ЭФФЕКТИВНОСТЬ СТАТИСТИЧЕСКИХ МЕТОДОВ СТЕГАНОАНАЛИЗА ПРИ ОБНАРУЖЕНИИ ВСТРАИВАНИЯ В ВЕЙВЛЕТ ОБЛАСТЬ ИЗОБРАЖЕНИЯ

Сивачев А.В.[1]

Предметом исследования данной работы является оценка эффективности современных статистических методов стеганоанализа, разработанных для обнаружения встраивания в пространственную область изображения, при обнаружении встраивания в вейвлет область изображения. С появлением формата JPEG2000, использующего дискретное вейвлет преобразование (ДВП), активно разрабатываются методы встраивания информации в вейвлет область изображения с помощью стеганографии. Существующие методы стеганоанализа показывают низкую эффективность обнаружения факта встраивания в вейвлет область. В связи с низкой эффективностью обнаружения встраивания в вейвлет область изображения в данной статье рассматривается возможность адаптации и использования статистических методов стеганоанализа пространственной области изображения для обнаружения встраивания в вейвлет область изображения. Проведенный в работе анализ позволяет оценить практические возможности использования статистических методов стеганоанализа для обнаружения встраивания в вейвлет область изображения. При оценке эффективности статистических методов стеганоанализа используется принцип создания равных условий. Стеганографическое воздействие моделируется путем изменения значений предпоследних значащих бит коэффициентов ДВП всех областей по отдельности, получаемых при одноуровневом двумерном ДВП изображения. Для обеспечения высокой достоверности результатов используется коллекция с большим количеством изображений. Результатом проведенного исследования являются графики, показывающие эффективность обнаружения встраивания в различные области коэффициентов вейвлет области. Основным результатом работы является предлагаемый метод стеганоанализа, с использованием машинного обучения, обеспечивающий высокую эффективность обнаружения встраивания в LL область изображения, основанный на использовании корреляции между изображением и LL областью данного изображения.

***Ключевые слова**: стеганография, стеганоанализ, статистические методы, вейвлет область, дискретное вейвлет преобразование (ДВП), низкочастотная область ДВП, принцип создания равных условий, машинное обучение, бинарная классификация, пассивное противодействие*

DOI: 10.21681/2311-3456-2018-1-72-78

Введение

Сегодня стеганография применяется для организации скрытых каналов передачи информации и ряда других задач [1]. Возможность создания скрытых каналов связи с использованием стеганографии привлекла внимание со стороны криминальных элементов, террористических организация и разведывательных служб [2]. Одним из наиболее популярных видов контейнеров для создания скрытого канала связи являются неподвижные цифровые изображения т.к. с одной стороны для цифровых изображений разработано большое количество стеганографических методов встраивания информации в цифровые изображения [3], а с другой стороны каждый день в интернете загружается и передается огромное количество изображений [4]. На сайте ФБР США можно найти примеры изображений, использованных для встраивания информации [5].

Использование стеганографии для создания скрытых каналов связи с противоправными целями привело к разработке методов стеганоанализа, позволяющих обнаружить факт встраивания информации в контейнер, для пассивного противодействия использованию стеганографии. В идеале метод стеганоанализа должен позволять безошибочно определять содержит данный контейнер встроенную информацию или нет. Существующие методы стеганоанализа для цифровых изображений имеют определенную погрешность. Например, метод стеганоанализа, предложенный в [6], имеет эффективность от 50% до 80% в зависимости от использованного метода встраивания. Согласно проведенному исследованию [7] современные методы стеганоанализа не позволяют эффективно обнаруживать встраивание информации в области коэффициентов LH, HL и LL ДВП изображения.

Низкая эффективность методов стеганоанализа при обнаружении встраивания в области коэффициентов ДВП изображения делает актуальным исследования направленные на повышение её эффективности. С точки зрения повышения эффективности обнаружения встраивания в вейвлет область целесообразно рассмотреть возможность адаптации и использования имеющихся методов

1 Сивачев Алексей Вячеславович, аспирант, Университет ИТМО, Санкт-Петербург, Россия. E-mail: sivachev239@mail.ru

обнаружения факта встраивания в пространственную область изображения. В частности, при обнаружении встраивания в пространственную область изображения хорошие результаты показывают статистические методы стеганоанализа [8]: 97,5% обнаруженных стегано изображений при менее 2.5% некорректно определенных оригинальных изображений.

