УНИВЕРСИТЕТ ИТМО

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2022 Том 22 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2022 Vol. 22 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК УНИВЕРСИТЕТ ИТМО
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2022-22-1-74-81
УДК 519.6

Редукция набора детекторов LSB с заданной достоверностью

Роман Александрович Солодуха[1]✉, Геннадий Вадимович Перминов[2], Игорь Викторович Атласов[3]

[1] Воронежский институт МВД России, Воронеж, 394065, Российская Федерация
[2] Российский государственный университет правосудия (центральный филиал), Воронеж, 394006, Российская Федерация
[3] Московский университет МВД России имени В.Я. Кикотя, Москва, 117997, Российская Федерация

[1] standartal@list.ru✉, https://orcid.org/0000-0002-3878-4221
[2] perminovgv@mail.ru, https://orcid.org/0000-0003-4170-2861
[3] mathematic1@rambler.ru, https://orcid.org/0000-0001-6270-6787

Аннотация

Предмет исследования. Предложено решение задачи сокращения набора стеганоаналитических методов определения размера вложения в пространственную область изображения, на примере количественных детекторов Least Significant Bits (LSB) стеганографии. Предположено, что методы могут отслеживать одни и те же закономерности в контейнерах, вследствие чего результаты их работы могут коррелировать. Представлены результаты разработки и тестирования методики редукции набора методов с учетом точности и достоверности для снижения вычислительной сложности стеганоаналитической экспертизы. **Метод.** Теоретическая база предложенного решения — приближение регрессии первого рода линейной регрессией второго рода для многомерных случайных величин. Для верификации результатов выполнен численный эксперимент. В качестве источника контейнеров применена коллекция BOSSbase. Вложения реализованы с шагом 10 % путем автоматизации стеганографических программ freeware-сегмента CryptArkan и The Third Eye с помощью AutoIt. Использованы стеганоаналитические методы Weighted Stego, Sample Pairs, Triples analysis, Asymptotically Uniformly Most Powerful detection, Pair of Values. Датасеты получены в среде MATLAB, программа реализована на языке Python. Для обеспечения воспроизводимости эксперимента датасеты и программный код представлены в Kaggle. **Основные результаты.** На основе экспериментальных данных рассчитаны интервальные оценки коррелированности методов для различных размеров стегановложения. Разработана методика в составе математической модели, алгоритма реализации модели и компьютерной программы. **Практическая значимость.** Предложенную методику целесообразно применять в задачах, где необходимо учитывать точность и достоверность. Такие оценки востребованы при осуществлении экспертно-криминалистической деятельности по вопросам, допускающим вероятностные выводы. С помощью данных оценок аналитик может варьировать количество методов в зависимости от доступных вычислительных мощностей и временных рамок исследования.

Ключевые слова

стеганоанализ, редукция набора методов, достоверность, LSB, стеганоаналитическая экспертиза, регрессия

Ссылка для цитирования: 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

Reduction of LSB detectors set with definite reliability

Roman A. Solodukha[1]✉, Gennadiy V. Perminov[2], Igor V. Atlasov[3]

[1] Voronezh Institute of the Ministry of Internal Affairs of Russia, Voronezh, 394065, Russian Federation
[2] Russian State University of Justice (Central Branch), Voronezh, 394006, Russian Federation
[3] Kikot Moscow University of the Ministry of Internal Affairs of Russia, Moscow, 117997, Russian Federation

[1] standartal@list.ru✉, https://orcid.org/0000-0002-3878-4221
[2] perminovgv@mail.ru, https://orcid.org/0000-0003-4170-2861
[3] mathematic1@rambler.ru, https://orcid.org/0000-0001-6270-6787

Abstract

The article focuses on decreasing the set of steganalytical methods that determine the payload value in the image spatial domain using quantitative detectors of Least Significant Bits (LSB) steganography. It is supposed that methods can trace the same image regularities, and hence their results can correlate. The work presents the results of the development and testing of the technique for reducing the set of steganalytical methods taking into account the accuracy and reliability to the diminution of the computational complexity of steganalytical expertise. The theoretical basis of the proposed solution is the approximation of regression of the first kind by linear regression of the second kind for multivariate random variables. To verify the results, the computational experiment was performed. The payloads were implemented in 10 % increments by automating the freeware steganographic programs CryptArkan and The Third Eye with AutoIt. The steganalytical methods, such as Weighted Stego, Sample Pairs, Triples analysis, Asymptotically Uniformly Most Powerful detection, Pair of Values, were used. The datasets were built in the MATLAB environment; the program was implemented in Python. For the experiment's reproducibility, the datasets and program code are provided in Kaggle. Interval estimates of methods correlation are calculated based on experimental data for various payload values. The developed technique includes a mathematical model, an algorithm for implementing the model, and a computer program. The proposed technique can be applied in those tasks where accuracy and reliability are taken into account. One of the subject areas demanding such assessments is computer forensics dealing with expertise with probabilistic conclusions. These estimates allow the analyst to vary the number of methods depending on the available computing resources and the time frame of the research.

Keywords

steganalysis, reduction of methods set, reliability, LSB, steganalytical expertise, regression

For citation: 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

Введение

В настоящее время в области анализа цифровой стеганографии изображений разработаны десятки методов, отвечающие на вопрос о размере вложения. Стеганоаналитические методы существуют «с открытым кодом» или в компилированном виде, т. е. представляются аналитику как «черный ящик». С учетом того, что разные методы могут отслеживать одни и те же закономерности в пустых и заполненных контейнерах, поэтому в их результатах может наблюдаться коррелированность, и один метод может включать векторы признаков других методов. С учетом вычислительной сложности стеганоаналитических методов возникает необходимость отбросить коррелированные методы или методы, являющиеся частными случаями других методов.

