Було розглянуто проблему розробки стеганографічного методу приховування інформації, стійкого до аналізу за моделлю Rich (що включає в себе декілька різних субмоделей) із використанням статистичних показників розподілу пар коефіцієнтів дискретного косинусного перетворення (ДКП) із різними значеннями. Сутність даного виду аналізу полягає у тому, що обчислюється кількість пар коефіцієнтів ДКП, координати яких у частотній області відрізняються на фіксовану величину (зсув). На основі цих значень для певної достатньо великої вибірки даних тренується класифікатор, який на основі розподілу пар коефіцієнтів ДКП окремого зображення визначає наявність додаткової інформації у ньому.

Для зменшення ймовірності виявлення прихованого повідомлення запропоновано метод, заснований на попередній модифікації контейнеру перед вбудовуванням повідомлення. Для модифікації було використано так звану генеративну змагальну мережу (ГЗМ), що складається з двох пов'язаних нейронних мереж – генератора та дискримінатора. Генератор створює модифіковане зображення на основі вихідного контейнера, а дискримінатор перевіряє, наскільки отримане модифіковане зображення близьке до заданого, та надає зворотній зв'язок для генератора.

За допомогою ГЗМ на основі вихідного контейнера генерується модифікований таким чином, щоб після вбудовування відомого стеганографічного повідомлення розподіл пар коефіцієнтів ДКП був максимально наближений до показників вихідного контейнера.

Було проведено комп'ютерне моделювання роботи запропонованої модифікації, на основі результатів моделювання обчислено ймовірності вірного виявлення прихованої інформації у контейнері при проведенні його модифікації та за умови її відсутності. Результати моделювання показали, що застосування модифікації, заснованої на сучасних інформаційних технологіях (таких, як машинне навчання та нейронні мережі) дозволяє помітно зменшити ймовірність виявлення повідомлення та підвищити стійкість до стеганографічного аналізу

Ключові слова: статистичні показники, машинне навчання, нейронна мережа, цифрова стеганографія, приховування інформації

UDC 004.056.5
DOI: 10.15587/1729-4061.2020.201731

A STEGANOGRAPHIC METHOD OF IMPROVED RESISTANCE TO THE RICH MODEL-BASED ANALYSIS

N. Kalashnikov
Postgraduate Student*
E-mail: kalashnikov_n.v@ukr.net
O. Kokhanov
Doctor of Technical Sciences,
Head of Department*
E-mail: okokhanov@gmail.com
O. Iakovenko
Postgraduate Student*
E-mail: iakovenko.oleksandr@gmail.com
N. Kushnirenko
PhD, Associate Professor**
E-mail: infsec2011@gmail.com
*Department of Radio Engineering Devices***
Department of Informatics and Information Systems Protection Management*
***Odessa National Polytechnic University
Shevchenka ave., 1, Odessa, Ukraine, 65044

Received date 06.02.2020
Accepted date 24.03.2020
Published date 30.04.2020

1. Introduction

Steganography solves the tasks related to hiding the fact of secret data existence during their transfer, storage, or processing. A hidden message is embedded in a certain container, an object that is not attracting attention, and can be freely transmitted to the addressee. Steganographic algorithms (SAs) are widely used to prevent unauthorized access to information, as well as to embed digital watermarks into digital media in order to warrant copyright protection [1].

Convenient organization of the hidden information transfer by using existing open communication channels renders logic to the application of files of common formats as a steganographic container as the fact of their transfer would not attract unnecessary attention.

When sending the hidden information over the Internet, it may be advisable to select the static JPEG image files as a container. In contrast to text files or a webpage code, an image file can provide a larger container capacity. In this case, the JPEG format is fairly common – it is used by approximately 70.5 % of the Web sites on the Internet [2].

