Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2025-23.2.00187

RESEARCH ARTICLE

Software Implementation of Detection of Hidden Information in Multimedia Files

O. Safaryan[1], I. Alferova[1], N. Gapon[1], M. Zhdanova[1], A. Romanov[1], E. Semenishchev[2]

[1]Don State Technical University. (Russian Federation)
[2]Moscow State University of Technology «STANKIN» (Russian Federation)

ARTICLE INFO

Received: Jul 21, 2025

Accepted: Sep 13, 2025

Keywords

Steganography
Steganalysis
Chi-Square Method
Koch-Zhao Method
Method Of Expanding the Palette of A GIF Container
Method of Using an Additional Comment Block of A GIF Container
Multimedia Files

***Corresponding Author:**

narong@pi.ac.th

ABSTRACT

The aim of this study is software implementation of detection of hidden information in multimedia files. Research methods: method based on analysis of Chi-square statistics, Koch-Zhao method, method of expanding the palette of GIF container, method of using an additional block of GIF container comments. Result: software for the task of detecting hidden information in multimedia files has been developed. In addition, the article provides a brief overview and analysis of the state of the art of steganalysis. The proposed methods allow detecting the presence of hidden information embedded in multimedia files by various popular steganography tools with sufficiently high accuracy. The theoretical significance of this work lies in the development of a fairly promising approach to steganalysis. The practical significance lies in the developed software product, as well as in the confirmatory methods of steganalysis in relation to detecting hidden information in multimedia files.

INTRODUCTION

Currently, files and data in graphic formats occupy a significant part of network traffic and are found everywhere. They are presented not only in the form of images posted on various online resources, but also in the form of elements of graphic interfaces and design solutions, as well as in more complex data formats and in other forms.

In recent years, there has been a noticeable increase in the number of attacks on information systems using steganography methods. These methods allow the transmission of a secret message to be hidden. These attacks pose a threat because they can be carried out using ordinary data files, especially graphic formats, which are used to deliver malicious code or leak confidential information.

Steganalysis is a section of steganography that studies methods for detecting hidden information in an analyzed object, as well as methods for extracting hidden information when it is detected in the absence of the necessary input data. Hidden information is usually data encrypted using steganographic methods.

Steganalysis methods should be actively used by both security specialists during incident investigations and automated security systems with integrated data file analysis modules. An important feature of such work is the need to study various elements and forms of presentation of container files. This includes checking not only the color values of pixels in images, but also frequency characteristics. This raises questions related to the best practices for using steganalysis algorithms and the correct interpretation of their results.

Despite the variety of algorithms for hiding information in graphic files, almost all of them are based on several basic steganography methods. These include methods using the least significant bits of pixels, as well as the Koch-Zhao method, which encodes information in the frequency domain of the image.

Most other steganographic methods are modifications or variations of these two methods.

To detect information hidden using these methods, special steganalysis methods have been developed. Their software implementation allows the analysis process to be automated and carried out without human intervention.

2. The Operating Principle of Systems for Concealing and Detecting Information in Multimedia Files

With the development of technology in the late 20th century, computer steganography received a new impetus. Currently, this field of science relies on the methods and achievements of cryptography, digital signal processing, communication and information theory.

The main task of steganographic methods is not only to conceal confidential information, but also to conceal the very fact of the existence of this information during transmission, storage or processing [1,2].

Digital steganography is a method of transmitting secret data, in which information is hidden in media files. Unlike traditional steganography methods, digital steganography allows you to hide data in media files so that it is impossible to detect them without special tools.

The threat of digital steganography is that media files can be safely transmitted over open communication channels without fear of detection of hidden information. This makes digital steganography a dangerous tool for violating the fundamental principles of restricted information.

The list of steganographic methods is updated annually, more reliable and original methods of hiding information are invented, the neutralization of which will require new, effective methods of analysis [3].

In steganography, a so-called container is used to hide information. This can be any file or data stream, the structure and size of which allow the necessary data to be hidden. Most often, images, text files, audio and video files are used as containers.

