UDC 621.327:621.391

IMPLEMENTATION OF THE METHOD OF ENCODING SERIES LENGTHS TO PROVIDE PROCEDURES STEGANOGRAPHIC IMAGE INSERTION

Honcharov Mykyta, Melkozerova Olha

V. N. Karazin Kharkiv National University, St. Svobody Square, 4, Kharkiv, 61022, Ukraine
m.honcharov@student.karazin.ua, olha.melkozerova@karazin.ua

Received: on October 2023. Accepted: on November 2023.

Abstract: *The purpose of this article is to introduce the principle of encoding series lengths to provide inter-block data multiplexing of a hybrid steganoalgorithm. As part of the modeling, it is made the assumption that the attacker has successfully determined one of the two parameters of content processing: the size of the basic blocks (BB) and the implemented principle of series scanning. The modeling was performed on the example of one test halftone image of the «city» type. Samples of the attacked test image obtained for a short sample stack (length in 4 series) are presented. An analysis of the results of attack content when using different ways of smoothing the source images is carried out. The implementation of different smoothing ways allows to improve in the combinatorics of multiplexing series and the number of formed BB. It is emphasized that with an increased dimensionality of the BB, the combinatorics of the multiplex of series parameters is limited. The dimension of the BB and the way of organizing the series scanning are elements of the composite key of the data extractor. The use of the principle of encoding series lengths significantly reduces the computational complexity of the algorithm and creates conditions for the implementation of inter-block multiplexing procedures. It is concluded that the use of the parameter of series lengths destroys the correlation relations of the source data more than in the case of using only BB. The main elements at the stage of inter-block content processing are BB and their parameters of series lengths. The modeling results confirm the key role of the «series length» parameter in the procedure of legitimizing content extraction. The variability of the sampling order of the used parameters of the BB series significantly enhances the resistance of the content to unauthorized extraction attempts. Attention is drawn to the fact that the principle of encoding series lengths is limited in the background areas of images, which makes it possible to preserve highly informative image areas and determine the structure of visual artifacts. According to the modeling results, it is concluded that the use of the method of encoding series lengths in inter-block multiplexing procedures additionally strengthens the protection and complicates the work of the steganoanalyst and determines the further course of the attack.*

Keywords: *Encoding Series lengths, Steganography, Content, Hacking, Stack.*

1. Introduction

The specialists in the field of image processing are well aware of the method of series length encoding, which is differed by simplicity of realization and has a low computational complexity. Its use in video data compression systems and formats of graphic information representation allows to obtain good results when processing images with limited color or brightness palette (*grayscale images*), and/or containing extended background areas with more or less homogeneous fill [1-4].

Taking into consideration the specificity of the human visual system and the features of the method of the series lengths encoding when processing images with different statistical characteristics [1, 3-5], an assumption is made about the possibility of its use to provide procedures inter-block data multiplexing, within the framework of the chosen concept of realization of a low-resource hybrid steganoalgorithm [6]. This possibility is due to the presence of 3 important circumstances characteristic of this method: 1 - undemanding to hardware resources (*orientation on mobile gadgets*); 2 - high speed of processing (*support of real-time mode*); 3 - creation of conditions for realization of procedures of inter-block porting of parameters of series lengths, as a tool of counteraction to attempts at unauthorized extraction of content.

The procedure of inter-block multiplex of series parameters, precedes the stage coding with transform [1,4] for all base blocks (*BB*) of the image-content [6], which reduces the overall operating time of the algorithm. The inter-block data multiplex is provided by decomposing the original array of *BB* and corresponding series length values, through mutual permutations of these elements

