту f_1. Из всего множества комбинаций векторов признаков безусловный критерий предпочтения позволяет выделить подмножество решений, определяющих множество Парето.

В качестве целевых функций целесообразно выбрать среднеквадратическую ошибку и

■ ***Таблица 6.*** Парето-оптимальные комбинации наборов векторов признаков

■ ***Table 6.*** Pareto-optimal combinations of feature vectors sets

Целевая функция	Комбинация векторов признаков			
	HC+AUMP+WS	HC+AUMP+SP+WS	HC+AUMP+WS+T	HC+AUMP+SP+WS+T
q_1	5,68	5,62	5,62	5,6
q_2	35	40	40	45

количество признаков: $q_1 = RMSE(f_i) \to \min$, $q_2 = card(f_i) \to \min$, ограничение на достоверность ввести посредством коэффициента детерминации $\gamma \geq R^2$. Примем $\gamma \geq 0{,}97$, тогда оптимальными по Парето являются комбинации табл. 6.

На основании парето-оптимальных комбинаций аналитик может ввести весовые коэффициенты (в зависимости от задачи, отдав приоритет точности или достоверности) на элементы $\mathbf{Q}$ и получить скалярный критерий выбора.

Заключение

На основе трасологического анализа алгоритма BPCS выявлены перспективные векторы признаков для стеганоанализа с применением машинного обучения. Выполнена модификация детекторов LSB для целей анализа BPCS. Разработан 25D вектор признаков, базирующийся на гистограмме сложности блоков битовых плоскостей.

Осуществлен эксперимент по эффективности определения размера вложения, выполненного стеганографической программой Qtech-HV02. Использована технология машинного обучения, реализованная в среде MatLab — SVM-регрессия. Для каждого вектора признаков или их комбинации получены оценки коэффициента детерминации и среднеквадратической ошибки.

Найдены комбинации векторов признаков (от 30D), обеспечивающие меньшую ошибку регрессии, чем SPAM (686D). Сокращение количества признаков позволяет уменьшить ресурсоемкость и размер обучающей выборки, что важно для выявления стеганографии в рамках стеганоанали-

тических подсистем DLP-систем. Полученные парето-оптимальные комбинации векторов признаков позволяют настраивать соотношение достоверность/ресурсоемкость процессов противодействия утечки данных. Разработанный вектор признаков HC (25D) на основе гистограммы сложности битовых плоскостей входит во все парето-оптимальные комбинации.

Таким образом, материалы статьи развивают математическое и алгоритмическое обеспечение DLP-систем применительно к задаче обнаружения BPCS-стеганографии.

В дальнейших исследованиях по данной тематике предполагается провести аналогичный программный эксперимент на полноцветных изображениях с применением ансамблевых методов.

