SIMULATION MODELLING OF ARTIFICIAL NEURAL NETWORKS FOR THE PURPOSES OF STEGANALYSIS

*Yoana Ivanova**

Department of Telecommunications, New Bulgarian University
Bulgaria

* Corresponding Author, e-mail: yivanova@nbu.bg

Abstract: This paper is considered to be a continuation of previous publications devoted to applications of simulation modelling in cybersecurity and in particular the use of Artificial Neural Networks (ANNs) with Backpropagation of Error for digital recognition (IJITS, №4, 2021). It aims to present advanced steganographic methods that are applicable for the purposes of steganalysis to ensure the information protection. The empirical study and the experimental analysis made contribute to explaining various steganographic approaches that are applicable in security.

Key words: steganography, steganalysis, histogram equalization, LSB, Virtual Steganographic Laboratory, Artificial Neural Networks.

1. INTRODUCTION

In cybersecurity cryptography and steganography are applicable separately or jointly depending on the main goal – the text in a message to be transformed into an unintelligible sequence of characters by symmetric or asymmetric cryptographic algorithms and hash functions, as well as the message itself to be hidden by embedding in raster images, vector graphics and multimedia files using specialized steganography software.

"Steganography is the art of secret communication in order to exchange a secret message" [1] hidden into a text or a multimedia object. There are also steganographic approaches expressed in use the free space in the storage devices to hide information. For example, the protocol suite TCP/IP is characterized by "*unused space in the packet headers"* – the TCP packet header has 6 and the IP packet header– 2 reserved bits [2].

Actually, steganography and steganalysis are two separate but interdependent fields because steganography aims to hide digital data into a text, computer graphics or multimedia *"without depiction of any misgiving" [3],* while steganalysis aims to detect the presence of steganography. Because steganography is not always a mean of strengthening cybersecurity. It could be also a serious threat about it in the hands of malicious individuals. There are advanced techniques for detecting the presence of hidden message in a raster or vector digital image [4] such as Artificial Neural Networks (ANNs) and steganalysis based on different levels of luminance in the image [5].

The content of this paper consists of the following sections:

- **Section 2** includes a basic classification some of the widely used steganographic techniques and methods of steganalysis.
- **Section 3** presents the experimental part of this empirical study which is expressed in steganographic analysis using the software Virtual Steganographic Laboratory. It includes summary evaluation and analysis of the results supported by examples of other methods applicable in LSB steganography.
- **Section 4** represents extended research in a simulation environment SIMBRAIN related to proposing a method for building an ANN with Backpropagation of Error for detection and recognition of raster stego-images.

2. ADVANCED STEGANOGRAPHIC TECHNIQUES AND METHODS OF STEGANALYSIS

Steganography and steganalysis are the two major parts of steganology that ensures the information security and especially the protection of confidential information from unauthorized access. Actually, steganography is subdivided into image, video, audio and text depending on the data. A basic classification of the digital image steganography based on main techniques is:

- ***LSB*** – is also known as right-most bit or this is the lowest bit in a binary code. The preliminary encrypted text can be hidden in a raster image by a substitution of the lowest bit of the image with bits of the text. Bur this technique is vulnerable to attacks, because it is basic and widely used.
- ***MSB*** – its principle is analogous to that of LSB, since in this case the left-most bit is replaced by bits from the encrypted message.
- ***BPCS (Business Planning and Control System)*** – the digital image can be divided into two regions: *"informative region"* and *"noise-like region"*. The integrity of the encrypted message is saved although it can be hidden in all bits of *"the noise-like region"* [6].