Using steganography, a secret message is embedded in an inconspicuous object that is sent to the addressee or placed in a publicly accessible area. The recipient of the message, knowing the steganographic key, decrypts the message [4].

Figure 1 shows a diagram of steganographic transformations.

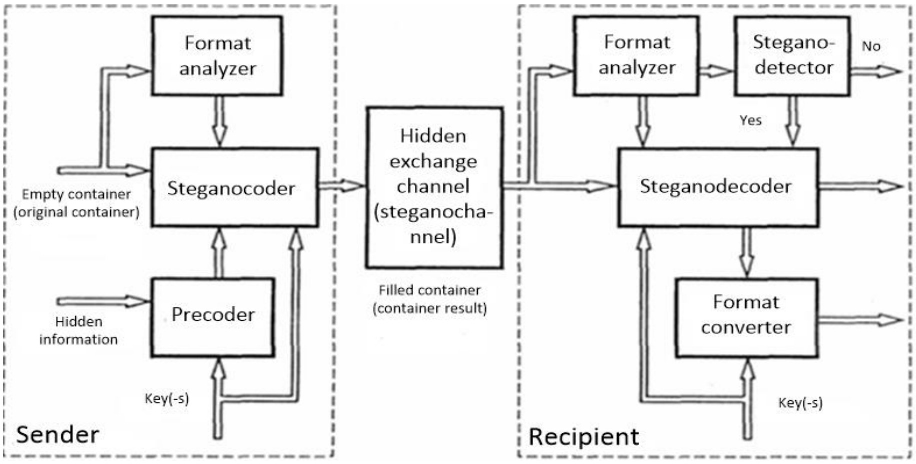


Figure 1. Steganosystem diagram

The main components of a steganosystem are:

- empty container – a container that does not contain a secret message;

- filled container (stegocontainer) – a container that contains a secret message;

- attachment – data to be hidden;

- key (stegokey) – a secret key needed to hide the stegocontainer.

There are two types of keys in stegosystems: closed (secret) and open. If a stegosystem uses a closed key, it must be created either before the start of message exchange or transmitted over a secure channel. A stegosystem that uses a public key must be designed in such a way that it is impossible to obtain a private key from it. In this case, the public key can be transmitted over an unsecured channel;

- format analyzer – a program for checking a container for the possibility of its use for steganography (format, potential size of the attachment);

- steganocorder – a software module implementing a steganographic algorithm, taking into account the characteristics of the container (converts the attachment into a stegan-embedding);

- precoder – a software module designed to convert the hidden message into a form convenient for embedding in the container;

- hidden exchange channel – a Steg container transmission channel;

- steganodecoder – a software module that restores the stegan-embedding (without decrypting and/or unzipping);

- format converter – a tool for converting data from one format to another. Usually with the preservation of the main logical and structural content of the information.

Steganalysis is the process of detecting and analyzing hidden information in digital data. It is used to detect the presence and characteristics of steganographic algorithms that hide data within other types of data, such as images, audio, or video files.

The main goal of steganographic analysis (steganalysis) is to study the resistance of a steganographic embedding scheme to various types of attacks. The main tasks of steganalysis are:

1. Detecting hidden information. Steganalysis can be used to attempt to detect the presence of hidden information in digital data. This can be useful for detecting unauthorized use or distribution of confidential data.

2. Calculating the size and location of hidden information.

3. Extracting hidden information. Steganalysis can help extract hidden information from digital data. This is useful for analyzing the content of hidden information and determining its source.

4. Destroying hidden information. In some cases, it may be necessary to destroy secret information to prevent its further use or distribution.

5. Analyzing the strength of existing steganographic algorithms. Steganalysis can be used to assess the vulnerabilities of existing steganographic algorithms and determine how susceptible they are to attack.

6. Developing new methods for detecting hidden information. Based on the results of steganalysis, new methods for detecting hidden information can be developed that will be more effective.

At the moment, the main achievements in the field of steganalysis are related to the solution of the first problem. In this case, statistical methods of studying the container with the embedded message are mainly used.