When converting an original image to the JPEG format (with losses), the so-called discrete cosine transform (DCT) is used. The presence of losses in conversion ensures a less noticeable distortion that occurs when embedding hidden information in a container. To hide information, the frequency domain of such images employs the SAs that use the difference of DCT coefficients to encode the bits of a steganographic message (SM), such as Koha Zhao and Hsu and Wu algorithms [1]. A SA is also applied with the modification of the smallest significant bits (SSB) of the DCT coefficients, for instance, the Jpeg-Jsteg software [1]. The use of such algorithms can provide the level of the container image distortion, which is invisible to the human eye.

At the same time, there are statistical methods of steganography analysis based on that the distribution of values for the DCT coefficients in an empty container and in a container with embedded additional information would be different. For example, hiding an SM can be discovered based on a change in the distribution of the DCT coefficients values or a change in the ratio of the number of even and odd DCT coefficients in the image.

Accordingly, the number of pairs of the DCT coefficients with certain values changes, both within the DCT coefficients block and for the pairs of coefficients in different blocks. Pairs denote here the pairwise selected DCT coefficients, the difference in coordinates of which in the frequency domain (the offset) is fixed. The analysis of the distribution of the number of coefficients pairs with certain values, obtained using different offsets, is more effective than the analysis of the distribution of the values of individual coefficients, and is used, in particular, in the Rich steganoanalytic model [3].

A steganoanalytic method based on the Rich model provides reliable detection of the fact of embedding an SM in a container for a significant number of modern SAs [4]. Therefore, it is a relevant task to build such steganographic methods that could make it possible to embed additional information resistant to the steganographic analysis based on the Rich model. In this case, it appears promising to apply machine learning technologies, as well as neural networks. This would ensure the adaptive modification of a container when embedding the message and the reduction in the change of the distribution of the number of DKP coefficients pairs, and, accordingly, the probability of detecting the hidden information.

2. Literature review and problem statement

The methods of steganographic analysis have been widely used for the detection of hidden information. For example, the steganographic analysis of JPEG images using the Rich model was suggested [3, 4]. Such an algorithm ensures the reliable detection of a steganographic message even when a container is partially filled by changing the distribution of the number of DCT coefficients pairs with different values. However, if the distribution of the DCT coefficients pairs is preserved when the message is embedded, no reliable detection could be possible. The steganoanalytic algorithm based on the difference in the imbalance of the even and odd DCT coefficients in a container [5, 6] is simpler to implement and it is faster. It was shown that such an algorithm ensures the detection of a hidden message at its embedding into the smallest significant bits of the DCT coefficients. However, if the container is filled only partially, or the steganographic method implies the correction of the DCT coefficients values when embedded, the reliable detection may not be ensured.

Thus, works [7, 8] consider the possibility to detect the fact of container modification based on the analysis of individual blocks, by using its own numbers. The proposed algorithms provide for the reliable detection of a hidden message in case the container is partially filled. At the same time, as the container filling coefficient increases, the efficiency of these algorithms may decrease.

Another proposal is the method of steganographic analysis using machine learning, which is an alternative to the Rich model usage [9, 10]. This method allows for the more efficient detection of a hidden message, compared to the Rich model-based techniques, but provides lower performance.

Many existing steganography methods have certain constraints that limit their resistance to the analysis using the Rich model or restrict their use to certain containers only. Thus, paper [11] reports the results of studying the modification efficiency of the steganographic algorithm JSTEG with the rearrangement of DCT coefficients for each block of image pixels. It is shown that the proposed rearrangement makes it possible to reduce the visible curvature of the container in comparison with the s steganographic algorithm JSTEG. However, such a modification does not ensure the preservation of the distribution of values for individual DCT coefficients, which can be found in the statistical analysis of the container. One way to reduce distortions in the distribution of the DCT coefficients may be to hide additional information by adding a noise-like message that mimics the photographic sensor noise [12]. At the same time, such a solution limits the choice of a container to photographic images. Thus, the considered steganographic algorithms do not provide the possibility of embedding a message into an arbitrary container, resistant to the statistical steganographic analysis, for example, based on the Rich model. Therefore, it appears expedient to investigate the possibility of such an embedding and develop a steganographic method that would preserve the distribution of values of the DCT coefficients.