For example, the joint application of LSB and MSB is one of the possibilities to strengthen security. But this case should be considered on the other side as well due to the potential risk for security if the steganographic techniques are used by attackers. In terms of steganalysis is necessary the attacks which determine the possible methods of steganalysis to be classified as follows:

- ***Stego-only attack*** – the stego-image is the key object of analysis.
- ***Known cover attack*** – the original and the stego-image are available in the software for steganalysis.
- ***Known message attack -*** the secret message is known for the attacker who analyzes the stego-image for patterns, because of his aim to determine the steganographic algorithm.
- ***Known stego attack –*** except that the algorithm is known, the original and the stego-image are available for steganalysis.
- ***Chosen stego attack*** – the algorithm and the stego-image are known and available for analysis.
- ***Chosen message attack –*** the approach used by the steganalyst is expressed in generating a stego-image from a steganography algorithm which aims to detect corresponding patterns in the stego-image [7].

3. CONDUCTING EXPERIMENTS IN A SOFTWARE ENVIRONMENT FOR STEGANOGRAPHIC ANALYSIS

For this part of the research the author has chosen a Java-based software environment Virtual Steganographic Laboratory (VSL) which is a *“graphical block diagramming tool. It allows complex using, testing and adjusting different steganographic techniques and provides GUI along with modular, plug-in architecture” [8].* Some of the main functional modules in VSL are explained based on the Global Information Assurance Certification Paper of SANS institute [9] as follows:

- ***Noise with gaussian distribution*** – *“uniform noise inserts colored pixels that closely resemble the pixels in the original image.”*
- ***Gaussian blurring*** – softens significant transitions and reduces contrast *“by averaging the pixels next to the hard edges of defined lines and areas”.*
- ***Sharpen –*** the addition of a sharpening mask is displayed as the opposite of the blurring effect.
- ***Cropping of an image –*** if a region of pixels is cut from the stego-image the message could not be repaired [10].
- ***Salt (white) and Pepper (black) noise.***
- ***Image resizing.***
- ***Median filtering.***
- ***Gamma correction.***
- ***JPEG compression.***

Actually, the role of the distortions in VSL shown on the left side of Fig. 1 is to make the message unusable. The other used modules in this empirical research are selected on the right side of the same figure.

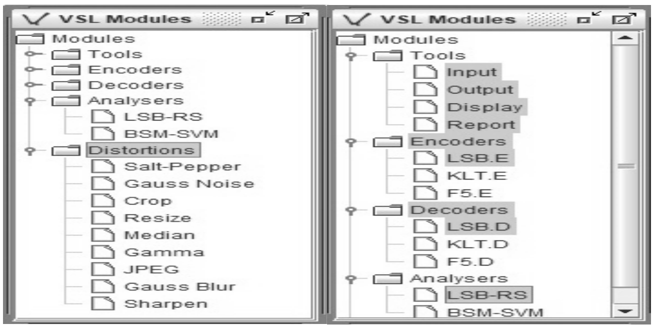


Fig. 1. The functional module "Distortions: in VSL.

The principle of work of such products is based on invisible distortion in the original image. It should be noted that these modifications can be detected by other tools for identify the malicious source. If the noise regions predominate in a digital image, then this is the first symptom that it has been modified. In this case the change in the image under LSB is related to the disturbing the smooth transitions between colors. Those pixels that are no longer next to other pixels with a color of the same palette will be visualized as a noise region. One of the effective methods used by developers of steganography software is to cover the color images with 256 gray-scale images, because the color change increases gradually instead of sharply.

A basic example of displaying a digital image before and after LSB steganography is shown in the first scheme in Figure 2. The second example is improved by distortions (***Gauss Noise*** and ***Image Resizing***), LSB analysis and generating a report. The third scenario represents an example of the combination of ***Salt-Pepper noise*** and ***Crop*** distortions. Before starting the experiment, it is required to import a digital image into each input and to specify a message for every steganographic encoder.

In all scenarios the same digital image in grayscale mode saved in JPEG file format is imported, as well as the same 48-bits message "AVATAR" saved in a TXT file format and encrypted by 7-zip is embedded. In Fig. 3 are shown the original and the stego-images as a result of LSB in Scenario 1 and 2.