All these methods are based on the idea that the message injection leads to changes in the statistical characteristics of the container, which can be detected by analyzing various distributions.

Thus, there is a random nature of the distribution of the least significant bits of the blue component, and on its basis the chi-square criterion is used to detect steganography. This method gives good results with uniform filling of the container.

However, statistical methods are not enough to solve the problems of calculating the size and extraction, and it is necessary to use intelligent algorithms. For example, to analyze the least significant layer, the hierarchy analysis method is used, which allows you to determine the size and position of the embedded information with high accuracy.

There are several steganalysis methods, each of which has its own characteristics and approaches. Here are some of them:

1. Statistical Analysis. This method is based on the analysis of statistical properties of media files. Steganalysis compare statistical characteristics of files with and without embeddings to detect anomalies. For example, in images, the distribution of pixel values can be analyzed to detect changes caused by the embedding of hidden information. Used to detect steganographic methods that change the statistical properties of an image, audio, or video.

2. Frequency Response Analysis**.** This method involves analyzing the frequency components of media files, such as the Fourier Transform or Wavelet Transform. Changes in frequency characteristics can indicate the presence of hidden information. Often used in audio files, where the embedding of information can change the frequency spectrum.

3. Machine Learning-Based Methods. Machine learning algorithms can train models based on large data sets so that they can classify files as containing or not containing hidden information. These models can take into account many factors, including statistical and frequency characteristics. Effective in automating the steganalysis process and improving detection accuracy. 4 Visual artifact analysis. This method involves visually analyzing images for artifacts that may indicate the presence of steganographic embedding. For example, changes in texture or color may be a sign of embedding. It is used mainly for images where visual changes may be noticeable.

5. Comparison-based methods. This approach involves comparing the media file with its original version (if available) to detect changes. This can be done using various algorithms, such as differential analysis. It is effective when the original file is available, allowing for accurate detection of changes.

6. Metadata analysis. Some steganographic methods may leave traces in the metadata of files. Metadata analysis can help identify the presence of hidden information. It is used for audio and video files, where the metadata may contain information about encoding and modifications.

When creating stego attack algorithms, it is essential to understand what embedding methods are currently in use. Currently, most stego attacks are targeted attacks. This means that each individual attack algorithm is designed to detect an embedding made by a specific method or group of steganography methods with similar characteristics. Thus, with knowledge of popular embedding algorithms, it is possible to develop stego attacks that will be effectively used in the modern world.

In addition, to create steganalysis algorithms, it is necessary to understand which media files most often serve as containers for transmitting hidden information. It is important to note that most stego attacks focus on identifying various statistical characteristics of a media file with an embedding and their deviations from the properties of a clean file (a file without an embedding).

Media files of different formats have their own unique statistical characteristics. Although it is currently impossible to collect accurate data on statistical distributions for all media file formats due to the complexity of computer file statistics, stegoattack algorithms are still based on certain approximated properties of media files. Understanding which file formats can serve as containers allows us to develop automated stegoattack methods for detecting media files with attachments in large amounts of information. Thus, steganalysis is an important tool for ensuring the security of digital data. It allows us to identify and destroy hidden information, as well as to develop new methods for protecting data from unauthorized use.

3. Development of System Modules

Figure 2 shows the general scheme of the proposed system. Developing a steganalysis system capable of detecting hidden information in digital objects requires a modular approach that provides flexibility, scalability, and the ability to adapt to new steganographic methods.