Since the existing steganographic algorithms necessarily alter the distribution of DCT coefficients when embedding a message, it is necessary to investigate the possibility of reducing such changes. That could reduce the likelihood of detecting the fact of embedding by the methods based on an analysis of the distribution of such pairs of DCT coefficients, for example, based on the Rich model. To reduce changes in the values of the DCT coefficients, it is possible to use a preliminary container modification so that after the message is embedded, the distribution of values is approximate to the distribution of the original container. Such a method for improving the container stability to statistical steganoanalysis has not been addressed by available studies as it follows from our analysis.

Therefore, it is expedient to build a new steganographic method, which would make it possible to embed a message in such a way that the statistical analysis based on the distribution of values for the DCT coefficients pairs is not capable of ensuring its reliable detection. Thus, the new method would improve the efficiency of the hidden transfer or storage of confidential information and could be applied in the field of information security.

3. The aim and objectives of the study

The aim of this study is to build a method for embedding additional information into a container image with its preliminary modification in order to maintain the distribution of the number of the DCT coefficients pairs with certain values. This would give an opportunity to improve the resistance of hidden information against a statistical analysis of the container, thereby providing more effective protection of confidential information.

To accomplish the aim, the following tasks have been set:

– to select the neural networks and describe the sequence of activities to perform the modification of a container;

– to explore the effectiveness of detecting additional information in a non-modified container using a classifier based on the Rich model;

– to explore the effectiveness of detecting additional information in a container with the proposed modification using the Rich model-based classifier.

4. Initial materials and methods to study the resistance of a steganographic method with container modification

4. 1. The initial data acquisition and the construction of a Rich model of a container image

The algorithm for determining the distribution of values for the DCT coefficients pairs and constructing the Rich model of a container image consists of the following steps:

1. Select the original image $C_{(i,j)}$, where i, j is its size in pixels.

2. Build a two-dimensional DCT image to obtain the array (plane) of the DCT coefficients of dimensionality $i \times j$ with the coefficients' values (except for the constant components of the block) from $-Q$ to Q.

3. Quantize the DCT coefficients to obtain the integer values.

4. Build a square matrix M of dimensionality $2Q$ and fill it with zero values.

5. Find, for each non-zero coefficient of DCT $K_{i,j}$, which is not a constant component, where i, j are the coordinates of the coefficient on the plane, a pair coefficient $K_{i+x,j+y}$. Hereafter, the values (x,y) are referred to as the *offset*.

6. Determine the value of the DCT coefficients (m, n) for each pair and increase the value in a matrix cell with coordinates $M_{(m,n)}$ by 1 for each pair with the coefficients' values (m, n).

7. Since the images can have different number of pairs, after filling the matrix its values are normalized relative to the number of blocks of the DCT coefficients 8×8:

$$M_n = M / \left(\frac{i}{8} \times \frac{j}{8} \right). \quad (1)$$

8. Repeat steps 4–6 for each selected offset of the total number N.

9. Merge all the resulting square matrices into an overall R array of dimensionality $2Q \times 2QN$, which would characterize the image.

4. 2. The software and original data used in the experiment

The implementation and simulation of the operation of the selected steganographic methods, a statistical classifier, which determined a change in the distribution of the number of DCT coefficients pairs and the presence of a message, the neural networks used for the adaptive modification of the container, were carried out using the MATLAB programming environment. The quantization matrix used was the IJG standard table [13]; the image quality score Q_f=85. The steganographic message applied was a pseudo-random bit sequence with a uniform distribution of values. The containers employed were a sample of arbitrary JPEG images, acquired from the Internet. The sample size N=1,000 images.

The analyzer used was a neural network, the type of a simple perceptron with one hidden layer. The number of sensor elements is 511 (based on Q=255), the number of neurons in a hidden layer is 15. All links are with direct propagation, the number of links among the layers is 7,665 and 15, respectively. The output of the classifier yields a binary value – "1", if the container has an embedded message, and "0", if the container is empty.