Литература

1. **Герлинг Е. Ю., Ахрамеева К. А.** Обзор современного программного обеспечения, использующего методы стеганографии. *Экономика и качество систем связи*, 2019, № 3 (13), с. 51–58. http://nirit.org/wp-content/uploads/2020/01/51-58.pdf (дата обращения: 27.04.2023).
2. **Орлов В. В., Алексеев А. П.** Активная стеганография в сетях TCP/IP. *Инфокоммуникационные технологии*, 2009, т. 7, № 2, с. 73–78.
3. **Закалкин П. В., Иванов С. А., Вершенник Е. В., Кирьянов А. В.** Способ маскирования передаваемой информации. *Тр. ИСП РАН*, 2020, т. 32, вып. 6, с. 111–126. doi:10.15514/ISPRAS–2020–32(6)–9
4. **Карпухин Е. О.** *Методы скрытой передачи информации*. М., Горячая линия-Телеком, 2020. 80 с.
5. **Шипулин П. М., Козин В. В., Шниперов А. Н.** Метод организации скрытого канала передачи информации на основе протокола потоковой передачи данных. *Научно-технический вестник информационных технологий, механики и оптики*, 2018, № 5, с. 834–842. doi:10.17586/2226-1494-2018-18-5-834-842
6. **Пономарев И. В., Строкин Д. И.** Стеганографические методы встраивания и обнаружения сокрытых сообщений, использующие gif-изображения в качестве файлов-контейнеров. *Изв. Алтайского государственного университета*, 2022, № 1 (123), с. 112–115. doi:10.14258/izvasu(2022)1-18
7. **Мельман А. С., Петров П. О., Шелупанов А. А., Аристов А. В., Похолков Ю. П.** Встраивание информации в jpeg-изображения с маскировкой искажений в частотной области. *Докл. Томского государственного университета систем управления и радиоэлектроники*, 2020, т. 23, № 4, с. 45–50. doi:10.21293/1818-0442-2020-23-4-45-50
8. **Воронцова Н. В., Миляева И. В.** Стеганографическая защита информации. *Изв. Тульского государственного университета. Технические науки*, 2020, № 12, с. 86–95.
9. **Радаев С. В., Басов О. О., Мясин К. И., Мотиенко А. И.** Встраивание стеганографических сообщений в видеофайлы формата MPEG-4. *Экономика. Информатика*, 2018, т. 45, № 4, с. 769–781. doi:10.18413/2411-3808-2018-45-4-769-781
10. **Солодуха Р. А.** Концепция формирования системы противодействия стеганографическим каналам в компьютерных сетях органов внутренних дел. *Вестник Воронежского института МВД России*, 2021, № 1, с. 131–142.
11. **Андрианов В. И., Сивков Д. И., Юркин Д. В.** Методика внедрения системы предотвращения утечек информации (DLP) в коммерческую организацию для информационной сети с использованием больших данных. *Вестник Брянского государственного технического университета*, 2020, № 6, с. 38–48. doi:10.30987/1999-8775-2020-6-38-49
12. **Вильховский Д. Э.** Обзор методов стеганографического анализа изображений в работах зарубежных авторов. *Математические структуры и моделирование*, 2020, № 4 (56), с. 75–102. doi:10.24147/2222-8772.2020.4.75-102
13. **Gutiérrez-Cárdenas J. M.** Steganography and data loss prevention: An overlooked risk? *International Journal of Security and its Applications*, 2017, (11) 4, pp. 71–84. doi:10.14257/ijsia.2017.10.4.06
14. **Kawaguchi E., Eason R.** Principle and applications of BPCS-steganography. *Multimedia Systems and Applications*, 1998, vol. 3528, pp. 464–473.
15. **Vipul Patel, Neha Soni.** Uncompressed image steganography using BPCS: Survey and analysis. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 2013, vol. 15, iss. 4, pp. 57–64.
16. **Радаев С. В., Орлов Д. В., Басов О. О.** Комбинированный стеганографический алгоритм встраивания конфиденциальной информации в цифровые изображения формата JPEG. *Экономика. Информатика*, 2017, т. 44, № 23 (272), с. 185–192.
17. **Michiharu Niimi, Richard O. Eason, Hideki Noda, Eiji Kawaguchi.** Intensity histogram steganalysis in BPCS-steganography. *Security and Watermarking of Multimedia Contents III*, 2001, vol. 4314, pp. 555–565. doi:10.1117/12.435440
18. **Солодуха Р. А.** Статистический стеганоанализ фотореалистичных изображений с использованием градиентных путей. *Вопросы кибербезопасности*, 2022, № 1 (47), с. 26–36. doi:10.21681/2311-3456-2022-1-26-36
19. **Солодуха Р. А.** Формирование градиентных путей изображения как предварительный этап стеганоанализа. *Вестник Воронежского института МВД России*, 2020, № 1, с. 97–106.