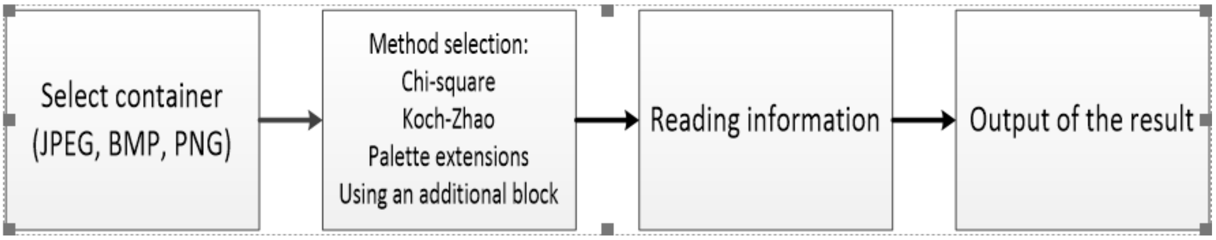


Figure 2. General diagram of the system

Each module should be focused on analyzing a specific type of data or applying a specific steganalysis algorithm, which allows dividing a complex task into smaller, more manageable ones. Such a modular design simplifies the development, testing, and integration of individual components, and facilitates adding new features and updating existing ones. This section develops steganalysis systems, including image analysis modules.

3.1. Stegoanalysis using the Chi-square statistical analysis method

The method of attacking a Steg system using the Chi-square analysis criterion was proposed and described by Andreas Westfeld and Andreas Pfitzmann in 1999. This method is aimed at detecting information hidden by the least significant bit (LSB) method [5, 6].

During the operation, the following steps are performed:

1. Splitting the image into pixel rows.
2. Splitting each pixel row into color components and extracting the least significant bits (RGB) from them.
3. The empirical distribution value is found for each row.
4. The expected distribution value is specified for each row.
5. The Chi-square criterion and significance level are calculated for each row.
6. A histogram is constructed. The abscissa axis is the row number of the analyzed container image; the ordinate axis is the significance level value for a specific row.

The significance level value gives us the probability of obtaining a Chi-squared value at least as extreme as the observed value if the null hypothesis is true. In this context, the null hypothesis is the assumption that the bits are randomly distributed (i.e. the probability of '0' and '1' is 0.5) or that there is no information embedded in the analyzed container image. The lower the significance level value, the higher the probability that the image contains a hidden message.

The significance level value is determined based on the Chi-squared distribution with the corresponding number of degrees of freedom (df). In this context, $df = 1$ (number of categories minus 1). The significance level value is calculated using the Chi-squared cumulative distribution function (CDF).

The formulas for calculating Chi-square and the significance level are presented in expression (1) and (2), respectively:

$$\mathrm{x}^2 = [(\mathrm{O}_0 - \mathrm{E}_0)^2/\mathrm{E}_0] + [(\mathrm{O}_1 - \mathrm{E}_1)^2/\mathrm{E}_1] \quad (1)$$

$$P(x^2 \geq x_{OBS}^2 | df = 1) \quad (2)$$

The following steps are performed during the execution of the function shown in Figure 15:

1. Splitting the image into pixel rows.
2. Splitting each pixel row into color components and extracting the least significant bits (RGB) from them.
3. Finding the empirical distribution value for the entire container image.
4. Setting the expected distribution value for the entire container image.

Calculating the Chi-square criterion and the significance level for the entire container image/

Two functions were implemented based on the above steps. In both cases, the container image is input. The output for the first function is a histogram showing the significance level value for each image row; for the second function, the Chi-square value and the significance level value for comparison between the original image and the assumed filled container.

3.2. Koch-Zhao steganalysis

This type of analysis is designed to detect the embedding of a message in an image container using the Koch-Zhao method. This method searches for information encoded in the frequency representation of an image [4].

The image is represented in the frequency domain by calculating the coefficients of the discrete cosine transform (DCT). To do this, the image is divided into blocks of 8×8 pixels. Then, a two-dimensional DCT is performed for each block, resulting in a matrix consisting of 64 coefficients.

In the resulting matrix, the coefficient located in the upper left corner, corresponding to the zero frequency (the matrix element with indices (0; 0)), is called the DC coefficient. It determines the main color shade (average color intensity) of the entire block. The remaining coefficients are called AC coefficients and reflect the change in color intensity in different directions of the selected block (horizontally and vertically).

Thus, each matrix of DCT coefficients is divided into three groups: low-frequency, mid-frequency, and high-frequency (from the upper left corner to the lower right corner of the matrix).

Low-frequency coefficients have the greatest impact on the color intensity of pixels. Therefore, any changes and transformations of DCT coefficients occur in the mid- or high-frequency regions.