Данная задача особенно актуальна в рамках компьютерной экспертизы по определению наличия стегановложения при известном стеганографическом алгоритме/программе или в любом исследовании, допускающем вероятностные выводы. Согласно [1] при выполнении экспертизы эксперт не должен применять все методы, так как среди них могут присутствовать дублирующие друг друга. Потому на практике этот вопрос решается исходя из достаточности метода для решения задачи с необходимой надежностью. При этом должны учитываться такие факторы, как экономия средств и времени.

Если в распоряжении эксперта имеется несколько стеганоаналитических методов, то целесообразно предварительно проверить их на коррелированность (схожесть результатов), но с учетом требуемой достоверности, градация которой облегчила бы оценку и использование выводов эксперта следствием и судом [2]. Иначе говоря, отбрасывать какой-либо метод эксперт должен в соответствии с методикой, оценивающей достоверность результатов. При этом принятие решения на основании коэффициента корреляции требует проверки гипотезы о значимости различия между коэффициентами корреляции выборки и генеральной совокупности, а затем проверки гипотез о конкретных, интересующих аналитика, значениях генеральной совокупности [3]. Также для легитимности коэффициента корреляции результаты работы методов должны иметь совместное нормальное распределение, что накладывает дополнительное условие на выборку.

Возникает задача разработки методики редукции множества стеганоаналитических методов с возможностью априорной установки схожести методов с заданным уровнем достоверности при отсутствии требований совместного нормального распределения результатов их работы. Теоретическая база для решения данной задачи — приближение регрессии первого рода линейной регрессией второго рода для многомерных случайных величин. Теоретические вопросы для общего случая рассмотрены в работах [4, 5], а в настоящей работе приведено попарное сравнение методов для частного случая.

Предложенный алгоритм предполагается выполнять перед стеганоанализом, а также по мере появления новых стеганоаналитических методов для выявления и отбрасывания схожих методов в рамках заданных достоверности и точности.

Постановка задачи

Рассмотрим выборку из n групп по g файлов в каждой. Файлы являются пустыми и заполненными контейнерами с шагом вложения 10, т. е. размер вложения $s = 0, 10, \ldots, 100$ в процентах от максимально возможного. В данной работе предполагается, что используется m стеганоаналитических методов.

Обозначим результаты работы второго метода применительно к первой файловой группе $\mathbf{x}_1^2 = \{x_{1,i}^2\}_{i=1}^{g}$, обобщенный вектор результатов $\mathbf{x}_n^m = \{x_{n,i}^m\}_{i=g(n-1)+1}^{gn}$.

Построим n-размерную случайную величину для каждого m-го метода:

$$\zeta_m = (\zeta_{m,1},\zeta_{m,2}, \ldots, \zeta_{m,n-1}, \zeta_{m,n}) = \\ = \left(\frac{1}{\sqrt{g}}\sum_{j=1}^{g} x_{1,j}^m, \ldots, \frac{1}{\sqrt{g}}\sum_{j=(n-1)g+1}^{gn} x_{n,j}^m\right). \quad (1)$$

Согласно центральной предельной теореме для многомерных случайных величин, случайные величины ζ_k, $(k = 1, \ldots, m)$ для $\tilde{\mathbf{x}} = (x_1, \ldots, x_m)$ имеют нормальное распределение с неизвестной функцией распределения

$$f_{\eta_k}(\tilde{\mathbf{x}}) = \frac{\sqrt{|\mathbf{A}|}}{(2\pi)^{\frac{m}{2}}}\exp\left(-\frac{1}{2}\langle\mathbf{A}(\tilde{\mathbf{x}} - \mathbf{h})^T, (\tilde{\mathbf{x}} - \mathbf{h})\rangle\right) = \\ = \frac{\sqrt{|\mathbf{A}|}}{(2\pi)^{\frac{m}{2}}}\exp\left(-\frac{1}{2}\sum_{i,j}a_{ij}(x_i - h_i)(x_j - h_j)\right),$$

где вектор $\mathbf{h} = (h_1, \ldots, h_m)$ и $\mathbf{A} = \{a_{ij}\}_{i,j=1}^m$ — симметричная положительно-определенная матрица, $\langle\rangle$ — обозначение скалярного произведения. Значение g выбирается из соображений обеспечения нормальности распределения ζ_m [6].

Применительно к стеганоанализу можно сказать, что есть методы, опирающиеся на одинаковые закономерности естественного контейнера и использующие схожий математический аппарат. Результаты их работы статистически идентичны, поэтому один из таких методов можно отбросить. Заметим, что условие

$$f(\zeta_1) = \zeta_2,$$

где f — измеримая функция, которую можно заменить условием математического ожидания $M(x)$:

$$\min_f M(\zeta_2 - f(\zeta_1))^2 = M(\zeta_2 - f_{21}(\zeta_1))^2 = 0.$$

Также можно приближенно считать, что для некоторых a_{21} и b_{21}

$$f_{21}(x) \approx a_{21} + b_{21}x,$$

поэтому

$$\min_{a,b} M(\zeta_2 - a - b\zeta_1)^2 \approx \min_f M(\zeta_2 - f(\zeta_1))^2 = 0.$$

Окончательно, если для методов 1 и 2 выполнено условие

$$M(\zeta_2 - a_{21} - b_{21}\zeta_1)^2 < \gamma,$$

то случайные величины ζ_1 и ζ_2 можно считать схожими.