Цель работы

В работе проводится анализ эффективности использования статистических методов стеганоанализа, разработанных для обнаружения встраивания в пространственную область изображения, для вейвлет области изображения.

Исследование предметной области

Статистические методы стеганоанализа [9] разрабатывались для обнаружения характерных искажений, возникающих в изображении вследствие стеганографического встраивания в пространственную область. Результатом применения статистического метода является значение оценки количество искаженных пикселей в изображении. В тоже время встраивание информации в коэффициенты ДВП изображения тоже вызывает определенные искажения в изображении – рассмотрим возможность обнаружения данных искажений с использованием статистических методов стеганоанализа. Для этого проанализируем стегано изображения, в которых смоделировано встраивание в каждой из возможных областей (LL, LH, HL, HH) коэффициентов ДВП изображения, с помощью статистических методов стеганоанализа.

В качестве методов статистического стеганоанализа выберем методы: RS-анализ (RS), Sample pair analysis (SPA), Triples analysis (TR) и Weighted stego (WS), описанные в [9], показавшие хорошие результаты при обнаружении факта встраивания в пространственную область изображения [8]. В качестве коллекции изображений используется коллекция BOWS2. Для оценки эффективности методов статистического стеганоанализа используем полученные и представленные на рисунке 1 графики соотношения количества верно классифицированных стегано изображений, содержащих встраивание в соответствующую область ДВП изображения, от неверно классифицированных оригинальных изображений (т.е. ошибочно классифицированных как стегано изображения) .

Анализ графиков, приведенных на рисунке 1, показал неэффективность методов стеганоанализа [9] при встраивании в области коэффициентов

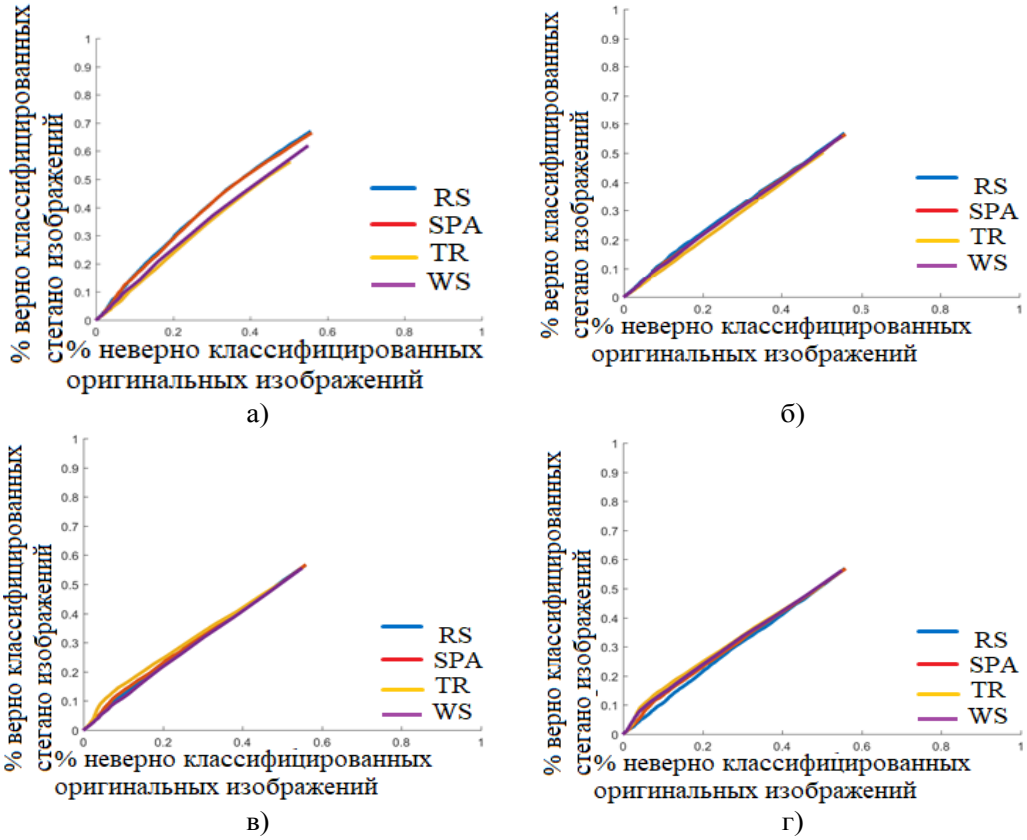


***Рис. 1**. График соотношения количества верно классифицированных стегано изображений от неверно классифицированных оригинальных изображений при встраивании в: а) LL области; б) LH области; в) HL области; г) HH области*