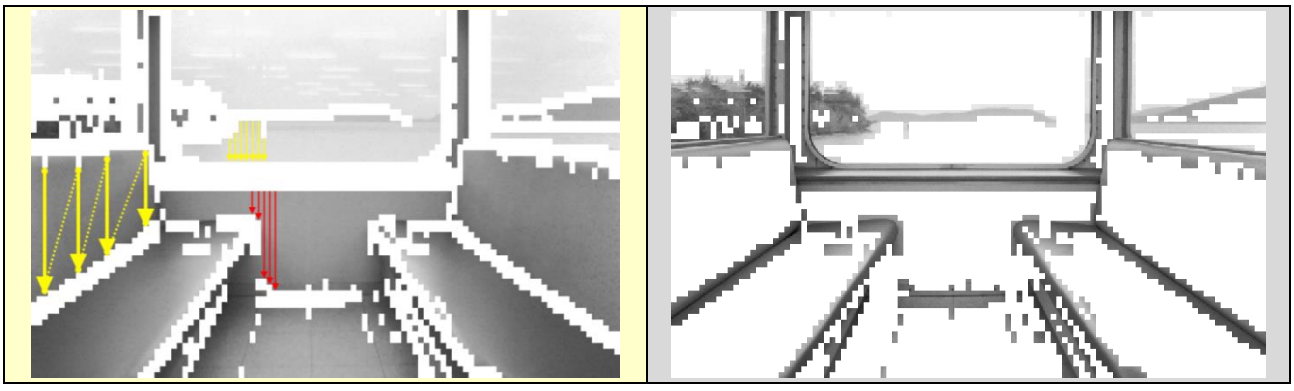
within the current combinatorics of the mixing mask. The number of formed BB depends on: - the realized way of smoothing original images and the threshold value of the difference «P_Z» between the elements of image blocks [6]; - the set dimensionality of blocks at the stage of formation of the series array [7]; - statistical characteristics and type of image-content [3-5, 8].

It is important to emphasize that in addition to the block dimensionality, the most important parameter providing legitimization of the procedure of content extraction is the way used to organize the scanning of the series. These two parameters are elements of the composite key of the data extractor and determine the order of realization of content encoding-decoding procedures at the level of inter-block processing of video data-content [6].

2. Main part

This work aims to demonstrate the possibility of implementing the series length encoding method [1-2] to provide procedures for two-level multiplexing of original data [6-7]. Multiplexing of the active parameters of the series length array provides the solution of two problems: 1 - reduces the computational complexity of the whole algorithm; 2 - counteracts attempts of illegitimate extraction of data (image-content) from the container. The way of forming series from blocks of the original image used in the modeling, is schematically represented in Fig. 1.

a) Original test image;



*b) Marking of **BB**;* *c) Marking of **All Series**;*

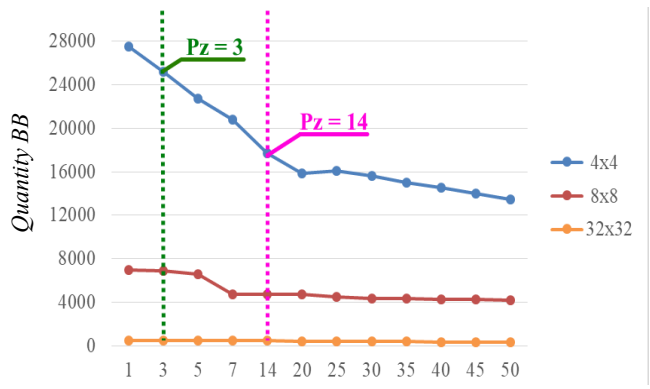
Fig. 1 - Scanning of blocks by columns and marking of the main elements of the series array (***BB*** *- white blocks on the sample (b);* ***All Series*** *- white areas on the sample (c);*
The working size of blocks - 8×8 el.)

