Research of the characteristics of a steganography algorithm in images when using different alphabet

Veselka Stoyanova
Artillery, AD and CIS faculty
National Military University Vasil Levski
Shumen, Bulgaria
veselka_tr@abv.bg

Abstract. **Steganography can be defined as a method of hiding data in cover media so that others are not aware of its existence. Steganographic systems play an important role in the covert transmission of information even in the presence of a steganalyser. The article deals with the steganography system which hides text inside images without losing data in components of RGB model. The secret message is hidden in the cover image using Least Significant Bit algorithm. The statistical characteristics of stego-images with the embedded information in Cyrillic and Latin are investigated.**

The aim of the study is to determine whether there is a change in the qualitative characteristics of the stego-image, when it is hidden the same information, but was used different alphabets. The comparative results for the proposed algorithm are very promising for Cyrillic alphabet. To evaluate steganography system properties are used the measures like Signal-to-Noise Ratio, Peak Signal-to-Noise Ratio, Mean Squared Error and Structural Similarity Index for measuring.

Keywords: Alphabets, information hiding, Steganography, text hiding

I. Introduction

Nowadays in the world of information technologies static and unprotected data transmission is tantamount to suicide. There are many methods by which organizations can protect themselves and to certify their right to use the transmitted confidential information. An increasingly common way to protect transmitted information is its invisibility. The use of steganography, in close connection with cryptography and securing with static passwords in the authentication process protects against the high risk of information security.

Steganography conceals the existence of secret information in the cover carrier [1]. Steganography can use several tapes of cover media (i.e., audio, video, image, text and network. According to [2] in Bulgaria for 2023г. 45,1% use email and 65,4% use social media or implement real-time messaging (Viber, WhatsApp, Messenger, Snapchat, Skype, Discord, Telegram), where the main communication is realized in Cyrillic. The situation in the EU is similar, 84% users [3] have online activities of Internet. They could use mobile steganography systems or apps.

The Cyrillic alphabet to be use by approximately 250 million people. Cyrillic alphabet is the 6th most popular writing script on the planet and used across 50 languages. The Cyrillic has been the third official alphabet of the European Union alongside the Latin and the Greek alphabets [4].

II. Materials and Methods

A. *Basic characteristics of steganography algorithms evaluation*

The stego-image characteristics of the presented LSB-based method enable its use by users in different languages around the world. This article examines and compares hidden information in Cyrillic and Latin.

The two authors from [5], tell us that the stego-file can be attacked in two ways: A visual attack and statistical attack. When we do the visual analyze we uses the human vision to detect the differences between the original object from the stego-object, whereas the statistical analyze using steganalysis algorithms based on mathematical theories [6].

When comparing two images four major statistical properties which describe the degree of similarity between the images are calculated: *MSE* (Mean Squared Error), *PSNR* (Peak Signal-to-Noise Ratio) and images entropy. Matlab [7] is so easy to calculate the differences between the original object from the stego-object. Ther are statistical function: MSE, PSNR and entropy. In [8] was presented the experiments in Matlab which are carried out with the fuzzy logic tool.

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2024vol4.8236

The characteristics studied are represented by formulas (1) and (2), the PSNR is based on values obtained for the MSE:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2, \qquad (1)$$

Where m and n are the width and height of the image; I (i, j) and K (i, j) are relevant pixels with coordinates (i, j) in the original stego-image.

$$PSNR = 10.\log_{10}(\frac{max^2}{MSE}) = 10.\log_{10}(\frac{max}{\sqrt{MSE}}), \qquad (2)$$

where MAX = 255 for 8 bit images.

The degree of similarity of the images before and after the embedding of the data, measured by the *MSE* and the *PSNR*, determines the quality the stego-image [9], [10].

Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image. The entropy of an image can be calculated by calculating at each pixel position (i,j) the entropy of the pixel-values within a 2-dim region centered at (i,j).

$$entropy = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p(i,j) log \log_b p(i,j) \; □ \; (3),$$

where ***n*** and ***b*** are again the number of gray levels and the base of the logarithm function, respectively, and ***p(i, j)*** stands for the probability of two pixels separated by the specified offset having intensities *i* and ***j***.