A cut region of the original image as a result of Scenario 3 (Fig. 4) appears to be a segment of its white background with added noise. The experimental results obtained by the functional modules for report and analysis are originally generated in a CSV file directly from the software. The items to report are: *Number of iterations, Number of Input, Module IN, Module OUT, Image filename, Image size, Message size, Peak signal-to-noise ratio* and etc.

It should be noted that in Scenario 1 the algorithm is unknown, but the block diagram includes two blocks for displaying the original and the stego-object, which can be analyzed under the impact of a known cover attack. There is no option the original object to be displayed in Scenarios 2 and 3, because this is an attack of the type stego-only.

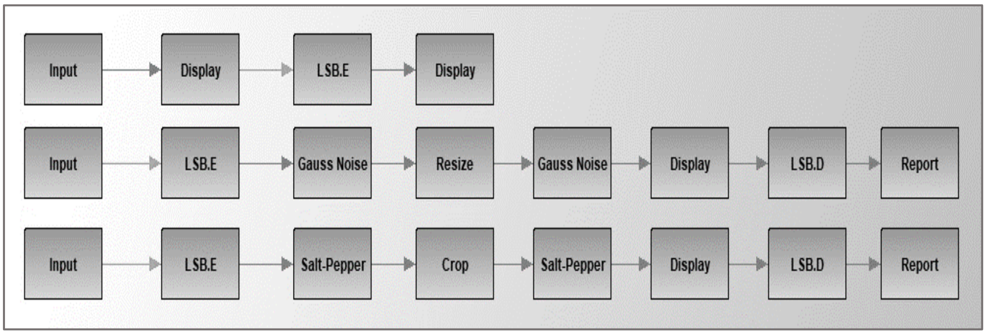


Fig. 2. Various scenarios of LSB steganographic embedding using VSL.

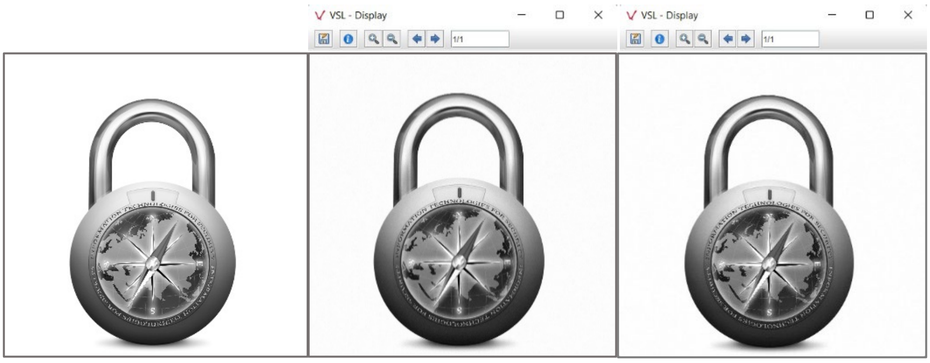


Fig. 3. A comparison between the original image and the stego-images in Scenarios 1 and 2.

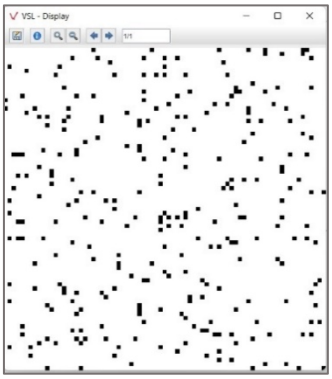


Fig. 4. The stego-image in Scenario 3.