Если выполнено условие

$$M(\zeta_2 - a_{21} - b_{21}\zeta_1)^2 > \gamma,$$

то считаем, что случайная величина ζ_2 принципиально отличается от ζ_1 (результаты работы методов 1 и 2 различны). Значение γ представляет собой меру схожести методов и выбирается аналитиком.

Обобщим данное утверждение. Пусть $1 \le k \le m$ произвольное натуральное число. В соответствии с [7], если уравнение

$$\zeta_{k+1} = f(\zeta_1, \ldots, \zeta_k) \quad (2)$$

верно, то случайная величина ζ_{k+1} статистически значимо связана с остальными случайными величинами. Метод $k + 1$ не привносит никакого улучшения в эту группу методов и может быть отброшен. Заметим, что условие (2) можно заменить условием

$$M(\zeta_{k+1} - f_{k+1}(\zeta_1, \ldots, \zeta_k))^2 = \min_f M(\zeta_{k+1} - f(\zeta_1, \ldots, \zeta_k))^2 = 0.$$

Примем, что для некоторых a_i и $b_{k+1,i}$

$$f_{k+1}(\mathbf{x}_1, \ldots, \mathbf{x}_k) \approx a_i + \sum_{i=1}^{k} b_{k+1,i}\mathbf{x}_i.$$

Следовательно

$$M\left(\zeta_{k+1} - a_i - \sum_{i=1}^{k} b_{k+1,i}\zeta_i\right)^2 \approx \min_f M(\zeta_2 - f(\zeta_1))^2 = 0.$$

Окончательно, если для некоторого $\gamma > 0$ (степень схожести) условие

$$M\left(\zeta_{k+1} - a_i - \sum_{i=1}^{k} b_{k+1,i}\zeta_i\right)^2 < \gamma$$

верно, то случайная величина ζ_{k+1} может быть отброшена.

Если условие

$$M\left(\zeta_{k+1} - a_i - \sum_{i=1}^{k} b_{k+1,i}\zeta_i\right)^2 > \gamma$$

верно, то случайная величина ζ_{k+1} фундаментально отличается от $\{\zeta_i\}_{i=1}^m$ и не может быть отброшена.

Математическая модель редукции

В основе предлагаемой методики лежит сравнение каждого метода с другим методом, с парой методов, с тройкой методов и т. д. для выявления оптимальной комбинации методов. Ограничимся рассмотрением самого простого частного случая (1-to-1) — сравнение двух случайных величин. Для каждой случайной величины ζ_i обозначим $m_i = M(\zeta_i)$. Выполним дальнейшие построения на примере двух первых случайных величин:

$$\min_a M(\zeta_2 - a - b_{21}\zeta_1)^2 = M((\zeta_2 - m_2) - \\ - b_{21}(\zeta_1 - m_1) + (m_2 - b_{21}m_1 - a_{12}))^2 = \\ = M(\zeta_2 - m_2)^2 - 2b_{21}M((\zeta_2 - m_2)(\zeta_1 - m_1)) + \\ + b_{21}^2 M(\zeta_1 - m_1)^2 + (m_2 - \beta_{21}m_1 - a_{12})^2 = \\ = M(\zeta_2 - m_2)^2 - 2b_{21}M((\zeta_2 - m_2)(\zeta_1 - m_1)) + \\ + b_{21}^2 M(\zeta_1 - m_1)^2. \quad (3)$$

Равенство (3) выполняется при удовлетворении условия

$$a = a_{21} = m_2 - b_{21}m_1.$$

Минимизируя по b_{21}, имеем

$$b_{21} = \frac{M((\zeta_2 - m_2)(\zeta_1 - m_1))}{M(\zeta_1 - m_1)^2},$$

$$M((\zeta_2 - m_2)(\zeta_1 - m_1)) = M(\zeta_1 - m_1)^2 b_{21}.$$

Согласно (3), имеем

$$\min_a M(\zeta_2 - a - \beta_{21}\zeta_1)^2 = M(\zeta_2 - m_2)^2 - \beta_{21}^2 M(\zeta_1 - m_1)^2. \quad (4)$$

Для оценки дисперсии и ковариации зададим уровень значимости, например, α = 0,95. Найдем коэффициенты β, входящие в выражение (4). Для всех $k = 1, \ldots, n$ рассмотрим m-мерные нормальные независимые случайные векторы $\zeta_k = (\zeta_{1k}, \ldots, \zeta_{mk})$, плотности вероятности которых для всех $k = 1, \ldots, n$ совпадают и равны

$$f_{\eta_k}(\tilde{\mathbf{x}}) = \frac{\sqrt{|\mathbf{A}|}}{(2\pi)^{\frac{m}{2}}} \exp\left(-\frac{1}{2}\langle \mathbf{A}(\tilde{\mathbf{x}} - \mathbf{h})^T, (\tilde{\mathbf{x}} - \mathbf{h})\rangle\right),$$

где $\mathbf{A} = \{a_{ij}\}_{i,j=1}^{m}$ — симметричная положительно-определенная матрица.