4. 3. Selecting the indicators to assess the effectiveness of the proposed method

To assess the effectiveness of the proposed algorithm, the probabilities of an error of the first kind α and the second kind β [14] were used. The indicator α demonstrates the degree of probability at which the steganographic analysis of an empty container would show that it is filled. The indicator β demonstrates the degree of probability at which the steganographic analysis of a container with the message would show that the container is empty. The closer the value of β to 0.5 (the result of random guessing), the less effective the analysis, and the greater the efficiency of the proposed method. The formulae for computing α and β when using a certain set of test containers N:

$$\alpha = 1 - \frac{N_{np}}{N_{tp}}, \quad (2)$$

$$\beta = 1 - \frac{N_{ne}}{N_{te}}, \quad (3)$$

where N_{np} is the number of containers for which the classifier correctly detected the absence of a message; N_{tp} is the total number of empty containers in the set; N_{te} is the number of containers for which the classifier detected the message; N_{te} is the total number of containers with a message in the set.

4. 4. Modifying the container to reduce the statistical visibility of the fact of embedding additional information

The methodology of finding the number of pairs of the DCT coefficients with certain values while using different offsets, as well as the algorithm for constructing the Rich model of a container image and for analyzing hidden information, are described in detail in [3, 4].

To solve the task of unnoticeable embedding of additional information in the container, it is necessary to solve the issue of changing the container statistical parameters. At the same time, after embedding a preset message M in the original container C, the values of the selected predictors for the received container C_e do not exceed the recognition threshold. In addition, there should be no visually noticeable distortion of the container image. The modification of a container before embedding a hidden message should be performed in such a way that the following condition is met:

$$E(C, M) = C_e, R(C_e) \to 0, \quad (4)$$

where E is the function of embedding additional information in a container, C is the original container (image), M is the steganographic message, C_e is the steganographic container with additional information, R is the final decision on the presence of additional information in the container ("1" – additional information has been detected, "0" – additional information has not been detected).

5. Results of studying the resistance of a steganographic method with container modification

5. 1. Construction of a steganographic method with container modification

A possible way to solve the task is to create an artificial image-container C`, visually similar to C but such that the condition (4) is met for it.

To create such a container image, it is possible to use the so-called generative adversarial networks (GANs), which are widely used in image processing [15]. Structurally, a GAN consists of two separate neural networks. The first one (generator, G) is used to create artificial images based on certain input data (such as an array of pseudo-random values). After that, the second network input (discriminator, D) is iteratively sent the artificial images created by the generator and the images from the original dataset. The output of the discriminator produces the final decision whether both images that were submitted to the network input coincide. When training is completed, a GAN can create artificial images, visually similar to the original ones, whose artificial nature is almost impossible to detect without the presence of a discriminator from the GAN that was involved in creating these images.

The generator used was a neural network, the type of a "multi-layer perceptron with an error back propagation" with two hidden layers. The number of links between the layers is 102,301, 10,201, 102,301 respectively. The network described in chapter 4.2 was used as a discriminator.

In order to generate the steganographic containers for messages known in advance, a GAN circuit can be modified. Instead of pseudo-random values, the input of a generative network should be sent a container image, modified as follows:

$$C^{-1} = D(C,M), E\left(C^{-1},M\right) = C, \quad (5)$$

where C^{-1} is the modified container, C is the original steganographic container, D is the modification function of the original container C for obtaining C^{-1}, M is the steganographic message, E is the function of embedding additional information in a container.

The image containers, generated in this way, would be, after embedding additional information, closer to the original container C than when creating them based on the random data.

Fig. 1 shows the general scheme of a steganographic system employing a GAN. A more detailed scheme of the GAN is shown in Fig. 2.