20. **Pevný 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. doi:10.1109/TIFS.2010.2045842
21. **Dumitrescu S., Wu X., Memon D.** On steganalysis of random LSB embedding in continuous-tone images. *IEEE International Conference on Image Processing*, 2002, vol. 3, pp. 641–644.
22. **Fillatre L.** Adaptive steganalysis of Least Significant Bit replacement in grayscale natural images. *IEEE Transactions on Signal Processing*, 2012, vol. 60, no. 2, pp. 556–569. doi:10.1109/TSP.2011.2174231
23. **Ker A., Böhme R.** Revisiting weighted stego-image steganalysis. *Proc. of SPIE*, 2008, vol. 6819, pp. 681905. doi:10.1117/12.766820
24. **Ker A.** A general framework for structural steganalysis of LSB replacement. *Lecture Notes in Computer Science*, 2005, vol. 3727, pp. 296–11. doi:10.1007/11558859_22
25. **Солодуха Р. А., Перминов Г. В., Атласов И. В.** Редукция набора детекторов LSB с заданной достоверностью. *Научно-технический вестник информационных технологий, механики и оптики*, 2022, т. 22, № 1, с. 74–81. doi:10.17586/2226-1494-2022-22-1-74-81
26. **Westfeld A., Pfitzmann A.** Attacks on steganographic systems: Breaking the steganographic utilities EzStego, Jsteg, Steganos and S-Tools-and Some Lessons Learned. *Lecture Notes in Computer Science*, 2000, vol. 1768, pp. 61–76. doi:10.1007/10719724_5
27. **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 ВГУ. Сер.: Системный анализ и информационные технологии*, 2021, № 1, с. 33–52. doi:10.17308/sait.2021.1/3369
28. **Монарев А. И., Пестунов В. А.** Эффективное обнаружение стеганографически скрытой информации посредством интегрального классификатора на основе сжатия данных. *Прикладная дискретная математика*, 2018, № 40, с. 59–71. doi:10.17223/20710410/40/5
29. **Сивачев А. В.** Эффективность статистических методов стеганоанализа при обнаружении встраивания в вейвлет область изображения. *Вопросы кибербезопасности*, 2018, № 1 (25), с. 72–78. doi:10.21681/2311-3456-2018-1-72-78
30. **Mo Chen, Mehdi Boroumand, Jessica Fridrich.** Deep learning regressors for quantitative steganalysis. *Proc. IS&T Int'l. Symp. on Electronic Imaging: Media Watermarking, Security and Forensics*, 2018, pp. 160-1–160-7, doi:10.2352/ISSN.2470-1173.2018.07.MWSF-160
31. **Парасич А. В., Парасич В. А., Парасич И. В.** Формирование обучающей выборки в задачах машинного обучения. *Информационно-управляющие системы*, 2021, № 4, с. 61–70. doi:10.31799/1684-8853-2021-4-61-70
32. **Дикий Д. И.** Метод обнаружения DoS-атак на прикладном уровне в сетях «издатель-подписчик». *Информационно-управляющие системы*, 2020, № 4, с. 50–60. doi:10.31799/1684-8853-2020-4-50-60
33. **Лебедев И. С.** Адаптивное применение моделей машинного обучения на отдельных сегментах выборки в задачах регрессии и классификации. *Информационно-управляющие системы*, 2022, № 3, с. 20–30. doi:10.31799/1684-8853-2022-3-20-30
34. **Носков С. И., Рязанцев А. И.** Двухкритериальная транспортная задача. *T-Comm: Телекоммуникации и транспорт*, 2019, т. 13, № 2, с. 59–63. doi:10.24411/2072-8735-2018-10237
35. **Подиновский В. В., Ногин В. Д.** *Парето-оптимальные решения многокритериальных задач*. М., Наука, 1982. 256 с.
36. **Пименов В. И., Пименов И. В.** Анализ и визуализация данных в задачах многокритериальной оптимизации проектных решений. *Информатика и автоматизация*, 2022, № 3 (21), с. 543–571. doi:10.15622/ia.21.3.4

UDC 519.6
doi:10.31799/1684-8853-2023-2-27-38
EDN: DXURBZ

Steganalysis of Bit Plane Complexity Segmentation algorithm

R. A. Solodukha[a], PhD, Tech., Associate Professor, orcid.org/0000-0002-3878-4221, standartal@list.ru
[a]Voronezh State University of Engineering Technologies, 19, Revolucii Ave., 394036, Voronezh, Russian Federation

Introduction: Bit Plane Complexity Segmentation (BPCS) steganographic algorithm allows to payload up to 50% of the container size. Therefore, BPCS-based software is preferred by an insider to transport information from isolated corporate or departmental computer networks. At the same time, modern data leak prevention systems for corporate networks do not have a feature set related to the detection of digital steganography. One of the reasons is these systems are not provided with appropriate methodical, algorithmic and program support. **Purpose:** To develop a feature vector for BPCS steganalysis. To compare the effectiveness of steganalytical feature vectors that are adequate for our task. Using experimental data to obtain the Pareto optimal combinations of feature vectors. **Results:** We perform the tracology analysis of the BPCS algorithm. We develop the feature vector based on intensity histogram of the bit planes. We apply a regression model for machine learning procedure. Datasets are obtained by MatLab. To ensure reproducibility of the experiments the datasets and scripts are presented in Kaggle. Using experimental data we calculate the effectiveness metrics of the combinations of

feature vectors for BPCS-steganalysis. Finally, we obtain the Pareto-optimal combinations of feature vectors. **Practical relevance:** The dependence of the regression error for combinations of different dimensions feature vectors for BPCS-steganalysis is shown. With the help of the obtained estimates an analyst can vary the reliability/dimension of the feature vectors depending on the available computing power and the size of the training set.

Keywords — steganalysis, feature vector, BPCS-steganography, data leak prevention, steganography channel, machine learning, SVM, regression.