Для $1 \le i,j \le m$ и $1 \le k \le n$ определим случайную величину:

$$\bar{\zeta}_i = \frac{1}{n}\sum_{k=1}^{n} \zeta_{ik}. \quad (5)$$

Обозначим массив коэффициентов ковариации $\mathbf{\Theta} = \{\bar{\theta}_{ij}\}_{i,j=1}^{m}$, тогда

$$\bar{\theta}_{ij} \overset{\text{def}}{=} \frac{1}{n}\sum_{k=1}^{n} \zeta_{ik}\zeta_{jk} - \bar{\zeta}_j\bar{\zeta}_i = \frac{1}{n}\sum_{k=1}^{n} (\zeta_{ik} - \bar{\zeta}_i)(\zeta_{jk} - \bar{\zeta}_j). \quad (6)$$

Пользуясь уровнем значимости α и n найдем q, такое, чтобы выполнить условие

$$\int_{\frac{n}{(1+q)^2}}^{\frac{n}{(1-q)^2}} f_{\xi^2}^{n-1}(x)dx = \alpha, \quad (7)$$

где $f_{\xi^2}^{n-1}$ — плотность вероятности распределения ξ-квадрат с $n - 1$ степенями свободы. В этом случае, при $0 < q < 1$ для найденного n, величина $M(\zeta_k - m_k)^2$ с вероятностью α находится в промежутке функции вероятности $P(x)$

$$P(\bar{\theta}_{kk}(1 - q) < M(\zeta_k - m_k)^2 < \bar{\theta}_{kk}(1 + q)) = \int_{\frac{n}{(1+q)^2}}^{\frac{n}{(1-q)^2}} f_{\xi^2}^{n-1}(x)dx = \alpha. \quad (8)$$

При $q \ge 1$, для найденного n, величина $M(\zeta_k - m_k)^2$ с вероятностью α находится в промежутке

$$P(0 < M(\zeta_k - m_k)^2 < \bar{\theta}_{kk}(1 + q)) = \int_{\frac{n}{(1+q)^2}}^{\infty} f_{\xi^2}^{n-1}(x)dx = \alpha. \quad (9)$$

Величину $M((\zeta_i - m_i)(\zeta_j - m_j))$ рассчитаем по формуле

$$M((\zeta_i - m_i)(\zeta_j - m_j)) = \beta_{ij} M(\zeta_j - m_j)^2.$$

Заметим, что индекс коэффициента β_{ij} формируется следующим образом: сначала номер случайной величины, которую хотим приблизить, затем номер величины, на которую устанавливается регрессия. Рассмотрим нахождение β_{21}.

Обозначим

$$w_{21} = \frac{\begin{vmatrix} \bar{\theta}_{11} & \bar{\theta}_{21} \\ \bar{\theta}_{21} & \bar{\theta}_{22} \end{vmatrix}}{\bar{\theta}_{11}} = \frac{\bar{\theta}_{11}\bar{\theta}_{22} - \bar{\theta}_{21}^2}{\bar{\theta}_{11}}, \quad (10)$$

$$u_{21} = \sqrt{\bar{\theta}_{11}}\left(\frac{\bar{\theta}_{21}}{\bar{\theta}_{11}} - \beta_{21}\right) = \sqrt{\bar{\theta}_{11}}(b_{21} - \beta_{21}).$$

В соответствии с [8] случайная величина

$$\sqrt{n-2}\,\frac{u_{21}}{\sqrt{w_{21}}} = \frac{\sqrt{\bar{\theta}_{11}}(b_{21} - \beta_{21})\sqrt{n-2}}{\sqrt{w_{21}}}$$

имеет распределение Стьюдента с $n - 2$ степенями свободы.

Пользуясь уровнем значимости α, найдем t такое, чтобы выполнить условие

$$\int_{-t}^{t} f_{n-2}(x)dx = \alpha, \quad (11)$$

где f_{n-2} — плотность вероятности распределения Стьюдента с $n - 2$ степенями свободы. В этом случае для найденного t величина β_{21} с вероятностью α находится в промежутке

$$P\left(b_{21} - \frac{t\sqrt{w_{21}}}{\sqrt{\bar{\theta}_{11}}\sqrt{n-2}} < \beta_{21} < b_{21} + \frac{t\sqrt{w_{21}}}{\sqrt{\bar{\theta}_{11}}\sqrt{n-2}}\right) = \int_{-t}^{t} f_{n-2}(x)dx = \alpha. \quad (12)$$

Обобщенный алгоритм редукции

Рассмотрим методику работы обобщенного алгоритма редукции.

1. Определение количества групп.

   Исходя из требуемой достоверности α, в соответствии с формулой (9) подберем значения n и q. Выберем степень схожести γ.
2. Формирование исходных данных.
   - — Подбор требуемого количества файлов g.
   - — Реализация стегановложения с заданным шагом.
   - — Формирование группы контейнеров.
   - — Для каждого контейнера расчет стеганоаналитического вектора признаков.
   - — В соответствии с формулой (1) построение n-мерных реализаций случайной величины ζ_m. Проверена нормальность распределения. Если ζ_m распределена не нормально, то выполняется увеличение g и повтор данного пункта с новым файлом g.
3. Попарное сравнение случайных величин ζ_m.
   - — Вычисление элементов матрицы **W**. Элементы матрицы w_{ij} вычисляются по формулам (5), (6) и (10).
   - — Вычисление коэффициента β. В соответствии с выбранным значением определение значения t по формуле (11) и определение коэффициента β_{ij} по формуле (12).
   - — Оценка дисперсии $M(\zeta_k - m_k)^2$. В соответствии со значениями n, q, и α вычисление дисперсии по формулам (8) при $0 < q < 1$, или (9) при $q > 1$.
4. Принятие решения.

   Полученные значения $M(\zeta_k - m_k)^2$ и β_{ij} подставляем в формулу (4). Если максимальное значение $M(\zeta_k - m_i)^2 - \beta_{ij}^2 M(\zeta_j - m_j)^2 < \gamma$, то метод отбрасывается.

Схема программной реализации[1]

Схема программной реализации редукции набора стеганоаналитических методов с заданной достоверностью представлена на рис. 1.

Для программной реализации выбран язык программирования Python. В качестве среды разработки использовано приложение Jupyter Notebook, входящее в состав дистрибутива Anaconda. Все библиотеки, упоминаемые далее, входят в состав данного дистрибутива, если не указано иное.