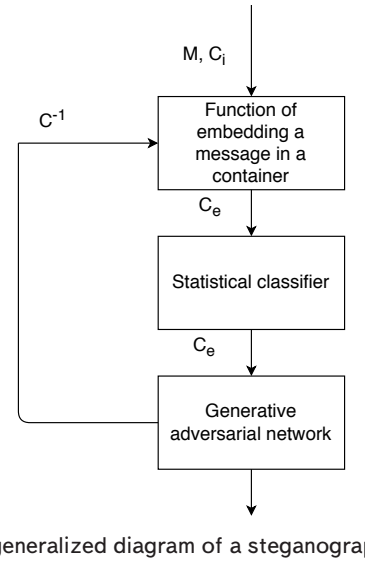

Fig. 1. The generalized diagram of a steganographic system

Fig. 2 shows that following the embedding a message in a container, the discriminator D analyzes the container with a message based on the Rich model. If we receive "0" at the output of the discriminator (the message is not detected), we believe that the message was embedded successfully and store the container for further transfer. If the output of the discriminator produces "1" (the message is detected), then we send, to the input of the generator G, the empty and filled containers to generate the modified container. Next, we embed the message into the already modified container.

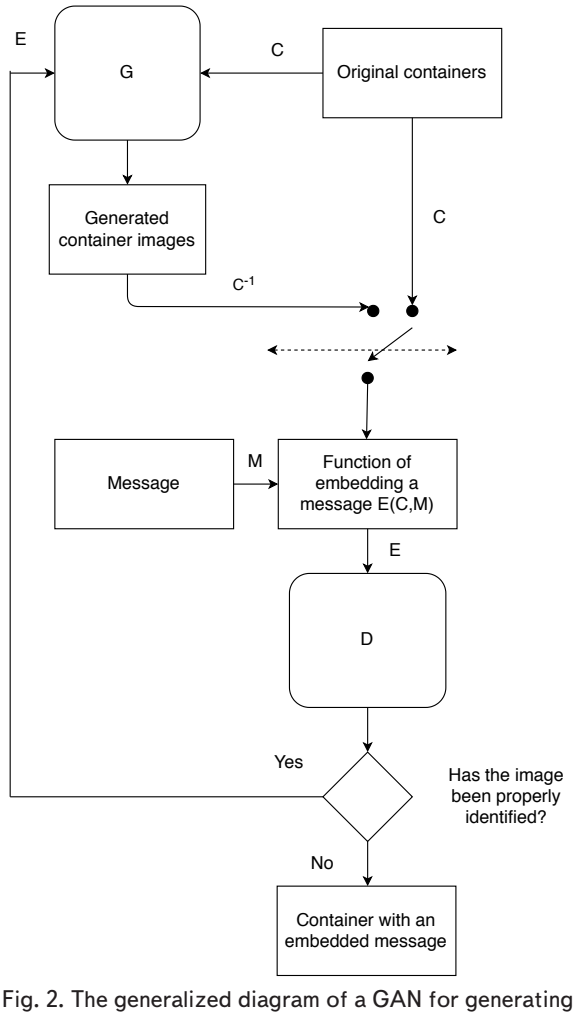

Fig. 2. The generalized diagram of a GAN for generating modified containers

The modification of a container and embedding a message involve the following:

1. Select a set of images to train the classifier. A separate set of images is required, which does not coincide with the set of containers for message embedding. Therefore, choose a separate training set of 300 images from the preliminary selected sample of 1,000 images.

2. A separate set of image containers (C_i) is chosen to embed a steganographic message.

3. Train the classifier using the set of images selected in paragraph 1, according to the algorithms described in [3] and [4].

4. The threshold values for the classifier are determined.

5. A GAN is constructed and its training is performed using a set of container image containers (C_i) for embedding a steganographic message.

6. Embed the selected steganographic message in the frequency domain of each image container. Embedding employs

the steganographic algorithm F5 [16]. The F5 embeds the bits of a message only in the smallest significant bits of the image DCT coefficients and leaves an unchanged number of the DCT coefficients of value 0. Therefore, its use would reduce the distortion of value distribution of the DCT coefficients for the container at embedding and avoid the loss of part of the message when using the JPEG format containers.