When trying to detect an embedding using the Koch-Zhao method, it is necessary to determine which DCT coefficients were used for the embedding. This is one of the key tasks. Since the application of the Koch-Zhao method involves hiding information in one set of mid-frequency components, the main analysis operations are performed for each of these sets separately [5].

Thus, steganalysis of the Koch–Zhao algorithm is reduced to the analysis of the dependence of sequences $C_i(j)$ $(j = 1, 2, 3;\ i = 1, \ldots, N)$ and the identification of the area of step changes.

During operation, the following steps are performed:

The container image is divided into blocks B_i $(i = 1, \ldots, N)$ of 8×8 pixels.

A discrete cosine transform is performed for each block B_i $(i = 1, \ldots, N)$ and the coefficient matrices D_i $(i = 1, \ldots, N)$, which also have a size of 8×8, are found.

The elements of the matrices D_i $(i = 1, \ldots, N)$ are analyzed. To do this, three sequences must be constructed:

$$C_i(1) = ||Di[3,4]| - |D_i[4,3]||, \quad i = 1, \ldots, N, \qquad (3)$$

$$C_i(2) = ||Di[3,5]| - |D_i[5,3]||, \quad i = 1, \ldots, N, \qquad (4)$$

$$C_i(3) = ||Di[4,5]| - |D_i[5,4]||, \quad i = 1, \ldots, N. \qquad (5)$$

At this stage, not just the difference in the absolute values of the coefficients is calculated, but their absolute values - this is due to the fact that the coding of bits is determined by overcoming the threshold value P for zero and $-P$ - for one [5].

Despite the possibility of fluctuations in the form of different peak values of C_i in blocks that were not used to code the bits of the hidden message, the blocks actually used for embedding are distinguished by a relatively long continuous sequential section of peak values. The input data are container images. The output data are three histograms.

3.3. Steganalysis of the Gif Container Palette Expansion Method

IF (Graphics Interchange Format) is a format for storing graphic images, capable of storing compressed data without loss of quality in a format of up to 256 colors. This format was developed in 1987 (GIF87a) by CompuServe for transmitting raster images over networks. In 1989, the format was modified (GIF89a), support for transparency and animation was added [6].

In GIF files, information is organized in blocks. These blocks always have the same length (or their length depends on certain parameters), so it is almost impossible to determine where each block begins and ends. The structure of the simplest non-animated GIF image of the GIF89a format is shown in Figure 3.

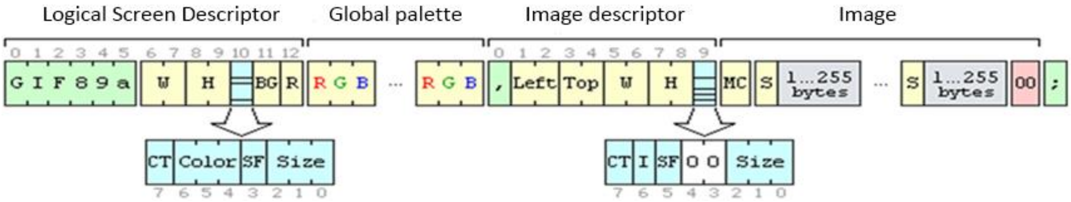


Figure 3. Structure of a GIF image in GIF89a format

Of all the blocks of the structure, in this case we need the global palette block and the parameters responsible for the palette:

- CT – presence of a global palette. If this flag is set, then the global palette should begin immediately after the logical screen descriptor [6].

- Size – the size of the palette and the number of colors in the picture [6].

- The values of this parameter are presented in Table 1.

Table 1. Size, number of colors, and occupied size in bytes

Size	Number of colors	Palette size, bytes
7	256	768
6	128	384
5	64	192
4	32	96
3	16	48
2	8	24
1	4	12
0	2	6

A palette expansion method that works only for the GIF structure. It will be most successful in container images with a small palette size. Its essence is that it increases the size of the palette, thereby providing additional space for writing the necessary bytes in place of the color bytes. Considering that the minimum palette size is 2 colors (6 bytes), the maximum size of the embedded message can be $256 \times 3 - 6 = 762$ bytes.