Рассмотрим шаги программной реализации.

Шаг 1. Ввод исходных данных из файла. Для работы с файлами использованы библиотеки xlrd и xlwt. Исходный файл start_data.xls содержит двумерную таблицу, в которой строки соответствуют методам (m), столбцы — группам файлов (n).

Шаг 2. Вычисление массива коэффициентов ковариации методов с помощью функции cov библиотеки numpy.

Шаг 3. Вычисление массива коэффициентов ширины диапазонов Omega.

Шаг 4. Вычисление матрицы отношений ковариации к дисперсии и помещение ее в массив **b**.

Шаг 5. Определение конечных точек диапазона, содержащего заданный процент распределения Стьюдента. Для этого использована функция interval библиотеки scipy.stats.t для работы с функциями распределения Стьюдента. В качестве исходных данных выступают: заданный процент, степени свободы, среднее значение, среднеквадратическое отклонение.

Шаг 6. Нахождение нижних и верхних границ диапазонов по Стьюденту. Матрицы Beta_from_Stud и Beta_to_Stud содержат границы диапазонов, в которых находится величина Beta для найденных на шаге 5 точек с заданной вероятностью.

Шаг 7. Вычисление Disper_from_Khi и Disper_to_Khi – матриц диапазонов дисперсий методов. Для вычисления использована функция q_fron_Khi с исходными данными: количество степеней свободы, требуемая достоверность, точность определения корней. Данная функция не входит в состав стандартных функций дистрибутива Anaconda, определена непосредственно в программе.

Шаг 8. Определение верхних диапазонов разностей выборочных средних по методам, взятых попарно. Результат помещен в массив Difference_max. Определены все возможные комбинации выборочных средних и выбраны максимальные из них.

Шаг 9. Определение отбрасываемых методов. Выполнено сравнение верхних диапазонов выборочных средних, полученные на шаге 8, с заданной степенью схожести Gamma, таким образом, чтобы разность между методами, взятыми попарно была меньше заданной.

Шаг 10. Коррекция исходных данных с учетом отброшенных методов и запись в файл.

[1] Исходный код доступен в Kaggle: [Электронный ресурс]. Режим доступа: https://www.kaggle.com/romansolodukha/batrlett-1-1 (дата обращения: 10.01.2022).

Программный эксперимент

Проведен эксперимент по верификации предложенной методики. В качестве стеганоаналитических методов для сравнения выбраны количественные детекторы LSB[2] (Least Significant Bits): Weighted Stego (WS) [9], Sample Pairs (SP) [10], Triples analysis (T) [11], Asymptotically Uniformly Most Powerful detection (AUMP) [12], Pair of Values (PoVs) [13]. WS использован в модификации без коррекции (WSn) и с коррекцией (WSy) смещения. Параметры AUMP, рекомендованные в работе [12]: размер блока пикселов — 16, степень полинома — 6. LSB представляет программу, которая вычисляет размер вложения на основании данных изображения. Приведенные детекторы используют один цветовой канал пространственной области изображения с глубиной цвета 8 бит.

В качестве источника контейнеров взята коллекция BOSSbase 1.01[3] (первые 1000 файлов), которая содержит 10 000 8-битных полутоновых изображений размером 512 × 512 пикселов в формате PGM (Portable Grey Map). Для дальнейшей работы файлы были преобразованы в формат BMP 8 бит с помощью XnConvert и в формат BMP 24 бит — пакетного преобразования FastStone Image Viewer. Поскольку все три цветовых плоскости контейнера одинаковые, для анализа использован только канал красного цвета. Контейнеры заполнены с шагом 10 % от максимального размера вложения от 9 до 99 %. Таким образом, выборка составила 11 000 контейнеров[4].

Выбраны следующие стеганографические программы, реализующие LSB пространственной области изображения:

1. CryptArkan 1.0 — не требует установки, позволяет задействовать 1, 2 или 4 LSBs файла, шифровать стегановложение. Вложение реализуется последовательно, что видно из рис. 2, *a–c*. Для эксперимента реализован режим LSB.
2. The Third Eye 1.0 — не требует установки, задействует плоскость младшего бита. Байты для вложения выбраны случайно, что видно из рис. 2, *d–f*.

Изображения на рис. 2 получены с помощью программы WinMerge[5] и иллюстрируют локализацию изменений изображения с размерами вложения 9, 49 и 99 %.

Для достоверности $\alpha = 0{,}95$, устойчивое соответствие нормальному распределению по критерию Шапиро–Уилка на уровне значимости 0,01 было достигнуто при количестве групп $n = 100$. Далее программный модуль редукции запускался с различными

[2] Structural LSB Detectors [Электронный ресурс]. Режим доступа: http://dde.binghamton.edu/download/structural_lsb_detectors/ (дата обращения 10.01.2022).

[3] Image Database BOSSbase 1.01 [Электронный ресурс]. Режим доступа: http://dde.binghamton.edu/download/ImageDB/BOSSbase_1.01.zip (дата обращения 10.01.2022).

[4] Датасет доступен в Kaggle: [Электронный ресурс]. Режим доступа: https://www.kaggle.com/romansolodukha/bartlett-reduction (дата обращения: 10.01.2022).

[5] WinMerge. Open Source инструмент сравнения и слияния для Windows [Электронный ресурс]. Режим доступа: https://winmerge.org/downloads/ (дата обращения: 10.01.2022).