7. The trained classifier checks each container with an embedded steganographic message. If the answer is negative (the derived output value is less than the recognition threshold), the resulting container is stored with additional information for subsequent transfer using an unprotected communication channel. If the answer is positive (the classifier was able to recognize the fact of embedding additional information in the container), further steps are carried out.

8. By using the trained GAN, based on the original container C and the filled container $C`$, the modified container C^{-1} (5) is created. The modification is carried out to ensure meeting the condition (4) after embedding a message in the modified container.

9. The original message is embedded in the resulting container C^{-1} similarly to point 6.

10. Re-check the resulting container with additional information similar to point 7.

11. Check the visibility of the fact of embedding additional information into the container by means of visual inspection and the use of common container distortion indicators [1].

5. 2. Studying the efficiency of detecting additional information in the container without modification

We detected the hidden information by using the Rich model-based classifier. The starting hypothesis H_0 was the assumption that the container is empty, the alternative hypothesis H_1 was the assumption that there is a hidden message in the container. The probabilities that these hypotheses hold are given in Table 1.

Table 1

Probabilities that the proposed hypotheses hold (without container modification)

Selected hypothesis	True hypothesis	
	H_0	H_1
H_0	0.923	0.077
H_1	0.077	0.923

The results from Table 1 results demonstrate that the embedding without modifying a container ensures a high probability of the proper detection of the message. The probabilities of errors of the first and second kind are small enough ($\alpha, \beta<0.08$).

5. 3. Studying the efficiency of detecting additional information in the container with modification

We detected the hidden information by using the Rich model-based classifier. The starting hypothesis H_0 was the assumption that the container is empty, the alternative hypothesis H_1 was the assumption that there is a hidden message in the container. The probabilities that these hypotheses hold are given in Table 2.

The results from Table 2 show that the embedding with the container modification significantly reduces the probability of proper recognition. In comparison with the embedding without modification, the probabilities of errors of the first and second type are much larger (α, β ~0.35). These error values do not allow the classifier to reliably detect the fact of embedding additional information in the container.

Table 2

Probabilities that the proposed hypotheses hold (with container modification)

Selected hypothesis	True hypothesis	
	H_0	H_1
H_0	0.654	0.346
H_1	0.346	0.654

6. Discussion of results of studying the effectiveness of detecting additional information in a container with modification

The results given in Tables 1, 2 show that modifying the container can reduce the likelihood of detecting a hidden message by ~0.27. This is explained by that both the preliminary modification of the container and embedding a hidden message introduces changes to the distribution of the DCT coefficients pairs. The modification is performed so that these changes are mutually compensated for, thereby reducing the resulting changes in the container.

Thus, the proposed method ensures smaller changes in the distribution of the DCT coefficients pairs in comparison with existing ones. This makes it possible to reduce the likelihood of detecting a hidden message in the Rich model-based steganographic analysis.

However, this paper has not shown the degree of change in the probability of detecting a message in the modified container if the container is filled only partially. In addition, the disadvantages of this work include a lack of research into the efficiency of the container modification depending on the type of neural networks used and their parameters. Therefore, it is possible to further improve the efficiency of the proposed method by selecting the alternative types of neural networks for GAN and optimizing the parameters of these networks. A potential difficulty for such an improvement could be related to an increase in the computational complexity of training the neural networks and testing their effectiveness after they have been improved.

7. Conclusions

1. To pre-modify a container, we have chosen a generative adversarial network (GAN), based on 2 neural networks (a multi-layer perceptron and a simple perceptron). The sequence of activities to be performed for this modification has been described.

2. We have simulated the efficacy of hidden message detection by means of the Rich model-based steganographic analysis without the prior modification of a container. It has been discovered that such an analysis reliably detects the hidden messages in a container, with a low probability of errors of the first and second kind ($\alpha, \beta<0.08$).