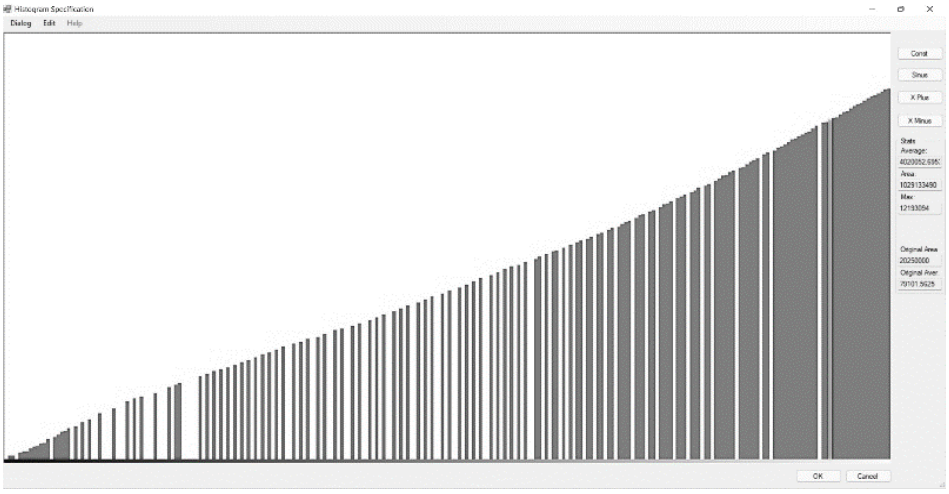
3.1. Assessment and analysis of the experimental results

An extended summary evaluation of the experimental results obtained can be made on the base of the visual information contained in the images in Fig. 3. The first of them is the same one imported in the block diagrams in Fig. 2. Obviously, while its luminance is lower than the other two, their contrast is stronger.

As it is known, the essence of this experiment is studying the impact of various attacks realized by the distortions on the stego-media and in this sense the barely noticeable change could be determined as a result of the integrated impact of the distortions selected in Scenario 1 and 2 if it was not established in scenario 1 already. In this case it should be noted that this stage precedes the application of additional distortions when in the block diagram is placed only the LSB encoder.

Actually, there is a practice the LSB steganography to be improved by histogram methods in order to ensure resilience of the stego-media against steganalysis. For example, the method of histogram equalization is characterized by image contrast enhancement and could affect the contrast of a stego-image.

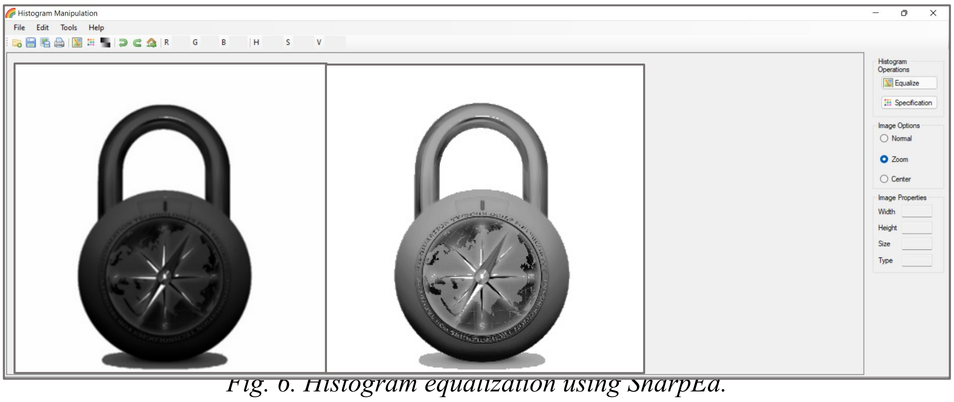
Although the histogram equalization cannot be directly selected from the list of distortions in VSL, in general it is possible to be used in the algorithms embedded in specialized software for steganography and steganalysis. Normally, this modification is barely noticeable with a human eye like in the images in Figure 5. This algorithm aims to distribute the pixel intensities evenly in the grayscale range as in this way regions with lower local contrast are transformed into regions with stronger local contrast.



Stats
Average:
4020052.695
Area:
1029133490
Max:
12193094
Original Area
20250000
Original Aver
79101.5625

Fig. 5. An example histogram obtained in SharpEd.