The disadvantage is low cryptographic security; the embedded message can be read using any text editor if the message has not been additionally encrypted.

During operation, the following steps are performed:

1. The number of unique colors in the image is read.

2. The size of the palette is determined.

3. The difference between the number of unique colors and the colors in the palette is analyzed.

The input data are GIF containers. The output data are whether concealment was detected/not detected, the number of matching colors in the palette and frame, the number of unique colors in the frame, and the number of unique colors in the palette.

3.4. Steganalysis of the method of using an additional comment block of a Gif container

The optional comment block contains text information that is not actually part of any GIF graphic image placed in the current data stream. The purpose of this block is to record comments on the graphic images, official information, descriptions, or any other type of information not related to the control or transmission of graphic data. The decoder may ignore the optional comment block or buffer it for further processing. However, the comment block must never interrupt the processing of the data stream or affect it in any way [7-10].

The original plan was to store information about the authors of GIF images in this block. However, this block is now usually skipped when viewing images.

According to the GIF specification, each such block should not exceed 255 bytes, but the number of blocks following each other is not limited.

This block is optional. There can be any number of comment blocks in the data stream.

During operation, the following steps are performed:

1. The GIF container is opened for reading.
2. Comment blocks are extracted from the GIF file. Comment blocks are stored in the info dictionary of the Image object under the "comment" key.
3. The sizes of the space occupied by the first comment block and the concatenation of the remaining comment blocks are calculated.

The input data are GIF containers. The output data is whether hiding was detected/not detected, the sizes of the first comment block, and the concatenations of the remaining comment blocks.

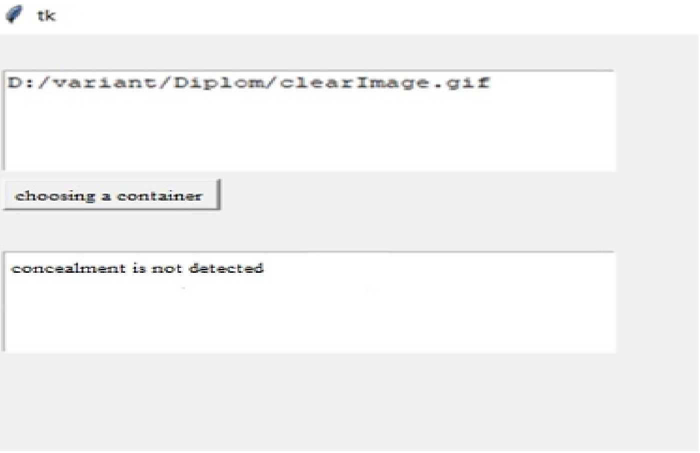
4. RESULTS

The areas of application of this software are information systems in which it is necessary to ensure security by checking multimedia objects for the presence of information hidden in them. The final software is intended for use in private companies, organizations and personally to protect confidential information. The functional purpose of the software is to detect the presence of hidden information using steganography methods in multimedia objects.

Requirements for the functional characteristics of the software:

- providing the ability to conveniently select a container image via a dialog box;
- working with JPEG, BMP, PNG, GIF file formats;
- steganalysis using a method based on the analysis of the Chi-square statistics;
- steganalysis using the Koch-Zhao method;
- steganalysis using the palette expansion method;
- steganalysis using the method of using an additional comment block.