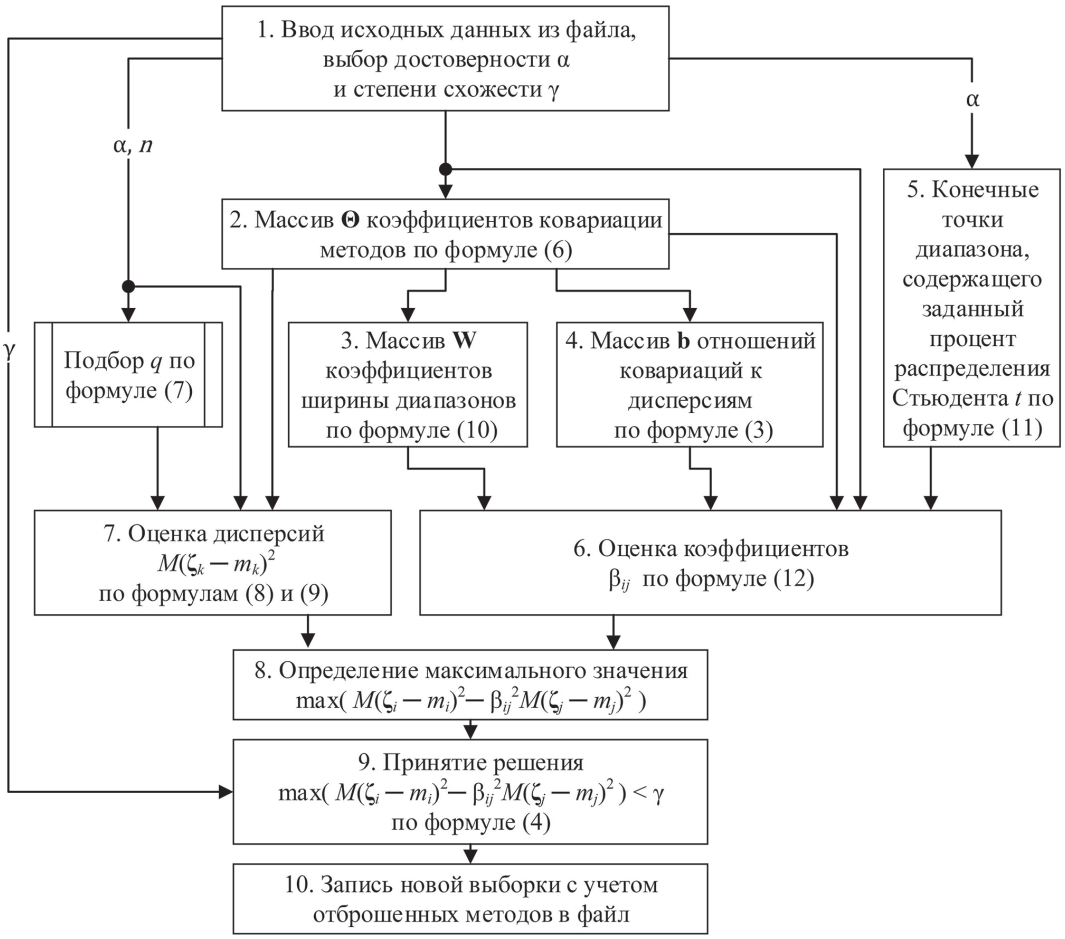

Рис. 1. Схема программной реализации редукции набора методов с заданной достоверностью
Fig. 1. Scheme of software implementation of methods set reduction with definite reliability

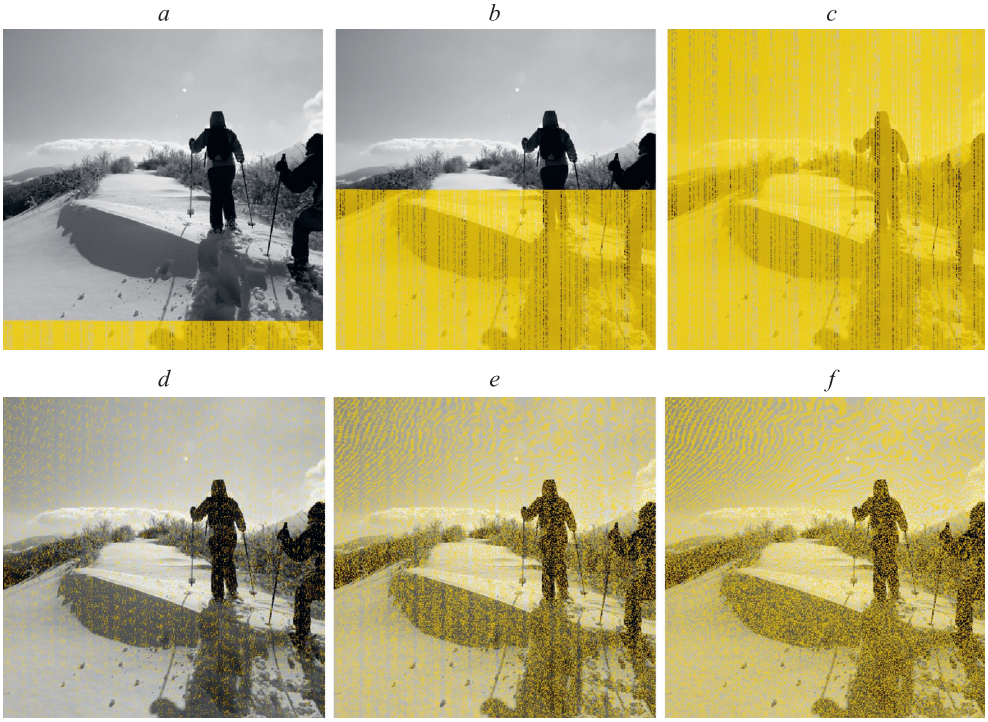

Рис. 2. Модификация пространственной области изображения программ CryptArkan (*a–c*) и The Third Eye (*d–f*)
Fig. 2. Modification of the image spatial domain by the CryptArkan (*a–c*) and The Third Eye (*d–f*)