For the purposes of a comparative analysis the author has chosen a specialized software for histogram equalization SharpEd (HistogramManipulation) to enhance contrast adjustment. The tool is suitable for image analysis using RGB or HSV (hue saturation value) histograms. In the histogram shown in Fig. 5 the pixels intensities are along the x axis and the number of pixels is along the y axis. The results of the direct application of histogram equalization to the image using the software's default settings and after a random change in the histogram in Fig. 5 are shown in Fig. 6.



○ Normal
● Zoom
○ Center

Fig. 6. Histogram equalization using SharpEd.

The mathematical algorithm of this method is demonstrated by an example of histogram equalization of 3-bits image [11]. The main parameters are explained and calculated in Table 1. The histogram equalization can be calculated by the formula:

$$G(r) = \left[\frac{T(r)}{n} L\right] - 1 = \left(\frac{L}{n}\right) T(r) - 1 \tag{1}$$

Table 1.

Parameters	***Designation***	***Results***							
Number of shades of the grayscale	*L = n^k^*	8							
Number of bits	*k*	3							
Number of pixels	*n*	0	1	2	2	3	3	4	4
		0	1	2	2	3	3	4	4
		0	1	2	2	3	3	4	4
		0	1	2	2	3	3	4	4
Types of shades of the grayscale	*r*	0	1	2		3		4	
Frequency (intensity) of each shade of the grayscale	*F (r)*	4	4	8		8		8	
Sum of all shades of the grayscale	*T (r)*	4	8	16		24		32	
The image after histogram equalization	*G (r)*	0	1	3		5		7	

The conclusion is that although the logical analysis made is based on plausible conditional assumptions, the unknown algorithm may be a constraint in terms of an unambiguous and accurate quantitative and qualitative analysis of the risk from a specific type of stego-attack. Therefore, it is recommended to be used ANNs with Backpropagation of Error. They are a powerful tool which is suitable to detect and recognize by a comparison between the original and the stego-images. In the security area these networks train themselves to recognize various symbols, segments of QR-codes and etc. Therefore, the current research can continue in this direction, using the method of simulation modeling in the first stage, preceding the physical implementation of networks.

In a sense this is the opposite approach to the one described in the previous Section 2, because the algorithm is known, but the encrypted message is not. The two experiments illustrate two different sides of the same study for greater clarity which is the main contribution of the joint application of the two methods of steganalysis.

4. A METHOD OF SIMULATION MODELLING OF ARTIFICIAL NEURAL NETWORKS FOR “STEGO-IMAGE” RECOGNITION

This method upgrades the basic algorithm presented in the previous paper of the author in IJITS for modelling the ANN because of additional steps included:

- ***choice of a steganographic technique*** – in relation with the previous experiment described in Section 3 this must be LSB steganography. The main difference is that in this case is known how this steganography changes the original image in a stego-image, because not only the original object and the stego-media are both available, but also the mechanism of changing the binary code as a result of LSB steganography. Consequently, this kind of attack can be classified as a known stego- instead of a known-cover attack.
- ***building a reference model of an ANN*** – this step is important for verification and validation of the simulation model in order to prove that the ANN is self-learned to recognize the original raster image filled or only contoured regardless of its spatial orientation and the optimization of the matrix of pixels which is recommended due to the need of reducing the number of the input neurons.
- ***modelling a separate ANN for “stego-image” recognition*** – this network must be able to establish that the stego-image is different compared to the original one even if only the right-most bit of it is replaced by bits of the encrypted message embedded. It is conditionally accepted that the correct feedback is obtained if the error is higher compared to the error in the reference model although the same parameters of ANN and matrix of pixels of the original and the stego-image.

The main spatial orientations of the original image after 4 rotations of 90° clockwise are shown in Fig. 7 based on which the binary codes are filled in the table Input data of the reference ANN in SIMBRAIN. The configuration of the reference ANN includes 64 input neurons, two hidden layers with 4 neurons each and 2 output neurons respectively for a filled and a contoured image. To each pixel of the matrix corresponds a neuron of the ANN (Fig. 8)

It should be noted that SIMBRAIN cannot be used directly for steganographic embedding of an encrypted message in the digital image, but the author makes a conditional assumption that the message is unknown and only the original and the stego-object can be analyzed in the simulation environment.