ДВП – обнаружено менее 70% стегано изображений и более 50% оригинальных изображений ошибочно классифицированы как стегано изображения. На практике это обозначает, что использование статистических методов для анализа изображений не позволяет отличить изображение оригинал от изображения со встроенной информацией в одну или несколько областей.

Низкая эффективность статистических методов стеганоанализа объясняется тем, что обратное ДВП, при котором из областей коэффициентов LL, LH, HL и HH получается исходное изображение, сглаживает искажения, возникающие вследствие встраивания в области коэффициентов ДВП. В тоже время получаемые при ДВП области LL, LH, HL и HH естественно взаимосвязаны с исходным изображением – т.е. могут сохранить закономерности характерные для естественных цифровых изображений.

Для оценки степени взаимосвязи между пикселями исходного изображения и коэффициентами областей LL, LH, HL и HH, полученными при ДВП изображения, рассчитаем коэффициент корреляции между ними по следующей формуле:

$$k = \frac{\sum_{x=1}^{X}\sum_{y=1}^{Y}(IMG(x,y)-\overline{IMG})*(W\left(RND(\frac{x}{2}),RND(\frac{y}{2})\right)-\bar{W})}{\sqrt{(\sum_{x=1}^{X}\sum_{y=1}^{Y}(IMG(x,y)-\overline{IMG})^2)*\left(\sum_{x=1}^{X}\sum_{y=1}^{Y}(W\left(RND(\frac{x}{2}),RND(\frac{y}{2})\right)-\bar{W})^2\right)}},$$

где k – коэффициент корреляции, $IMG(a,b)$ – значение пикселя изображения с координатами a и b, $\overline{IMG}$ – среднее значение пикселя в изображении, RND(a) – функция округление вверх значения a, W(a,b) – значение коэффициента одной из областей ДВП (LL, LH, HL, HH) с координатами a и b, $\bar{W}$ – среднее значение коэффициента области ДВП. Коэффициент корреляции $k \in [-1,1]$, при этом значения: «+1» – положительная прямолинейная корреляция; «-1» – отрицательная прямолинейная корреляция; «0» – отсутствие корреляции.

На рисунке 2 представлены гистограммы значения коэффициента корреляции для между исходным изображением и каждой из областей ДВП изображения (LL, LH, HL, HH).

На гистограммах, представленных на рисунке 2, видно, что корреляция между исходным изображением и областями LH, HL, HH практически отсутствует т.к. значение коэффициента корреляции в данном случае близко к нулю. В тоже время между исходным изображением и областью LL изображения существует прямолинейная положительная корреляции т.к. значение коэффициента корреляции близко к единице (среднее значение коэффициента корреляции для коллекции из 10000 изображений составляет 0,9803).