It is schematically presented the realized scheme of blocks scanning (*by columns, from top to bottom and from left to right*) and characteristic arrangement of BB in each new series (*marked in white color, Fig. 1(b)*). In Fig. 1(c), all formed series of BB, regardless of their content, are highlighted in white color. In general, Fig. 1: - visualizes the total number of blocks to be encoded with

transformation at the next stage of the algorithm (*i.e., everything that is NOT white in Fig. 1(b)*); – allows to judge the potential combinatorics of mutual permutations of the parameter of series lengths for *BB* of size 8×8 el *(Fig. 1(c)).*

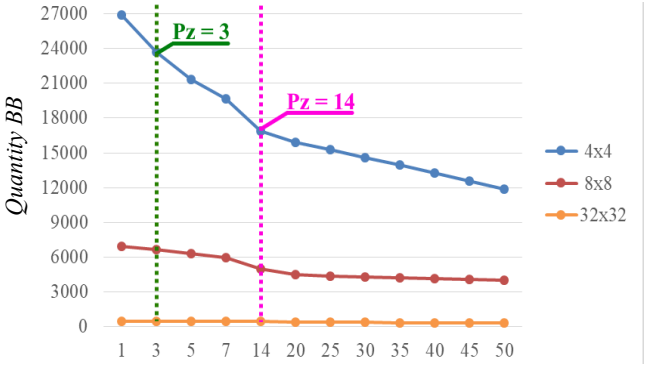
For presentation purposes of the obtained effects, the modeling results are presented on the example of one test halftone image of the type «landscape - city» (*see Fig. 2*). During the modeling process, the implementation of different variants of smoothing [6] of the original data was performed (*see Fig.5 (a-d)*). The graphs in Fig. 2 *(a, c, d)* reflect the nature of the dependence of the number of formed blocks (*respectively, and the combinatorics of permutations*) on the set values of the threshold of coarsening of the brightness of neighboring image elements *(P_Z)* and the dimensionality of the blocks themselves for different ways of smoothing the original images [5,6].

As can be seen from the presented dependences, the difference between different ways of data pre-processing practically disappears when choosing the P_Z value more than 14 brightness gradations (*dashed-cutoff P_Z = 14 in Fig. 2*).

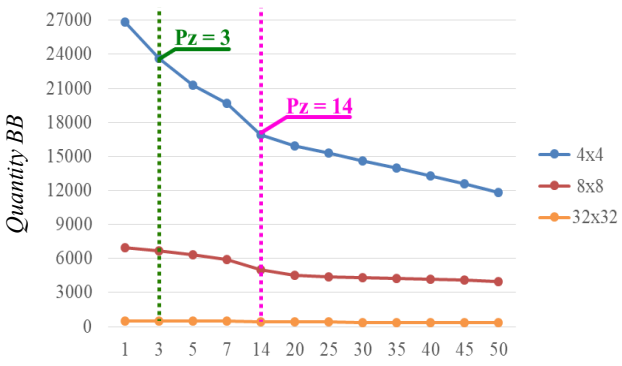


a) Number of BB without pre-processing;

b) Original test image;



c) Number of blocks for var.2 (P_Z = 3);



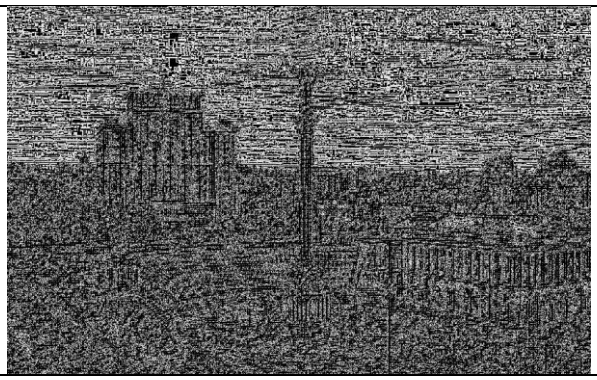
d) Number of blocks for var.3 (P_Z = 3);

Fig. 2 - Sample test image *(b)* and the number of formed blocks *(a, c, d)* for different values of P_Z, and ways of smoothing the original data.