In Matlab we can use function to calculate the entropy:

I = imread('lena.bmp');

J = entropy(I)

In [11] is said that the most important evaluation criteria for steganography algorithms are invisibility, capacity, robustness and security. They are presented in fig. 1.

Some of these criteria can be evaluated by calculation while others can be visualized [12].

The main difference between Steganography and cryptography is that, cryptography concentrates on keeping the contents of a message secret while steganography concentrates on keeping the existence of a message secret [13].

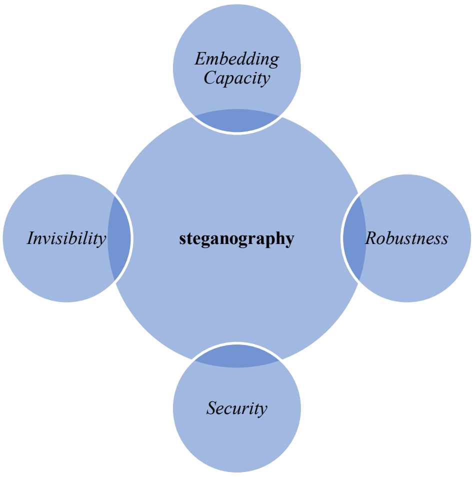


Fig.1. Evaluation criteria for text steganography algorithms

B. Steganography algorithm based on the LSB method of embedding Cyrillic and Latin information in images.

Least Significant Bit (LSB) replacement is the process of adjusting the most significant bits of the pixels of the cover image [10]. More details about how work LSB could be found in [14] - [19].

Proposed steganography algorithm for embedding information in images uses the last significant bits (LSB) in each color channel pixels. How this works presented on fig.2.

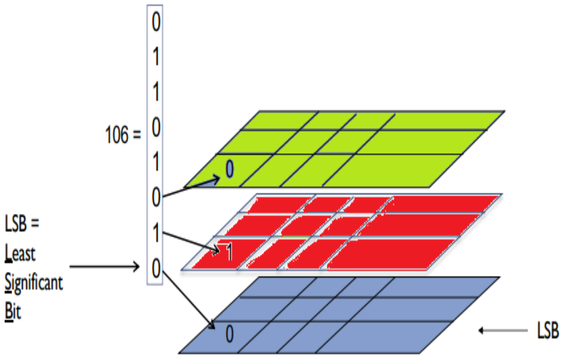


Fig.2. LSB Replacement embedding in color image

Figure 3 is a general block diagram of the algorithm, in which is checked whether there is a data entry or retrieval.

Thus is checked what operation will be realized and is proceeded to the incorporation or extraction of the confidential information. This is the main communication scheme in the steganographic environment.

The image thus extracted at the receiver's end is the same as the original image without any pixel value difference.