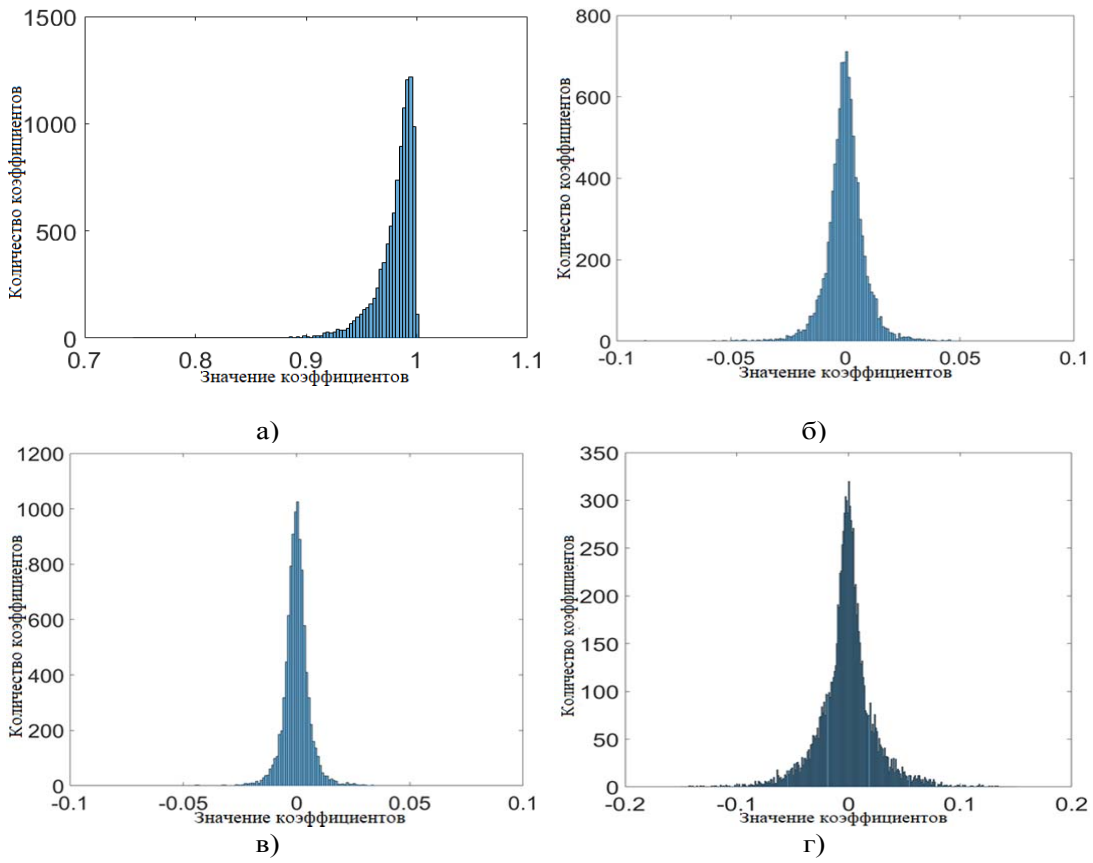
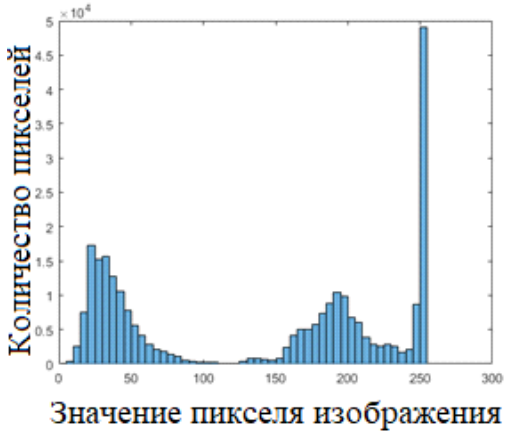


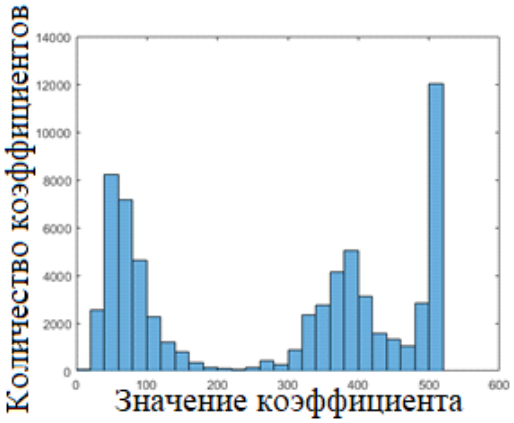
***Рис. 2**. Гистограммы значений коэффициента корреляции для массива исходных изображений и областей ДВП, где: а) LL области; б) LH области; в) HL области; г) HH области*

***Рис. 3а**. Изображение (пример)*

***Рис. 3б**. LL область, получаемая при ДВП изображения*

а)





б)

***Рис. 4**. Гистограмма значений: а) пикселей изображения (с рисунка 3а); б) коэффициентов LL области (с рисунка 3б)*

При визуальном сравнении исходного изображения и его LL области выявлено заметное сходство между ними. На рисунке 3 приведен пример изображения и LL области, получаемой в результате ДВП данного изображения. На рисунке 4 приведены гистограммы значений пикселей изображения и значений коэффициентов LL области.