Figs. 3-5 are presented images demonstrating the modeling results of attempts of illegitimate extraction/cracking of content on the assumption that the attacker knows only two processing parameters: - the size of the *BB* and the realized principle of scanning series [9], but he fails to determine the active parameters of the inter-block [10] data multiplex. This development of the situation predetermines three outcomes of events: 1 - incorrect determination of the valid *BB* arrangement parameter *(Fig. 3(a)); 2 - incorrect determination of the multiplex of the series length parameter (Fig. 3.c)*; 3 - mistake in determining two processing parameters at once, i.e., the *BB* and the series length parameter (*see Fig. 3(d)*). To demonstrate the characteristic structure of test image artifacts *(Fig. 3 (a, c, e))*, which are a consequence of unsuccessful content hacking, a short sampling stack *(length in 4 series)* was used. Naturally, this significantly limits the general combinatorics of per-

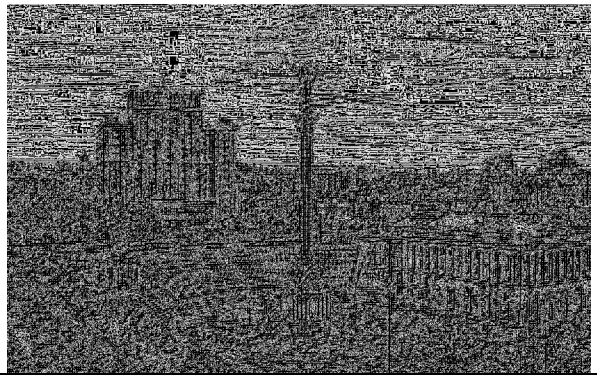
mutations of the investigated parameters, but the destruction of the original image, even on such a limited base, is more than indicative.

a) mix BB;



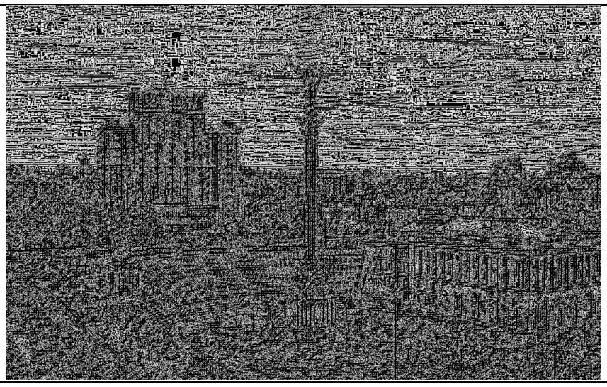
b) difference for mix BB;

c) mix Series;



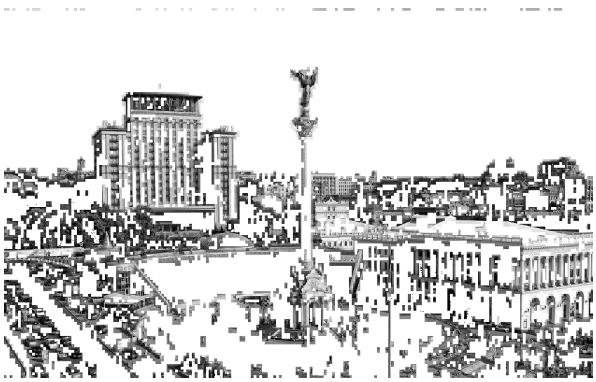
d) difference for mix Series;

e) mix BB + Series;



f) difference mix BB + Series;

g) marking BB in the original image (Fig.2(b));



h) marking Series in the original image (Fig.2(b));

Fig. 3 - Results of unsuccessful hacking of an image of the type city and visualization of artifacts for *BB* size 4×4 el. (*2nd var. processing;* $P_Z = 3$)

Naturally, the extension base of sampling will lead to a significant disruption of correlation relations between the active series *BB*. In turn, this will lead to intensification of the process of content defragmentation as a result of attempts of its unauthorized extraction [9,10].

Fig. 3*(b, d, e)* presents visualization of the difference between the original and «attacked» image-content, typical for the above three event outcomes. In this case, the darker the image fragments, the smaller the difference between the original and the obtained brightness values of their constituent elements and, accordingly, the brighter the element, the greater this difference. Samples *(g,h)* in Fig. 3 demonstrate the factual location of *BB* and their series in the test image when mistakenly selecting two parameters at once *(Fig. 3(e)). Conceptually, Fig. 3(g,h)* is similar to Fig. 1, but characterizes the situation with respect to the selected test image.