For citation: Solodukha R. A. Steganalysis of Bit Plane Complexity Segmentation algorithm. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 2, pp. 27–38 (In Russian). doi:10.31799/1684-8853-2023-2-27-38, EDN: DXURBZ

References

1. Gerling E., Ahrameeva K. The review of the modern software using steganography methods. *Ekonomika i kachestvo sistem svyazi*, 2019, no. 3 (13), pp. 51–58. Available at: http://nirit.org/wp-content/uploads/2020/01/51-58.pdf (accessed 23 April 2023) (In Russian).
2. Orlov V. V., Alekseev A. P. Active steganography in TCP/IP nets. *Infocommunication Technology*, 2009, vol. 7, no. 2, pp. 73–78 (In Russian).
3. Zakalkin P. V., Ivanov S. A., Vershennik E. V., Kir'yanov A. V. Method of masking transmitted information. *Proc. ISP RAS*, 2020, vol. 32, iss. 6, pp. 111–126 (In Russian). doi:10.15514/ISPRAS-2020-32(6)-9
4. Kharpukhin E. O. *Metody skrytoy peredachi informacii* [Methods of hidden information transfer]. Moscow, Goryachaya liniya — Telekom Publ., 2020. 80 p. (In Russian).
5. Shipulin P. M., Kozin V. V., Shniperov A. N. Covert channel technique based on streaming protocol. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics,* 2018, no. 5, pp. 834–842 (In Russian). doi:10.17586/2226-1494-2018-18-5-834-842
6. Ponomarev I. V., Strokin D. I. Steganographic methods for embedding and detecting hidden messages using GIF images as container files. *Izvestiya Altajskogo gosudarstvennogo universiteta*, 2022, no. 1 (123), pp. 112–115 (In Russian). doi:10.14258/izvasu(2022)1-18
7. Melman A. S., Petrov P. O., Shelupanov A. A., Aristov A. V., Pokholkov Y. P. Embedding information into JPEG images with distortion masking in frequency domain. *Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki*, 2020, vol. 23, no. 4, pp. 45–50 (In Russian). doi:10.21293/1818-0442-2020-23-4-45-50
8. Vorontsova N. V., Milyaeva I. V. Steganographic information protection. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki*, 2020, no. 12, pp. 86–95 (In Russian).
9. Radaev S. V., Basov O. O., Myasin K. I., Motienko A. I. Embedding steganographic messages into MPEG-4 video files. *Economics, Information Technologies*, 2018, vol. 45, no. 4, pp. 769–781 (In Russian). doi:10.18413/2411-3808-2018-45-4-769-781
10. Solodukha R. A. Conception of forming the steganographic channels counteraction system in the internal affairs computer networks. *The Bulletin of Voronezh Institute of the Ministry of Internal Affairs of Russia*, 2021, no. 1, pp. 131–142 (In Russian).
11. Andrianov V. I., Sivkov D. I., Yurkin D. V. Procedure for implementation of data leakage prevention (DLP) system to commercial company for information network using large database. *Vestnik Bryanskogo gosudarstvennogo tekhnicheskogo universiteta*, 2020, no. 6, pp. 38–48 (In Russian). doi:10.30987/1999-8775-2020-6-38-49
12. Vilkhovskiy D. E. A survey of steganalysis methods in the papers of foreign authors. *Matematicheskie struktury i modelirovanie*, 2020, no. 4 (56), pp. 75–102 (In Russian). doi:10.24147/2222-8772.2020.4.75-102
13. Gutiérrez-Cárdenas J. M. Steganography and Data Loss Prevention: An overlooked risk? *International Journal of Security and its Applications*, 2017, (11) 4, pp. 71–84. doi:10.14257/ijsia.2017.10.4.06
14. Kawaguchi E., Eason R. Principle and applications of BPCS-steganography. *Multimedia Systems and Applications*, 1998, vol. 3528, pp. 464–473.
15. Vipul Patel, Neha Soni. Uncompressed Image Steganography using BPCS: Survey and Analysis. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 2013, vol. 15, iss. 4, pp. 57–64.
16. Radaev S. V., Orlov D. V., Basov O. O. Steganographic combination algorithm of embedding the confidential information into the JPEG. *Economics, Information Technologies*, 2017, vol. 44, no. 23 (272), pp. 185–192 (In Russian).
17. Michiharu Niimi, Richard O. Eason, Hideki Noda, Eiji Kawaguchi. Intensity histogram steganalysis in BPCS-steganography. *Security and Watermarking of Multimedia Contents III*, 2001, vol. 4314, pp. 555–565. doi:10.1117/12.435440