Наличие корреляции между изображением и LL областью, а также визуальное сходство, позволяет предположить, что закономерности характерные для цифрового изображения характерны и для LL области. Наличие таких закономерностей в LL области дает возможность применения статистических методов не только к изображениям, как это задумывалось авторами статистических методов, но и непосредственно к LL области с целью обнаружения факта встраивания в область LL.

Проверим эффективность методов статистического стеганоанализа, описанных в [9], применяя их не к изображению, а непосредственно к LL области, получаемой в результате ДВП. На рисунке 5 представлен график соотношения количества верно классифицированных стегано изображений от неверно классифицированных оригинальных изображений при применении статистических методов стеганоанализа непосредственно к LL области.

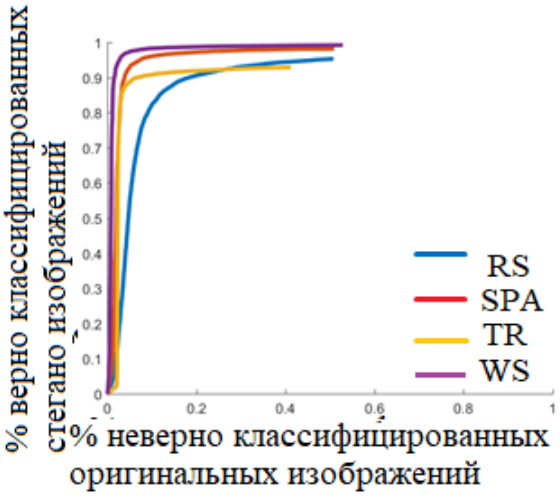


***Рис. 5**. График соотношения количества верно классифицированных стегано изображений от неверно классифицированных оригинальных изображений для LL области*

График, приведенный на рисунке 5, наглядно показывает возможность эффективного стеганоанализа изображений со встраиванием в коэффициенты LL области путем применения статистических методов стеганоанализа непосредственно к LL области. Применении методов статистического стеганоанализа непосредственно к LL области позволяют верно классифицировать более 95% стегано изображений при этом менее 5% оригинальных изображений будут классифицированы неверно.

Описание предлагаемый метод и условия проведения эксперимента

Значение оценки объема искаженных коэффициентов, полученное с помощью статистических методов стеганоанализа при их применении к LL области, можно использовать в качестве дополнительного параметра для метода стеганоанализа на основе машинного обучения. В качестве базовых параметров возьмем статистические моменты, используемые методами машинного обучения, которые рассматривались в [7], и добавим к ним предлагаемый дополнительный параметр для повышения эффективности обнаружения факта встраивания в область LL.

Таким образом предлагаемый метод стеганоанализа использует для классификации изображений с помощью машинного обучения следующий набор параметров:

- 1, 2, 3, 4 статистические моменты для областей коэффициентов LL, LH, HL, HH;
- значение количества искаженных коэффициентов, полученное с помощью применения статистического метода стеганоанализа к LL области.

В рамках статьи в качестве значения количества искаженных коэффициентов взят результат использования статистического метода стеганоанализа Weighted stego, который показал высокую эффективность (см. рисунок 5)

Для сравнения эффективности предлагаемого метода с другими выберем следующие методы:

- метод, предложенный Gireesh Kumar и другими [10];
- метод, предложенный Hany Farid [11];
- метод, предложенный Changxin Liu и другими [12];
- метод, предложенный Yun Q. Shi и другими [13].

Для моделирования стеганографического встраивания в область LL ДВП изображения производилась модификация предпоследних бит коэффициентов LL области.

В качестве коллекции изображений была выбрана коллекция BOWS2, которая часто используется авторами в работах по стеганографии и стеганоанализу [14, 15] и насчитывающая 10000 изображений с разрешением 512x512 пикселей. На основе выбранной коллекции было сформировано множество изображений, состоящее из оригинальных и стегано (15% измененных коэффициентов LL области) изображений.