In Fig. 4 presents an enlarged fragment of the upper corner (proportional to 1/6) of the test image in Fig. 2*(b)*. The presented labeled fragments give a more detailed view of the actual location of the *BB* and their series directly in the test image (highlighted in blue color) by analogy with Fig. 1. This example vividly demonstrates the difference in the impact of the existing parameters of the *BB* series (*series length and BB content*), emphasizing the importance of the accuracy of their recovery in the content extraction process.

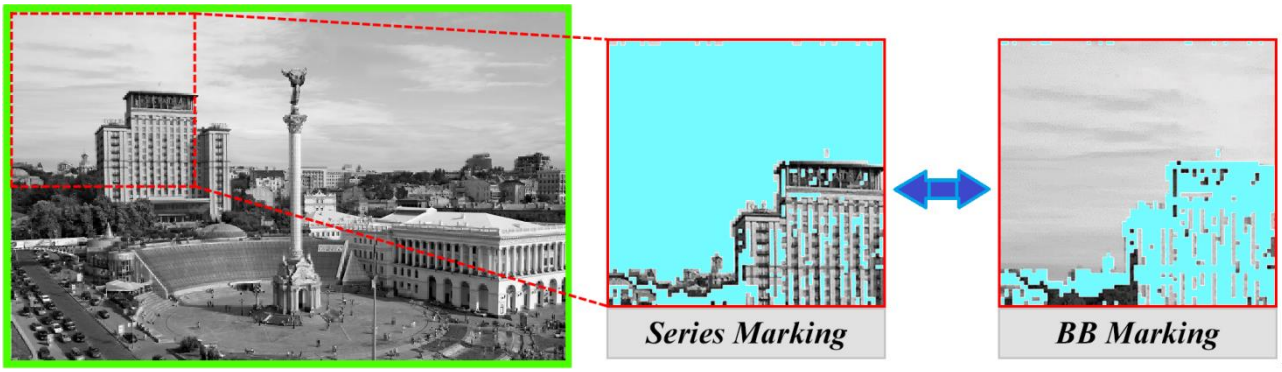


Fig. 4 - Enlarged fragment of a test image of the type city for *BB* size 8×8 el.
(*2nd var. processing;* $P_Z = 3$)

Results similar for the case of Fig. 3, but for different variants of content pre-processing *(i.e., smoothing)*, are presented in *Fig. 5(a-d)*. In this case, the modeling results are backed up by the corresponding values of *PSNR* (peak signal-to-noise ratio) and MSE *(mean square error)* that correspond to different options for their processing [6].

It should be noted that from the point of view of visual visibility of appearing artifacts, the value $P_Z = 14$ is a limiting value for almost all types of images. Therefore, in the interest of providing videodata steganographic insertion procedures, the most «interesting» results are obtained when using smoothing matrices of small dimensionality *(3×3 el.)*, *«*P_Z*»* values from 3 to 7 gradations and the dimensionality of the *BB* series in the range of *4÷8 el*. It is for such, processing parameters in *Figs. 2,3,5* presents the results of the attack of test content samples. In general, *Figs. 2,3,5* clearly show the possible consequences of selection the active parameters of the inter-block multiplex of BB series implemented on a small stack length of series sampling.

3. Conclusions

1. The use of the principle of series length encoding in the implementation of procedures for inter-block data multiplexing creates a good basis for counteracting attempts at unauthorized extraction of content.

2. Applying the principle of encoding series length decreases the computational complexity of the whole algorithm *(by reducing the total number of BB)* and creates the necessary original conditions for the implementation of inter-block data multiplexing procedures.

a) incorrect mix BB for Var.2;
(MSE = 52.08; PSNR = 30.964)

b) incorrect mix BB for Var.3;
(MSE = 52.952; PSNR = 30.892)

c) incorrect mix of Series parameter (Var.2);
(MSE = 110.176; PSNR = 27.71)

d) incorrect mix of Series parameter (Var.3);
(MSE = 106.367; PSNR = 27.863)

Fig. 5. Results of unsuccessful extraction *(i.e. attack)* of a test image of type city *(a-d)* for different smoothing variants of the original data *(P_Z = 3; BB 8×8 el.)*

3. Basic blocks and their parameters of series lengths are the main procedural elements in the inter-block content processing stage.