Таблица. Степени схожести количественных детекторов LSB
Table. Similarity of quantitative LSB detectors

Отброшенные методы	CryptArkan			The Third Eye		
	Размер вложения, %					
	0	49	99	0	49	99
WSn	0,31	0,33	0,36	0,32	0,34	0,39
WSn, AUMP	0,51	0,54	0,59	0,56	0,50	0,58
WSn, AUMP, SP	0,82	0,86	0,91	0,84	0,88	0,91
WSn, AUMP, SP, T	1,00	1,04	1,05	1,01	1,06	1,10
WSn, AUMP, SP, T, PoVs	1,12	1,14	1,18	1,13	1,16	1,21

значениями γ и отбрасывал коррелированные методы. Рассмотрены три уровня вложения 0, 49 и 99 %. Для каждой стеганопрограммы и уровня вложения методы отбрасывались в одинаковом порядке: WSn, AUMP, SP, Triples, PoVs. Результаты в виде значений γ приведены в таблице.

Анализ таблицы позволяет сделать вывод о том, что коррелированность работы методов не зависит от стеганопрограммы и уменьшается с увеличением размера вложения. Интерпретировать полученные результаты можно следующим образом: с вероятностью 0,95 квадрат разности значений отброшенного и какого-либо из оставшихся методов не превышает γ. В плане практического применения видно, что различия между самыми близкими методами WSn и WSy составляют $\gamma = [0{,}31, 0{,}39]$, и для задач с $\gamma < 0{,}4$ WSn может быть отброшен.

Выводы и перспективы

Представлена методика, включающая математическую модель, алгоритм реализации модели и компьютерную программу для редукции набора стеганоаналитических методов. В дальнейших исследованиях по данной тематике предполагается провести программный эксперимент не с методами, а многомерными стеганоаналитическими векторами признаков (SPAM (686D) [14], SRM (34671D) [15], CHEN (486D) [16], etc.), применив предложенный подход для снижения размерности признакового пространства.

Заключение

Экспериментально подтверждена гипотеза о возможности вычисления схожести стеганоаналитических методов с помощью интервальных оценок. Это позволяет использовать предложенную в работе методику в сфере, относящейся к судебной экспертной деятельности. В программном эксперименте в качестве методов использованы количественные LSB детекторы, степени их схожести рассчитаны, применительно к стеганопрограммам The Third Eye и CryptArkan. Распространение предложенного подхода на известные стеганопрограммы/алгоритмы и стеганоаналитические методы позволит уменьшить вычислительные затраты стеганоанализа с заданной достоверностью при проведении исследования на основании известного стеганоалгоритма/программы.

Литература

1. Жижина М.В. Судебно-почерковедческая экспертиза документов / под ред. проф. Е.П. Ищенко. М.: ЮРЛИТИНФОРМ, 2006. С. 110–116.
2. Усов А.И., Градусова О.Б., Кузьмин С.А. Использование вероятностно-статистических методов при оценке значимости результатов экспертного исследования в отечественной и зарубежной судебно-экспертной практике (сравнительный анализ) // Теория и практика судебной экспертизы. 2018. Т. 13. № 4. С. 6–15. https://doi.org/10.30764/1819-2785-2018-13-4-6-15
3. Фёрстер Э., Рёнц Б. Методы корреляционного и регрессионного анализа. Руководство для экономистов / пер. с нем. М.: Финансы и статистика, 1983. 304 с.
4. Атласов И.В., Солодуха Р.А. Стеганоанализ цифровых изображений: автоматизация, оптимизация, достоверность: монография. Электр. дан. и прогр. Воронеж: Воронежский институт МВД России, 2020. 171 с. [Электронный ресурс]. URL: www.kaggle.com/dataset/c736afc689f328127816c59961677b3106468ce9f8e4399f3d18a12745c9e94c (дата обращения: 10.11.2021).
5. Atlasov I., Solodukha R. Reduction of steganalytical methods set with determined reliability // Proc. 1st International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency (SUMMA). 2019. P. 126–131. https://doi.org/10.1109/SUMMA48161.2019.8947470

References

1. Zhizhina M.V. *Forensic Handwriting Document Expertise*. Moscow, JURLITINFORM Publ., 2006, pp. 110–116. (in Russian)
2. Usov A.I., Gradusova O.B., Kuz'min S.A. The use of probabilistic and statistical methods to test the significance of scientific evidence: comparative analysis of current forensic practices in russia and abroad. *Theory and Practice of Forensic Science*, 2018, vol. 13, no. 4, pp. 6–15. (in Russian). https://doi.org/10.30764/1819-2785-2018-13-4-6-15
3. Förster E., Rönz B. *Methoden der Korrelations — und Regressionsanalyse*. Berlin, 1979, 324 p.
4. Atlasov I.V., Solodukha R.A. *Steganography of Digital Images: Automatization, Optimization and Reliability*. Voronezh, Voronezhskij institut MVD Rossii. Available at: www.kaggle.com/dataset/c736afc689f328127816c59961677b3106468ce9f8e4399f3d18a12745c9e94c (accessed: 10.11.2021). (in Russian)
5. Atlasov I., Solodukha R. Reduction of steganalytical methods set with determined reliability. *Proc. 1st International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency (SUMMA)*, 2019, pp. 126–131. https://doi.org/10.1109/SUMMA48161.2019.8947470
6. Wishart J., Bartlett M.S. The distribution of second order moment statistics in a normal system. *Mathematical Proceedings of the*