Результатом классификации изображений с помощью методов машинного обучения является бинарная классификация изображения: содержит изображение встраивание или нет. В идеале метод машинного обучения должен сто процентов оригинальных изображений классифицировать как оригинальные изображение, а сто процентов стегано изображений как стегано изображений. В реальности результат классификации изображения может не совпадать с реальным состоянием изображение, т.е. оригинальное изображение может быть ошибочно классифицировано как стегано изображение и наоборот.

Результаты эксперимента

На рисунке 5 приведен график соотношения TN (истинно отрицательные, % верно классифицированных оригинальных изображений), TP (истинно положительные, % верно классифицированных стегано изображений), FN (ложноотрицательные, % неверно классифицированных стегано изображений), FP (ложноположительные, % неверно классифицированных оригинальных изображений), T (истинные, суммарный % верно классифицированных изображений), F (истинные, суммарный % неверно классифицированных изображений) для LL области при использовании описанного выше набора параметров, а также методов [10-13].

Выводы

График, представленный на рисунке 5, показывает, что предложенный набор параметров дает высокую эффективность обнаружения для LL области: предложенный метод дает 96,36% верно классифицированных изображений, в то время как метод [11] дает только 61,7% верно классифицированных изображений. Таким образом увеличение эффективности стеганоанализа составило более 30%.

Заключение

По итогам исследования можно сделать следующие выводы:

Статистические методы стеганоанализа, разработанные для стеганоанализа пространственной области изображения, неэффективны при анализе изображений, содержащих встраивания в коэффициентах LL, LH, HL и HH;

Наличие корреляции между исходным изображением и LL областью, получаемой в результате ДВП данного изображения, а также заметное визуальное сходство, делает возможным применение

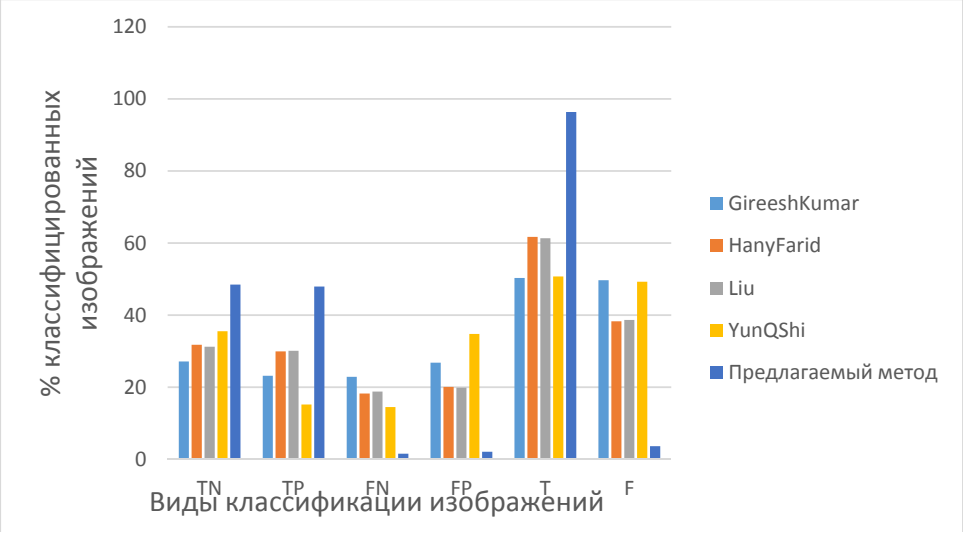


***Рис. 5**. График соотношения TN, TP, FN, FP, T, F при 5% объеме встраивания в LL область*

статистических методов стеганоанализа для обнаружения встраивания в коэффициенты области LL.

Использование объема искажений, получаемого при применении статистических методов к LL области, в качестве дополнительного параметра для методов стеганоанализа на основе машинного обучения позволяет повысить эффективность обнаружения встраивания в LL область изображения до 30%.