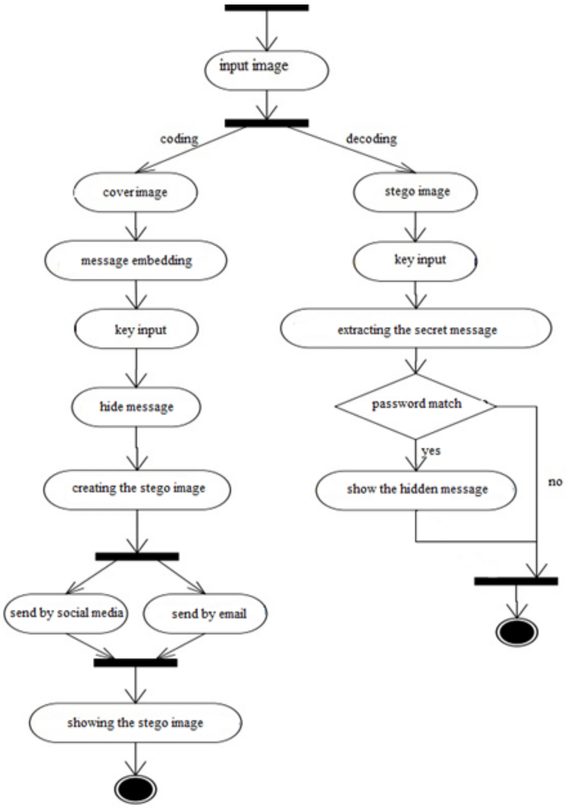


Fig.3. Communication scheme in the steganographic environment

III. Results and discussions

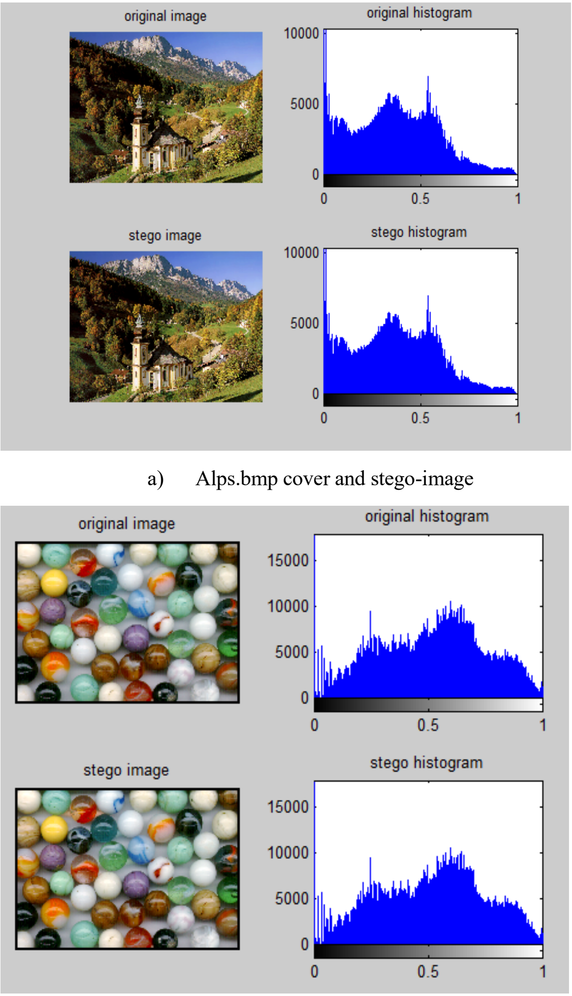
One of the main aims of the realized research of the steganography system is its practical applicability and evaluation in terms of the invisibility and size of the hidden data in the cover images. A comparison is implemented with respect to the type of alphabet used to create the secret message. Research has shown that when comparing the statistical characteristics of the same stego-images, the same secret text, but with a different alphabet, specifically Latin or Cyrillic, there is a difference in the compared statistical characteristics. On visual analysis, differences of this kind cannot be detected.

TABLE 1. HIDE 4KB LATIN AND CYRILLIC TEXT IN DIFFERENT IMAGES

Original image	Text type	MSEav	SNR	PSNR	entropy
Alps	lat.	0.0014	62.265	69.8157	7.5557
Alps	cyr.	0.0013	62.4801	70.0309	7.5549
Paradise	lat.	0.0011	60.1414	69.8151	6.456
Paradise	cyr.	0.00098	60.1312	69.8050	6.466
change	lat.	0.0012	64.1293	69.9402	4.9676
change	cyr.	0.0015	63.8985	69.7094	4.9698
Marbles	lat.	0.0013	66.9616	72.1563	6.9900
marbles	cyr.	0.0012	67.6274	72.8221	6.9874
Ice	lat.	0.0012	61.538	69.7777	6.5832
ice	cyr.	0.0013	61.5217	69.7613	6.5833
snow	lat.	0.0012	63.5232	69.7424	7.2032
snow	cyr.	0.0011	63.5473	69.7664	7.2032
Tahaa	lat.	0.0013	64.3824	69.7678	7.7233
Tahaa	cyr.	0.0011	64.4358	69.8212	7.7232

Table 1 presents the results of the qualitative characteristics of embedded text in Cyrillic and Latin with a size of 4 kB and seven cover digital image is used. Figura 4 a) - b) and figura 5 coresponded with table 1.

In Figure 4 can be seen histograms of original and stego-image obtained by embedding of 4 kB information at base settings of the steganography algorithm, i.e. successively embedding in the three-color components of pixels without using protection by stego-key in an Alps.bmp cover image (a) and Marbles.bmp cover image (b).