18. Solodukha R. A. Statistical steganalysis of photorealistic Images using gradient paths. *Voprosy kiberbezopasnosti*, 2022, no. 1 (47), pp. 26–36 (In Russian). doi:10.21681/2311-3456-2022-1-26-36
19. Solodukha R. A. Statistical steganalysis of photorealistic images using gradient paths. *The bulletin of Voronezh Institute of the Ministry of Internal Affairs of Russia*, 2020, no. 1, pp. 97–106 (In Russian).
20. Pevný 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. doi:10.1109/TIFS.2010.2045842
21. Dumitrescu S., Wu X., Memon D. On steganalysis of random LSB embedding in continuous-tone images. *IEEE International Conference on Image Processing*, 2002, vol. 3, pp. 641–644.
22. Fillatre L. Adaptive steganalysis of Least Significant Bit replacement in grayscale natural images. *IEEE Transactions on Signal Processing*, 2012, vol. 60, no. 2, pp. 556–569. doi:10.1109/TSP.2011.2174231
23. Ker A., Böhme R. Revisiting weighted stego-image steganalysis. *Proc. of SPIE*, 2008, vol. 6819, p. 681905. doi:10.1117/12.766820
24. Ker A. A general framework for structural steganalysis of LSB replacement. *Lecture Notes in Computer Science*, 2005, vol. 3727, pp. 296–311. doi:10.1007/11558859_22
25. Solodukha R. A., Perminov G. V., Atlasov I. V. Reduction of LSB detectors set with definite reliability. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 1, pp. 74–81 (In Russian). doi:10.17586/2226-1494-2022-22-1-74-81
26. Westfeld A., Pfitzmann A. Attacks on steganographic systems: Breaking the steganographic utilities EzStego, Jsteg, Steganos and S-Tools-and Some Lessons Learned. *Lecture Notes in Computer Science*, 2000, vol. 1768, pp. 61–76. doi:10.1007/10719724_5
27. Sirota A. A., Dryuchenko M. A., Ivankov A. Yu. Steganalysis of digital images by means of shallow and deep machine learning: existing approaches and new solutions. *Proc. of VSU. Series: Systems Analysis and Information Technologies*, 2021, no. 1, pp. 33–52 (In Russian). doi:10.17308/sait.2021.1/3369
28. Monarev A. I., Pestunov V. A. Efficient steganography detection by means of compression-based integral classifier. *Applied Discrete Mathematics*, 2018, no. 40, pp. 59–71 (In Russian). doi:10.17223/20710410/40/5
29. Sivachev A. V. Efficiency of statistical steganalysis methods in detection embedding in wavelet domain. *Voprosy kiberbezopasnosti*, 2018, no. 1 (25), pp. 72–78 (In Russian). doi:10.21681/2311-3456-2018-1-72-78
30. Mo Chen, Mehdi Boroumand, Jessica Fridrich. Deep learning regressors for quantitative steganalysis. *Proc. IS&T Int'l. Symp. on Electronic Imaging: Media Watermarking, Security, and Forensics*, 2018, pp. 160-1–160-7. doi:10.2352/ISSN.2470-1173.2018.07.MWSF-160
31. Parasich A. V., Parasich V. A., Parasich I. V. Training set formation in machine learning tasks. Survey. *Informatsionno-upravliaiushchie sistemy [Information and Control Systems]*, 2021, no. 4, pp. 61–70 (In Russian). doi:10.31799/1684-8853-2021-4-61-70
32. Dikii D. I. DoS attack detection at application level in publish-subscribe networks. *Informatsionno-upravliaiushchie sistemy [Information and Control Systems]*, 2020, no. 4, pp. 50–60 (In Russian). doi:10.31799/1684-8853-2020-4-50-60

33. Lebedev I. S. Adaptive application of machine learning models on separate segments of a data sample in regression and classification problems. *Informatsionno-upravliaiushchie sistemy [Information and Control Systems]*, 2022, no. 3, pp. 20–30 (In Russian). doi:10.31799/1684-8853-2022-3-20-30
34. Noskov S. I., Ryazantsev A. I. Two-criteria transport problem. *T-Comm*, 2019, vol. 13, no. 2, pp. 59–63 (In Russian). doi:10.24411/2072-8735-2018-10237
35. Podinovskiy V. V., Nogin V. D. *Pareto-optimal'nye resheniya mnogokriterial'nyh zadach* [Pareto-optimal decisions of multi-criteria problems]. Moscow, Nauka Publ., 1982. 256 p. (In Russian).
36. Pimenov V., Pimenov I. Data analysis and visualization in the tasks of the project solutions multicriteria optimization. *Informatics and Automation*, 2022, vol. 21, no. 3, pp. 543–571 (In Russian). doi:10.15622/ia.21.3.4