***Рецензент**: Коробейников Анатолий Григорьевич, доктор технических наук, профессор Санкт-Петербургского филиала Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова РАН. E-mail: korobeynikov_a_g@mail.ru*

Литература

1. Макаренко С.И. Эталонная модель взаимодействия стеганографических систем и обоснование на ее основе новых направлений развития теории стеганографии // Вопросы кибербезопасности. 2014. № 2 (3). С. 24-32.
2. Steganography: A Powerful Tool for Terrorists and Corporate Spies [Электронный ресурс]: Stratfor - Режим доступа: https://www.stratfor.com/analysis/steganography-powerful-tool-terrorists-and-corporate-spies.
3. Swain G., Lenka S.K. Classification of image steganography techniques in spatial domain: a study. [Текст] Int. J. Comput. Sci. Eng. Tech, (IJCSET) 2014 vol. 5 Pages 219-232.
4. Internet 2012 in numbers [Электронный ресурс]: Pingdom – Режим доступа: http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers/.
5. Steganography picture [Электронный ресурс]: Federal bureau of investigation – Режим доступа: https://vault.fbi.gov/ghost-stories-russian-foreign-intelligence-service-illegals/images/steganography-picture/view.
6. Gireesh Kumar T., Jithin R., Deepa D. Shankar Feature Based Steganalysis Using Wavelet Decomposition and Magnitude Statistics [Текст], Advances in Computer Engineering (ACE), 2010, pp. 298-300.
7. Сивачев А.В., Прохожев Н.Н., Михайличенко О.В., Башмаков Д.А. Эффективность стеганоанализа на основе методов машинного обучения // Вопросы кибербезопасности. 2017. № 2 (20). С. 53-60.
8. Prokhozhev N., Mikhailichenko O., Sivachev A., Bashmakov D., Korobeynikov A.G. Passive Steganalysis Evaluation: Reliabilities of Modern Quantitative Steganalysis Algorithms [Текст] // Advances in Intelligent Systems and Computing. 2016. V. 451. P. 89–94. doi:10.1007/978-3-319-33816-3_9
9. Rainer Böhme Advanced Statistical Steganalysis [Текст]. Springer Science & Business Media. 2010. 288 p
10. Gireesh Kumar T., Jithin R., Deepa D. Shankar. Feature Based Steganalysis Using Wavelet Decomposition and Magnitude Statistics, Advances in Computer Engineering (ACE) // International Conference on 2010. 2010. P. 298–300.
11. Farid Hany. Detecting Steganographic Messages in Digital Images // Technical Report TR2001-412, Dartmouth College, Computer Science Department, 2001.
12. Chunjuan Ouyang, Ming Guo, Huijuan Chen. Image Steganalysis Based on Spatial Domain and DWT Domain Features // Proceedings of the 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing. 2010. V. 1. P. 329–331.
13. Yun Q. Shi, Guorong Xuan, Chengyun Yang, Jianjiong Gao, Zhenping Zhang, Peiqi Chai, Dekun Zou, Chunhua Chen, Wen Chen. Effective steganalysis based on statistical moments of wavelet characteristic function // International Conference on Information Technology: Coding and Computing (ITCC'05). 2005. V. 2. P. 768–773
14. Rhythm Walia Steganography based on neighborhood pixels // Advances in Computing, Communications and Informatics (ICACCI). 2013. P. 203–206.
15. Jiaohua Qin, Xuyu Xiang, Yu Deng, Youyun Li and Lili Pan. Steganalysis of Highly Undetectable Steganography Using Convolution Filtering // Information Technology Journal. 2014. V. 13 P. 2588–2592.

EFFICIENCY OF STATISTICAL STEGNA ANALYSIS METHODS IN DETECTION EMBEDDING IN WAVELET DOMAIN

Sivachev A.[2]

Abstract. The subject of the study is the evaluation of the effectiveness of modern statistical methods of steganoanalysis, designed to detect embedding into the spatial domain, to detect embedding into the wavelet domain. Today methods of embedding information in the wavelet domain using steganography are being actively developed. The existing methods of steganoanalysis show a low efficiency of detecting the fact of embedding into the wavelet domain. Due to the low efficiency of detecting embedding in the wavelet domain, this article considers the possibility of adapting and using statistical methods of steganoanalysis of the spatial area to detect embedding in the wavelet domain. The analysis carried out in the work allows us to evaluate the practical possibilities of using statistical methods of steganoanalysis to detect embedding in the wavelet domain. When evaluating the effectiveness of statistical methods of steganoanalysis, the principle of creating equal conditions is used. The steganographic impact is modeled by changing the values of the penultimate significant bits of the DWT coefficients of all subbands separately, obtained with a single-level two-dimensional DWT. To ensure high reliability of the results, a collection with a large number of images is used. The result of the study is a graph showing the efficiency of detection of embedding into different subbands of the wavelet domain coefficients. The main result of the work is the proposed method of steganoanalysis, using machine learning, which provides high detection efficiency of embedding in the LL subband, based on the correlation between the image and the LL subband.