3. We have simulated the operation of a steganographic method of embedding additional information with the container modification. We have also simulated the effectiveness of detecting such messages in the Rich model-based

analysis, comparing the efficiency of the detection without modifying a container and with its modification. The results indicate that the preliminary modification of a container leads to a significant increase in the probability of errors of the first and second type at detection ($\alpha, \beta \sim 0.35$). Therefore, a noticeable decrease has been ensured in the effectiveness of detecting hidden messages in the Rich model-based steganographic analysis.

References

1. Konahovich, G. F., Puzyrenko, A. Yu. (2006). Komp'yuternaya steganografiya. Teoriya i praktika. Kyiv: «MK-Press», 288.
2. Usage statistics of JPEG for websites. Available at: https://w3techs.com/technologies/details/im-jpeg
3. Fridrich, J., Kodovsky, J. (2012). Rich Models for Steganalysis of Digital Images. IEEE Transactions on Information Forensics and Security, 7 (3), 868–882. doi: https://doi.org/10.1109/tifs.2012.2190402
4. Kodovsk , J., Fridrich, J. (2012). Steganalysis of JPEG images using rich models. Media Watermarking, Security, and Forensics 2012. doi: https://doi.org/10.1117/12.907495
5. Chechel'nickij, V. Ja., Kalashnikov, N. V., Jakovenko, A. A., Kushnirenko, N. I. (2016). Container`s statistic features considering for steganographic algorithm. Electrical and computer systems, 23 (99), 83–87. doi: https://doi.org/10.15276/eltecs.23.99.2016.13
6. Chechelnytskyi, V. J., Jakovenko, A. A., Kalashnikov, N. V., Kushnirenko, N. I. (2017). JPEG statistical detection of steganographic messages. Electrical and computer systems, 25 (101), 310–316. doi: https://doi.org/10.15276/eltecs.25.101.2017.36
7. Bobok, I. I., Kobozeva, A. A. (2019). Steganalysis method efficient for the hidden communication channel with low capacity. Radiotekhnika, 3 (198), 19–31. doi: https://doi.org/10.30837/rt.2019.3.198.02
8. Kobozeva, A. A., Bobok, I. I. (2019). Method for detecting digital image integrity violations due to its block processing. Radiotekhnika, 4 (199), 130–141. doi: https://doi.org/10.30837/rt.2019.4.199.16
9. Chen, M., Boroumand, M., Fridrich, J. (2018). Deep Learning Regressors for Quantitative Steganalysis. Electronic Imaging, 2018 (7), 160-1–160-7. doi: https://doi.org/10.2352/issn.2470-1173.2018.07.mwsf-160
10. Boroumand, M., Fridrich, J., Cogranne, R. (2019). Are we there yet? Electronic Imaging, 2019 (5), 537-1–537-13. doi: https://doi.org/10.2352/issn.2470-1173.2019.5.mwsf-537
11. Sheisi, H., Mesgarian, J., Rahmani, M. (2012). Steganography: Dct Coefficient Replacement Method andCompare With Jsteg Algorithm. International Journal of Computer and Electrical Engineering, 4 (4), 458–462. doi: https://doi.org/10.7763/ijcee.2012.v4.533
12. Denemark, T., Bas, P., Fridrich, J. (2018). Natural Steganography in JPEG Compressed Images. Electronic Imaging, 2018 (7), 316-1–316-10. doi: https://doi.org/10.2352/issn.2470-1173.2018.07.mwsf-316
13. Independent JPEG Group. Available at: http://www.ijg.org/
14. Oshibki I i II roda pri proverke gipotez, moshchnost'. Available at: http://statistica.ru/theory/oshibki-pri-proverke-gipotez-moshchnost/
15. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S. et. al. (2014). Generative Adversarial Networks. arXiv.org. Available at: https://arxiv.org/pdf/1406.2661.pdf
16. Westfeld, A. (2001). F5 – A Steganographic Algorithm. Lecture Notes in Computer Science, 289–302. doi: https://doi.org/10.1007/3-540-45496-9_21