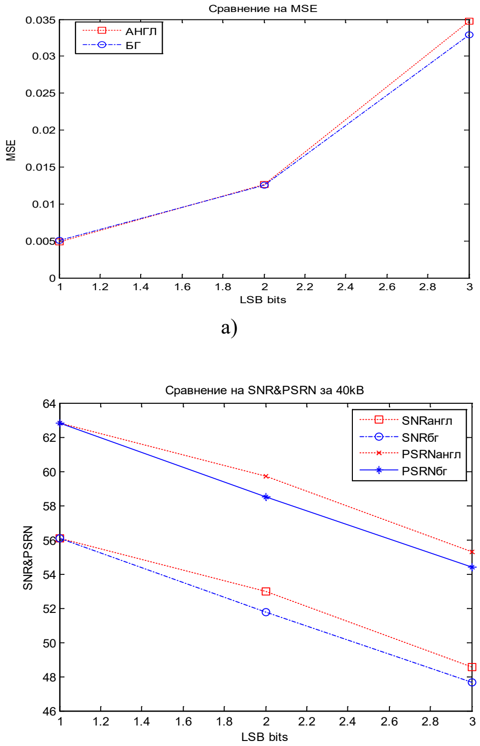
a) Alps.bmp cover and stego-image

b) Marbles.bmp cover and stego- image

Fig.4. Histogram of original and stego-image, which has 4 kB of embedded information

This image when compared with the payload does not reveal any differences in image quality and pixels.

Table 2 presents the results of the qualitative characteristics of embedded text files in Cyrillic and Latin with a size of 22b to 2 kB and cover digital image Lena.bmp is used. Matlab compares the statistical characteristics discussed earlier in the report (1) and (2).



a)

b

Fig.5. Comparison of hiding 4 kB information in the LSB of Latin and Cyrillic with indicators of (a) MSE и (b) SNR , PSNR by Tahaa.bmp

TABLE 2. THE RESULTS OF HIDING DIFFERENT SECRET INFORMATION IN LSB IN THE IMAGE (LENA.BMP) WRITTEN IN LATIN AND CYRILLIC

Original image	**Text size**	**MSE lat**	**MSE cyr**	**SNR lat**	**SNR cyr**	**PSNR lat**	**PSNR cyr**
Lena1	22b	$1.3987e^{-5}$	$2.5431e^{-5}$	82.5547	79.8547	87.6922	84.9923
Lena2	127b	$9.7911e^{-5}$	$1.7293e^{-4}$	75.0657	71.9935	80.2032	77.1311
Lena3	170b	$1.4750e^{-4}$	$2.848\ e^{-4}$	73.1877	70.0465	78.3252	75.1841
Lena4	300	$2.4033e^{-4}$	$4.3360e^{-4}$	71.3278	68.1009	76.4654	73.2385
Lena5	601b	$5.1880e^{-5}$	$8.9518e^{-4}$	68.1496	65.0523	73.2871	70.1899
Lena6	903b	$7.8837e^{-4}$	0.0019	66.3989	62.0186	71.5365	67.1561
Lena7	1.17kB	0.0011	0.0028	65.1329	60.2761	70.2705	65.4137
Lena8	1.46kB	0.0013	0.0037	64.1575	59.0956	69.2951	64.2332
Lena9	1.76kB	0.0016	0.0046	63.3459	58.1248	68.4834	63.2624
Lena10	2.05kB	0.0019	0.0055	62.6931	57.3308	67.8307	62.4684

CONCLUSION

In this paper, detailed security analysis has been provided on the novel algorithm using visual inspection, histogram analysis, mean squared error and peak signal-to-noise measure.

The results from the research can be summarized in the following conclusions:

- A normal human being cannot identify that a sensitive data is embedded in the image independ the alphabet
- In embedding in the same image of equal length messages in Cyrillic and Latin, the stego-images obtained have approximately 0,03% difference in the values of the parameters examined and his regularity is sometimes in favor of images containing text in Bulgarian, as the percentage is almost the same.
- Difference in the histograms is hardly observed.
- When entropy values are compared in most cases no differences are observed

This result can be attributed to the fact that the embedded message is not particularly large.