Actually, the second ANN for “stego-image” recognition compared to the filled and contoured original image has absolutely the same configuration, but the least significant bit of the stego-image is changed from 0 to 1 for its 4 spatial orientations as it is shown on the right side of Fig. 7 and in the relevant table Input data in Figure 9. In the first spatial orientation of the stego-image the least significant bit is relevant to pixel 64th and after the 3 rotations its position is changed respectively to 57th, 8th and 1st.

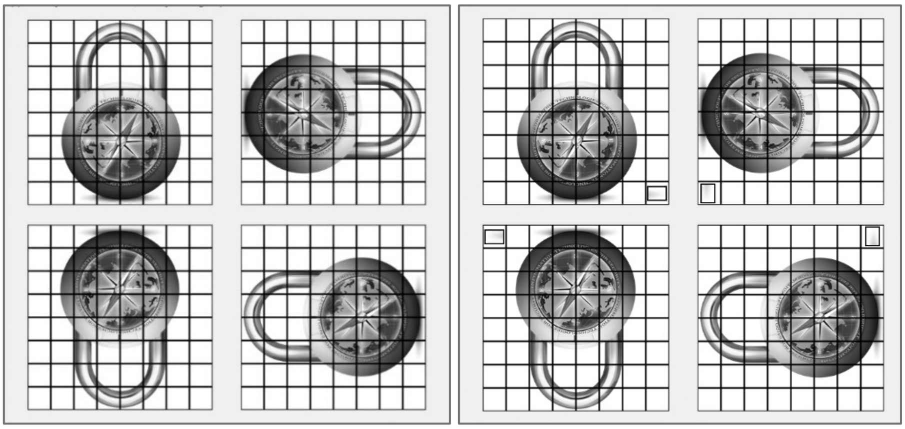


Fig. 7. Main spatial orientations of the original image and the stego-image represented in a matrix with a size 8x8.

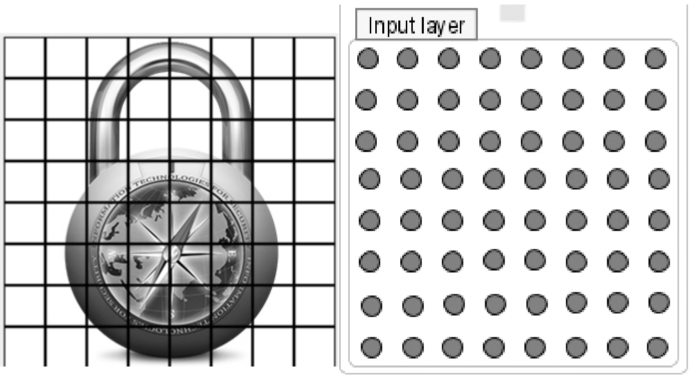
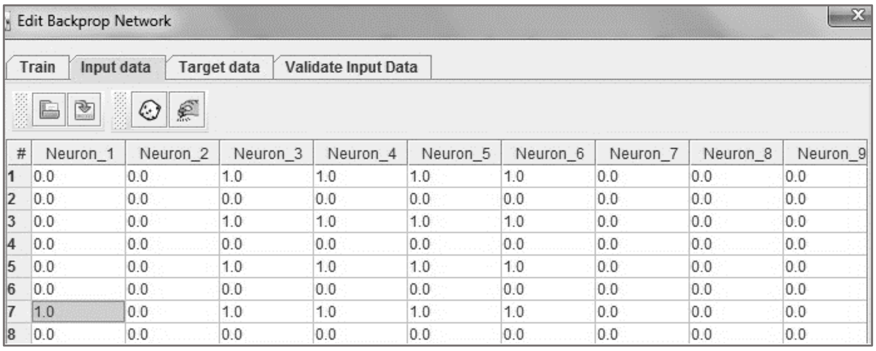


Fig. 8. Correspondence between the raster image matrix and the Input layer of the ANN.