Keywords*: steganography, steganalysis, statistical methods, wavelet domain, discrete wavelet transform (DWT), LL subband, principle of equal conditions, machine learning, binary classification, passive attack*

References

1. Makarenko S.I. Etalonnaya model' vzaimodeystviya steganograficheskikh sistem i obosnovanie na ee osnove novykh napravleniy razvitiya teorii steganografii, Voprosy kiberbezopasnosti. 2014. No 2 (3), pp. 24-32.
2. Steganography: A Powerful Tool for Terrorists and Corporate Spies [Elektronnyy resurs]: Stratfor - Rezhim dostupa: https://www.stratfor.com/analysis/steganography-powerful-tool-terrorists-and-corporate-spies.
3. Swain G., Lenka S.K. Classification of image steganography techniques in spatial domain: a study. [Tekst] Int. J. Comput. Sci. Eng. Tech, (IJCSET) 2014 vol. 5 Pages 219-232.
4. Internet 2012 in numbers [Elektronnyy resurs]: Pingdom – Rezhim dostupa: http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers/.
5. Steganography picture [Elektronnyy resurs]: Federal bureau of investigation – Rezhim dostupa: https://vault.fbi.gov/ghost-stories-russian-foreign-intelligence-service-illegals/images/steganography-picture/view.
6. Gireesh Kumar T., Jithin R., Deepa D. Shankar Feature Based Steganalysis Using Wavelet Decomposition and Magnitude Statistics [Tekst], Advances in Computer Engineering (ACE), 2010, pp. 298-300.
7. Sivachev A.V., Prokhozhev N.N., Mikhaylichenko O.V., Bashmakov D.A. Effektivnost' steganoanaliza na osnove metodov mashinnogo obucheniya, Voprosy kiberbezopasnosti. 2017. No 2 (20), pp. 53-60.
8. Prokhozhev N., Mikhailichenko O., Sivachev A., Bashmakov D., Korobeynikov A.G. Passive Steganalysis Evaluation: Reliabilities of Modern Quantitative Steganalysis Algorithms [Tekst], Advances in Intelligent Systems and Computing. 2016. V. 451. R. 89–94. DOI:10.1007/978-3-319-33816-3_9.
9. Rainer Böhme Advanced Statistical Steganalysis [Tekst]. Springer Science & Business Media. 2010. 288 p.
10. Gireesh Kumar T., Jithin R., Deepa D. Shankar. Feature Based Steganalysis Using Wavelet Decomposition and Magnitude Statistics, Advances in Computer Engineering (ACE), International Conference on 2010. 2010. R. 298–300.
11. Farid Hany. Detecting Steganographic Messages in Digital Images, Technical Report TR2001-412, Dartmouth College, Computer Science Department, 2001.
12. Chunjuan Ouyang, Ming Guo, Huijuan Chen. Image Steganalysis Based on Spatial Domain and DWT Domain Features, Proceedings of the 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing. 2010. V. 1. P. 329–331.
13. Yun Q. Shi, Guorong Xuan, Chengyun Yang, Jianjiong Gao, Zhenping Zhang, Peiqi Chai, Dekun Zou, Chunhua Chen, Wen Chen. Effective steganalysis based on statistical moments of wavelet characteristic function, International Conference on Information Technology: Coding and Computing (ITCC'05). 2005. V. 2. P. 768–773
14. Rhythm Walia Steganography based on neighborhood pixels, Advances in Computing, Communications and Informatics (ICACCI). 2013. P. 203–206.
15. Jiaohua Qin, Xuyu Xiang, Yu Deng, Youyun Li and Lili Pan. Steganalysis of Highly Undetectable Steganography Using Convolution Filtering, Information Technology Journal. 2014. V. 13 P. 2588–2592.

2 Aleksei Sivachev, postgraduate student, St. Petersburg National Research University of Information Technologies, Mechanics and Optics, St. Petersburg, Russia. E-mail: sivachev239@mail.ru