ACKNOWLEDGMENTS

The research was supported by the NSP DS program, which has received funding from the Ministry of Education and Science of the Republic of Bulgaria under the grant agreement no. Д01-74/19.05.2022.

REFERENCES

[1] D. Artz, "Digital steganography: hiding data within data," in *IEEE Internet Computing*, vol. 5, no. 3, pp. 75-80, May-June 2001, doi: 10.1109/4236.935180

[2] National Statistical Institute, NSI, [Online]. Available: https://www.nsi.bg/bg/content/2823/116-лица-използващи-интернет-по-цели-на-използване. [Accessed: Febr. 7, 2024].

[3] Statista, [Online].Available: https://www.statista.com/statistics/1238307/eu-european-union-internet-users-use-accessed-internet-daily [Accessed: Febr. 7, 2024].

[4] B. Dimitrov, Book Exhibition Dedicated to the Day of the Cyrillic Alphabet, May 19th, 2023, [Online]. Available: https://blogs.eui.eu/library/cyrillic-alphabet/ [Accessed: Febr. 1, 2024].

[5] A. Westfeld and A. Pfitzmann, Attacks on Steganographic Systems. In Proceedings of the 6th European Conference on Computer Vision, ECCV 2000, Dublin, Ireland, pp. 61–76.

[6] T.S. Reinel, R.P. Raul and I. Gustavo, "Deep Learning Applied to Steganalysis of Digital Images": A Systematic Review. IEEE Access 2019, 7, 68970–68990.

[7] Image Analysis - MATLAB & Simulink, [Online]. Available:http://www.mathworks.com [Accessed: Febr. 1, 2024].

[8] Kr.Slavyanov, Fuzzy Logic Procedure for Drawing up a Psychological Profile of Learners for Better Perception in Courses. In: the 12th International Scientific and Practical Conference, 2019, Vol.II, p. 140.

[9] K. R. Rao and P. C. Yip, The Transform and Data Compression, 1st ed.: CRC Press, 2001

[10] V. Stoyanova, Steganography System Using LSB Methods, ENTRENOVA. ENTerprise REsearch InNOVAtion Conference, September 6–8, 2018, Split, Croatia.

[11] T. Ahvanooey, Q. Li, J. Hou, R. Rajput and C. Yini, Modern Text Hiding, Text Steganalysis, and Applications: A Comparative Analysis. Entropy, 2019, 355.

[12] M. Aman, A. Khan, B. Ahmad and S. Kouser, "A Hybrid Text Steganography Approach Utilizing Unicode Space Characters and Zero-Width Character". Int. J. Inf. Technol. Secur, Jan 2017, 9, 85–100.

[13] H. Wang and S. Wang, “Cyber warfare: Steganography vs. Steganalysis”, Communications of the ACM, vol. 47, no. 10, 2004.

[14] A. M. Al-Shatnawi, "A New Method in Image Steganography with Improved Image Quality", Applied Mathematical Sciences, vol. 6, no. 79, pp. 3907 – 3915, 2012.

[15] C. K. Chan and L. M. Cheng Hiding data in images by simple LSB substitution, Pattern Recognition", vol. 37, pp. 469-474, 2004.

[16] C.C. Chang, J.Y. Hsiao and C.S. Cha "Finding optimal least-significant-bit substitution in image hiding by dynamic programming strategy", Pattern Recognition, vol. 36, pp. 1583-1595, 2003.

[17] C. H. Yang and S. J. Wang, "Transforming LSB Substitution for Image-based Steganography in Matching Algorithms", Journal of Information Science and Engineering, vol. 26, pp. 1199-1212, 2010.

[18] V. Stoyanova and Zh. Tasheva. "Research of the characteristics of a steganography algorithm based on LSB method of embedding information in images" Machines. Technologies. Materials., vol.9.7, pp. 65-68, 2015.

[19] E. Cole "Hiding in Plain Sight: Steganography and the Art of Covert Communication“, Wiley Publishing, Inc., Indianapolis, Indiana, 2003