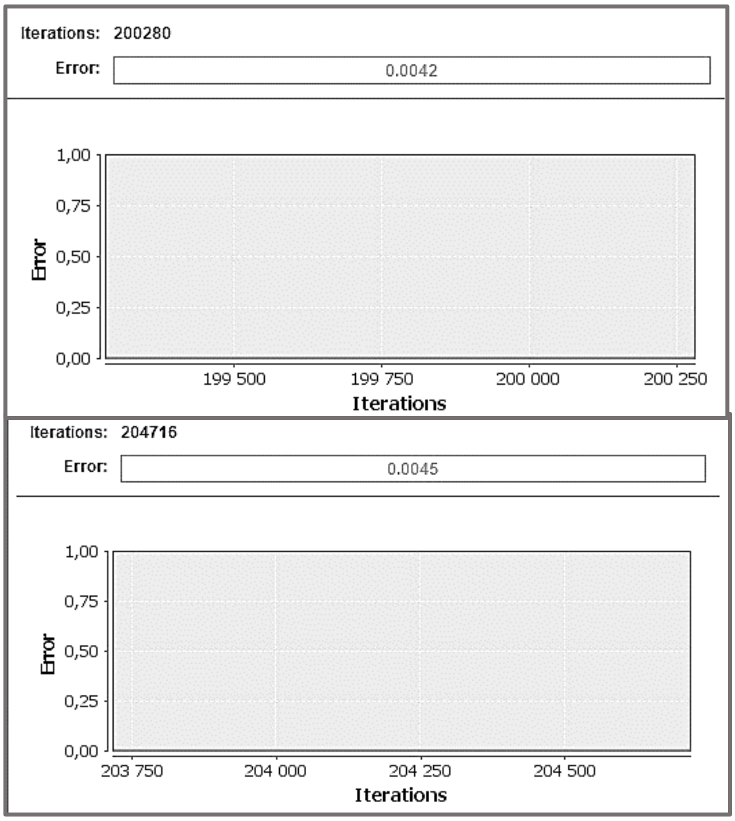


#	Neuron_1	Neuron_2	Neuron_3	Neuron_4	Neuron_5	Neuron_6	Neuron_7	Neuron_8	Neuron_9
1	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	1.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Fig. 9. Exampled binary codes inserted in the table Input data for the first 9 neurons.

Assessment and analysis of the simulation results

In Fig. 10 are shown selected screenshots taken during the simulation process respectively in the reference ANN and after the substitution of the bits from the original image with the bits of the conditionally determined stego-image.


Iterations: 200280
Error: 0.0042

Iterations: 204716
Error: 0.0045

Fig. 10. A screenshot taken during the simulation process in the reference model of ANN and the ANN for "stego-image" recognition.

As can be concluded the expected simulation results are achieved by the method proposed, because in the two ANNs the values of the error are within the range, but there is still a minimal difference in the results obtained. After a potential application of LSB steganographic technique the network still recognizes the filled or contoured image with a comparatively low error like in the reference ANN, but it looks as though it detects an unknown component. Therefore, the network continues to self-study even after more iterations than in the reference ANN while the error is still higher.

In a sense this is an example of an opposite approach to the one described in the previous Section 2, because the algorithm is known, but the encrypted message is not. The two experiments illustrate different sides of the same study for greater clarity which is the main contribution of a joint application of these methods of steganalysis.

5. CONCLUSION

In conclusion, it can be said that although the variety of steganographic tools, all steganographic methods *"modify the statistical characteristics of digital images"*. Steganalysis is considered as *"a classification problem"* which can be solved by using ANNs [see source 1].

The union of compression, encryption and steganography can be a powerful mean of ensure information protection. For example, the digital watermarking is an effective technique for protection of copyrights in case that an image may be used by third parties *"without the permission of the author"* [12].