4. The modeling results confirmed the leading role of the series length parameter *(see Fig.3 (a-e))* as a key element of the content extraction legitimization procedure. This thesis is confirmed already at a small stack length of the *BB* series sampling.

5. The inter-block multiplex of the series length parameter destroys the correlation relations of the original data to a much greater extent than when using only *BB*. The combined use of both parameters (*BB and Series lengths BB*) further enhances the required effect *(see Fig.3 (e))*.

6. An increase in the dimensionality of the *BB* significantly limits the combinatorics of the multiplex of series parameters, due to the reduction of the base stack sampling *(Fig. 2(a,c,d))*. The optimal dimensionality of the *BB* should be considered blocks in the range from *4×4* to *12×12 el.* with the value of $P_Z \leq 7$.

7. The use of various ways of content pre-processing allows to improve the combinatorics of the series multiplex, which is well confirmed by comparing the hacking results of the test image in *Fig.5(c, d)*.

8. The use of various ways of smoothing the input data, even with the same P_Z values, provides a significant difference in the number of formed *BB*.

9. The dimensionality of the *BB* and the way of organizing the scanning of series are the elements of the composite key of the data extractor [6], which determine the order of implementation of inter-block data processing procedures as a tool for legitimizing access to video content data.

10. The variability of the sampling order of the used parameters of the *BB* series when implementing any way of their scanning [9], strengthens the resistance of the content to attempts of its unauthorized extraction.

11. The use of the principle of series length encoding is largely limited in the background areas of images, which is clearly visible in the example images in *Fig.4*. Such data processing practically does not affect highly informative image areas, for instance, contour boundaries *(building details in Fig. 4)*. Therefore, for blocks forming similar image areas, it is necessary to use the intra-block data multiplexing mechanism *(for example, in parts of the transformation coefficients characterizing the average brightness of the BB)* [6,8].

References

[1] Прэтт У. (1985). Цифровая обработка изображений (Д. С. Лебедева, пер. с англ.). т. 1,2. Москва: Мир.

[2] Бутаков Е. А., Островский В. И., & Фадеев И. Л. (1987). Обработка изображений на ЭВМ. Москва: Радио и связь.

[3] Ярославский Л. П. (1979). Введение в цифровую обработку изображений. Москва: Сов. Радио.

[4] Зубарев Ю. Б., Дворкович В. П. Цифровая обработка телевизионных и компьютерных изображений. Москва: МЦНТИ, 1997. – 212 с.

[5] Красильников Н. Н. (1976). Статистическая теория передачи изображений. Москва: Связь.

[6] Лесная, Ю., Гончаров, Н., & Малахов, С. (2021). ОТРАБОТКА КОНЦЕПТА МНОГОУРОВНЕВОГО МУЛЬТИПЛЕКСА ДАННЫХ ГИБРИДНОГО СТЕГАНОАЛГОРИТМА. Збірник наукових праць SCIENTIA. Вилучено із: https://ojs.ukrlogos.in.ua/index.php/scientia/article/view/17666.

[7] Гончаров О., Лєсная Ю., Погоріла К., Богданова Є., Малахов С. Дослідження параметру «серій опорних блоків», як елементу композитного ключа екстрактора даних стеганоалгоритму // Problems of science and practice, tasks and ways to solve them. Proceedings of the XX International Scientific and Practical Conference. Warsaw, Poland. 2022. Pp. 779-785. URL: http://www.isg-konf.com/wp-content/uploads/2022/05/Problems-of-science-and-practice-tasks-and-ways-to-solve-them.pdf.