6. Wishart J., Bartlett M.S. The distribution of second order moment statistics in a normal system // Mathematical Proceedings of the Cambridge Philosophical Society. 1932. V. 28. P. 455–459. https://doi.org/10.1017/S0305004100010690
7. Фишер Р.А. Статистические методы для исследователей. М.: Госстатиздат, 1958. 268 с.
8. Bartlett M.S. On the theory of statistical regression // Proceedings of the Royal Society of Edinburgh. 1934. V. 53. P. 260–283. https://doi.org/10.1017/S0370164600015637
9. Ker A., Böhme R. Revisiting weighted stego-image steganalysis // Proceedings of SPIE. 2008. V. 6819. P. 681905. https://doi.org/10.1117/12.766820
10. Dumitrescu S., Wu X., Memon D. On steganalysis of random LSB embedding in continuous-tone images // IEEE International Conference on Image Processing. V. 3. 2002. P. 641–644.
11. Ker A. A general framework for structural steganalysis of LSB replacement // Lecture Notes in Computer Science. 2005. V. 3727. P. 296–311. https://doi.org/10.1007/11558859_22
12. Fillatre L. Adaptive steganalysis of Least Significant Bit replacement in grayscale natural images // IEEE Transactions on Signal Processing. 2012. V. 60. N 2. P. 556–569. https://doi.org/10.1109/TSP.2011.2174231
13. 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. V. 1768. P. 61–76. https://doi.org/10.1007/10719724_5
14. Pevný T., Bas P., Fridrich J. Steganalysis by subtractive pixel adjacency matrix // IEEE Transactions on Information Forensics and Security. 2010. V. 5. N 2. P. 215–224. https://doi.org/10.1109/TIFS.2010.2045842
15. Fridrich J., Kodovský J. Rich models for steganalysis of digital images // IEEE Transactions on Information Forensics and Security. 2012. V. 7. N 3. P. 868–882. https://doi.org/10.1109/TIFS.2012.2190402
16. Chen C., Shi Y.Q. JPEG image steganalysis utilizing both intrablock and interblock correlations // Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS). 2008. P. 3029–3032. https://doi.org/10.1109/ISCAS.2008.4542096

Cambridge Philosophical Society, 1932, vol. 28, pp. 455–459. https://doi.org/10.1017/S0305004100010690
7. Fisher R.A. *Statistical methods for research workers*. Oliver and Boyd. 12th ed., revised. Edinburg, London, Oliver and Boyd, 1954.
8. Bartlett M.S. On the theory of statistical regression. *Proceedings of the Royal Society of Edinburgh*, 1934, vol. 53, pp. 260–283. https://doi.org/10.1017/S0370164600015637
9. Ker A., Böhme R. Revisiting weighted stego-image steganalysis. *Proceedings of SPIE*, 2008, vol. 6819, pp. 681905. https://doi.org/10.1117/12.766820
10. Dumitrescu S., Wu X., Memon D. On steganalysis of random LSB embedding in continuous-tone images. *IEEE International Conference on Image Processing*, vol. 3, 2002, pp. 641–644.
11. Ker A. A general framework for structural steganalysis of LSB replacement. *Lecture Notes in Computer Science*, 2005, vol. 3727, pp. 296–311. https://doi.org/10.1007/11558859_22
12. 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. https://doi.org/10.1109/TSP.2011.2174231
13. 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. https://doi.org/10.1007/10719724_5
14. 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. https://doi.org/10.1109/TIFS.2010.2045842
15. Fridrich J., Kodovský J. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 2012, vol. 7, no. 3, pp. 868–882. https://doi.org/10.1109/TIFS.2012.2190402
16. Chen C., Shi Y.Q. JPEG image steganalysis utilizing both intrablock and interblock correlations. *Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 2008, pp. 3029–3032. https://doi.org/10.1109/ISCAS.2008.4542096

Авторы

Солодуха Роман Александрович — кандидат технических наук, доцент, доцент, Воронежский институт МВД России, Воронеж, 394065, Российская Федерация, SC 57201778092, https://orcid.org/0000-0002-3878-4221, standartal@list.ru

Перминов Геннадий Вадимович — кандидат технических наук, доцент, Российский государственный университет правосудия (центральный филиал), Воронеж, 394006, Российская Федерация, https://orcid.org/0000-0003-4170-2861, perminovgv@mail.ru

Атласов Игорь Викторович — доктор физико-математических наук, профессор, профессор, Московский университет МВД России имени В.Я. Кикотя, Москва, 117997, Российская Федерация, SC 57201775678, https://orcid.org/0000-0001-6270-6787, mathematic1@rambler.ru

Authors

Roman A. Solodukha — PhD, Associate Professor, Associate Professor, Voronezh Institute of the Ministry of Internal Affairs of Russia, Voronezh, 394065, Russian Federation, SC 57201778092, https://orcid.org/0000-0002-3878-4221, standartal@list.ru

Gennadiy V. Perminov — PhD, Associate Professor, Russian State University of Justice (Central Branch), Voronezh, 394006, Russian Federation, https://orcid.org/0000-0003-4170-2861, perminovgv@mail.ru

Igor V. Atlasov — D.Sc., Full Professor, Kikot Moscow University of the Ministry of Internal Affairs of Russia, Moscow, 117997, Russian Federation, SC 57201775678, https://orcid.org/0000-0001-6270-6787, mathematic1@rambler.ru

Статья поступила в редакцию 08.12.2021
Одобрена после рецензирования 20.12.2021
Принята к печати 28.01.2022

Received 08.12.2021
Approved after reviewing 20.12.2021
Accepted 28.01.2022