In addition, steganalysis contributes to cybersecurity reducing potential risks and vulnerabilities related to steganographic techniques used. Simulation modelling is a reliable method for various studies in the area of cybersecurity, included steganalysis using models of ANNs. Besides, the ANN' are very effective in detecting intrusions [13].

This research can be improved by exploration of compatible products for steganalysis and simulation of ANNs in order to an unambiguously interpretation of the experimental results for further use in evaluation and analysis.

REFERENCES

[1] Yedroudj, M. *Steganalysis and steganography by deep learnin*g. University of Montpellier, France, 2019 (NNT: 2019MONTS095).

[2] *Paladion, High Speed Cyber Defense, Steganalysi*s. Available at: https://www.paladion.net/blogs/steganalysis (visited on 20.03.2022).

[3] Rashid, K., A. Rashid, N. Salamat, M. M. S. Missen. Experimental Analysis of Matching Technique of Steganography for Greyscale and Colour Image, *International Journal of Computer Science and Information Technology*, No. 6, 2014, pp. 157-166.

[4] Kashyap, R., M. Ganesh, Securing Information for Commercial File Sharing by Combining Raster Graphic and Vector Graphic Steganographies, *International Journal of Engineering and Advanced Technology (IJEAT)*, ISSN: 2249-8958 (Online), Vol. 8, No. 6, 2019, pp. 788-795.

[5] Taburet, T., L. Filstroff, P. Bas, W. Sawaya. An Empirical Study of Steganography and Steganalysis of Color Images in the JPEG Domain. *IWDW, International Workshop on Digital Forensics and Watermarking*, Jeju, South Korea, 2018. Available at: https://hal.archives-ouvertes.fr/hal-01904482/document (visited on 12.03.2022).

[6] Khaire, S. S., S. L. Nalbalwar, Review: Steganography – Bit Plane Complexity Segmentation (BPCS) Technique, *International Journal of Engineering Science and Technology,* Vol. 2, No. 9, 2010, ISSN: 0975-5462, pp. 4862-4868.

[7] Gupta, R., Information Hiding and Attacks: Review, *International Journal of Computer Trends and Technology (IJCTT),* 2014, ISSN: 2231-2803, pp. 21-24.

[8] Węgrzyn, M., Virtual Steganographic Laboratory for Digital Images (VSL) Free tool for steganography and steganalysis, *SourceForge,* 2011. Availabe at: http://vsl.sourceforge.net/ (visited on 20.03.2022).

[9] Bartel, J., *GIAC certifications, Global Information Assurance Certification Paper,* Copyright SANS Institute, 2002. Available at: https://www.giac.org/paper/gsec/707/steganalysis-overview/101589 (visited on 24.03.2022).

[10] Johnson, N. F., Z. Duric, S. Jajodia, *Information Hiding: Steganography and Watermarking – Attacks and Countermeasures*, Boston, Massachusetts: Kluwer Academic Publishers, 2001.

[11] Chapter 3: *Histogram Equalization*, 2020. Available at: https://www.youtube.com/watch?v=5-7xskk3aeo (visited on: 25.03.2022).

[12] Hsu, C., S. Tu, Digital Watermarking Scheme for Copyright Protection and Tampering Detection. *International Journal on Information Technologies and Security*, ISSN 1313-8251, Vol. 11, No 1, 2019, pp. 107-119.

[13] Mustafaev, A. G., Application of Artificial Neural Networks in the Intrusion Detection System, *International Journal on Information Technologies and Security*, ISSN 1313-8251, Vol. 10, No 4, 2018, pp. 57-66.

Information about the author:

Dr. Yoana A. Ivanova – A chief teaching assistant in the Department of Telecommunications at NBU; Area of scientific research: Applications of Information Technologies in Security, Communication and Information Systems and Technologies in Security; Professional area: 5.3. "Communication and computer equipment"; Doctoral Program: "Automated Systems for Information processing and Management”.

Manuscript received on 07 April 2022