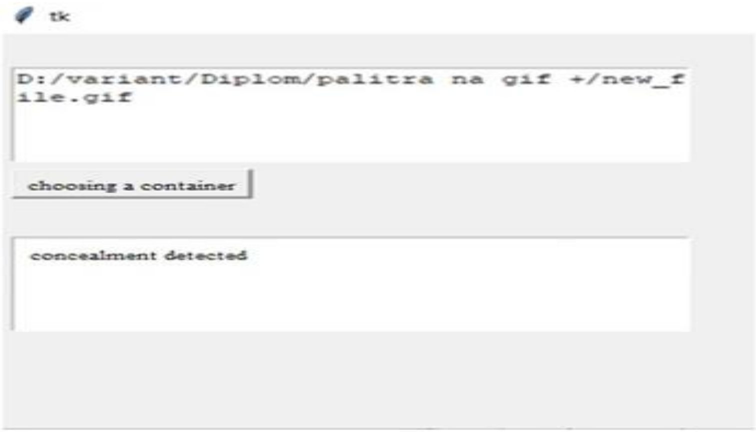
Reliability of operation is ensured by the correct functioning of the hardware, the availability of sufficient RAM to speed up work on large-format objects and free space in the storage of the technical device. The software is accessed by running the program file; as input data, the program receives a multimedia object in the JPEG, BMP, PNG, GIF format by selecting the required file via the graphical interface. Examples of outputs of the steganalysis function program for the method of using an additional comment block, empty and filled GIF containers are shown in Figures 4 and 5.



D:/variant/Diplom/clearImage.gif

concealment is not detected

Figure 4. Example of steganalysis of the method of using an additional comment block on an empty container



D:/variant/Diplom/palitra na gif +/new_file.gif

concealment detected

Figure 5. Example of steganalysis of the method of using an additional comment block on a filled container

Thus, a system for detecting hidden information in multimedia objects with all its constituent functions was implemented, a functional block diagram of the system was presented, and the operation of the software was demonstrated.

CONCLUSION

As a result, the main features and purpose of systems for detecting intentionally hidden information in multimedia files using steganography methods were considered. Information can be hidden in any data format. Thus, the use of special software is necessary to detect hidden information in multimedia objects.

The use of steganographic tools to hide information in files that are then transferred to multimedia content sites and file-sharing servers allows for the covert organization and coordination of various types of illegal actions.

In these conditions, the task of detecting hidden information, extracting and destroying it, as well as analyzing the resistance of existing steganographic algorithms, and developing new methods for detecting hidden information becomes especially relevant.

ACKNOWLEDGMENT

The Scientific Research was funded by the Ministry of Science and Higher Education of the Russian Federation under the Grant «Development of intelligent control methods for technological equipment using the example of semiconductor crystal bonding, and its automation through a machine vision system» (FSFS-2025-0009).

REFERENCES

Steganalysis in computer-technical expertise. URL : https://habr.com/ru/articles/791284/

Zakharova, O. A., Selikhina, A. V., Vezirov, T. G. (2020). Modeling an analytics system for industrial safety monitoring based on expert assessments. Advanced Engineering Research (Rostov-on-Don), 20(1), 100-105.

Romanov, A., Safaryan, O., Cherckesova, L., Revyakina, E. (2023). Implementation of steganographic software for hiding information in mpeg video files. In E3S Web of Conferences (Vol. 389, p. 07020). EDP Sciences.

Ponomarev I. V., Strokin D. I. Steganographic methods of embedding and detecting hidden messages using GIF images as container files // Bulletin of Altai State University. Mathematics and Mechanics. 2022. No. 1 (123) pp. 112-115.

Dryuchenko, M. A. (2007). Algorithms for detecting steganographic concealment of information in jpeg files. Bulletin of the Voronezh State University. Series: System Analysis and Information Technologies, (1), 21-30.

Chastikova, V. A., Abbasov, T. O., Abbasova, S. S. (2020). Methodology for Recognizing Hidden Information in Images Based on Steganography Algorithms. Bulletin of Adyghe State University. Series 4: Natural, Mathematical and Technical Sciences, (3 (266)), 40-45.

Grachev, Ya. L., Sidorenko, V. G. (2021). Steganalysis of information hiding methods in graphic containers. Reliability, 21(3), 39-46.

Belim, S., Vilkhovsky, D. Mathematical Structures and modeling. mathematical structures and modeling Founders: Omsk State University named after FM Dostoevsky, (2), 79-85.

Steganography in GIF . URL : https://habr.com/ru/articles/128327/

Graphic file GIF . URL : https : // www . clarionlife . net / graficheskiy - fayl - gif / # sect 12