[8] Гончаров, Н., & Малахов, С. (2022). Использование параметра длин серий, как элемента межблочного мультиплекса данных стеганоалгоритма. Збірник наукових праць ΛΌГОΣ, 180-187. https://doi.org/10.36074/logos-08.07.2022.050.

[9] Лєсная, Ю., Гончаров, М., Семенов, А. & Малахов, С. (2023) Моделювання розгортки серій опорних блоків зображення, як інструменту з протидії спробам несанкціонованої екстракції стеганоконтенту. Grundlagen der modernen wissenschaftlichen Forschung: coll. of scientific papers «ΛΌГОΣ» with Proceeding of the International Scientific and Practical Conf. (Pp.109-115). March 31, 2023, Zurich, Switzerland: BOLESWA Publishers & Europäische Wissenschafts plattform. Вилучено із: https://archive.logos-science.com/index.php/conference-proceedings/article/view/631

[10] Honcharov, M., Pavlova, L., & Lesnaya, Y. (2022). Modeling steganocontent extraction attempts with different lengths stack sampling series of images blocks. Computer Science and Cybersecurity, (2), 22-27. Вилучено із: https://periodicals.karazin.ua/cscs/article/view/21036/19743

Надійшла: Жовтень 2023. **Прийнята**: Листопад 2023.

Автори:
Гончаров Микита, аспірант кафедри безпеки інформаційних систем та технологій Харківського національного університету імені В.Н. Каразіна, Україна.
ORCID ID: https://orcid.org/0000-0002-9790-7260
E-mail: m.honcharov@student.karazin.ua

Ольга Мелкозьорова, к.т.н, доцент кафедри безпеки інформаційних систем і технологій Харківського національного університету імені В.Н. Каразіна, Україна.
ORCID ID: https://orcid.org/0000-0002-1134-2925
E-mail: olha.melkozerova@karazin.ua

Імплементація методу кодування довжин серій для забезпечення процедур стеганографічної вставки зображень.

Анотація. Метою даного матеріалу є ознайомлення з принципом кодування довжин серій для забезпечення міжблочного мультиплексу даних гібридного стеганоалгоритму. У рамках моделювання зроблено припущення, що атакуючий вдало визначив один із двох параметрів обробки контенту: розмір опорних (базових) блоків (ОБ) і реалізований принцип розгортання серій. Моделювання проводилося на прикладі одного тестового напівтонового зображення типу «місто». Представлені зразки атакованого тестового зображення отримані для короткого стеку вибірки (довжиною в 4 серії). Виконано аналіз результатів атаки контенту при використанні різних способів згладжування вихідних зображень. Реалізація різних способів згладжування дає змогу поліпшити комбінаторику мультиплексу серій та кількість сформованих ОБ. Підкреслено, що при збільшені розмірності ОБ обмежує комбінаторику мультиплексу параметрів серій. Розмірність ОБ і спосіб організації роз-

гортки серій є елементами складеного ключа екстрактора даних. Використання принципу кодування довжин серій значно зменшує обчислювальну складність алгоритму та створює умови для реалізації процедур міжблочного мультиплексування. Зроблено висновок, що використання параметра довжин серій руйнує кореляційні зв'язки вихідних даних більше, ніж у разі використання тільки ОБ. Основними елементами на етапі міжблокової обробки контенту є ОБ та їхні параметри довжин серій. Результати моделювання, підтверджують ключову роль параметра «довжин серій» у процедурі легітимації вилучення контенту. Варіативність порядку вибірки використовуваних параметрів серій опорних блоків, значно посилює стійкість контенту до спроб його несанкціонованого вилучення. Звернуто увагу, що принцип кодування довжин серій дещо обмежений у фонових областях зображень, що дає змогу зберегти високоінформативні області зображень та визначити структуру візуальних артефактів. За результатами моделювання зроблено висновок, що використання методу кодування довжин серій у процедурах міжблочного мультиплексування додатково підсилює захист та ускладнює роботу стеганоаналітика і визначає подальший хід атаки.

Ключові слова: *кодування довжин серій, стеганографія, контент, атака, стек.*