ON THE MATHEMATICAL MODEL OF STEGANOGRAPHY
AND SOME OF ITS APPLICATIONS

Abstract. One of the most important areas of information technology in modern times is information security. Steganography is one of the most important tools in protecting the security of electronic information. In this regard, research in the field of steganography is one of the most pressing issues of our time. Therefore, the publication of this article is relevant.

Keywords: Steganography, steganogram, container, JPEG format, discrete cosine transformation

Steganography is currently one of the fastest and most widely developing areas of information technology. From this point of view, experts predict a great future for steganography. Although there has been considerable research in the field of steganography in modern times, there is still no single theory of steganography. This, in a sense, hinders the normal and rapid development of steganography.

Currently, there are many terms and models in steganography, but most of them are not sufficiently substantiated, and some are not fully studied.

However, there are already some changes in this direction. Although steganography is not the main or only solution to ensure information security, there
are already serious attempts to use it as an auxiliary tool. In recent years, a major change in theory has taken place, mainly due to the formation of a mathematical model of steganography.

It should be noted that, on the other hand, the field of study of steganography is so wide that it has become an interdisciplinary science in this regard. If cryptography can be abstracted outside of devices and can only solve problems in the world of discrete mathematics, then a specialist in steganography must study the environment. Although, of course, there are a number of problems in building cryptosystems, such as attacking with external channels; but this is not the fault of the quality of the password. It should be noted that steganography will develop in line with the development of the study of the environment in which secret messages are transmitted. With this point of view, new directions in steganography appear in such areas as: "chemical steganography", "steganography in pictures", "steganography in the correction of errors", "food steganography" and so on.

From this point of view, not only mathematicians-cryptographers, but also linguists, philologists and chemists have recently become interested in steganography. This predicts that this area will develop on a very large scale.

Also, the rapid development of various Internet services in modern times has guaranteed the widespread development of steganography. The modern virtual world is full of texts, pictures, videos, etc., and it is constantly filling up. More than 100 hours of video are uploaded to a YouTube site every minute. During this time, only a few hours of various videos appear on YouTube. On the other hand, this fact is a great "ground" for hiding electronic information from the point of view of steganography. That is, "technically" the world has long been ready for steganography. Experts predict that in the near future, the fight against steganography and steganography will become as relevant as the problem of modern information technology BigData or the Internet of Things. From this point of view, research in the field of steganography has become widespread in modern times.

The mathematical model of steganography should be studied in detail in the development of various fields of steganography. It seems to me that the mathematical model of steganography will also give impetus to the creation of a
unified theory of steganography. The mathematical model of steganography further accelerates the development of steganography and increases its productivity. At the same time, the mathematical model of steganography will give rise to new fields of steganography. In this regard, the study of the mathematical model of steganography is one of the most pressing issues of modern times.

**About the mathematical model of steganography**

The formal mathematical model of the steganographic system has its own characteristics and capabilities.

In order to build a mathematical model of the steganographic system of information security, similar to the theory of cryptographic systems, it is necessary to consider the functional elements and mathematical operators that describe it in the abstract.

To do this, let's first take the following notation:

Denote by $M = \{M_1, M_2, ..., M_m\}$ possible messages intended for hiding, by $L = \{L_1, L_2, ..., L_l\}$ set of possible containers to be hidden and let us by $E = \{E_1, E_2, ..., E_n\}$ denote the number of possible filled containers (Containers full of hidden information are also called steganograms). Through the following

$$\varphi = \{\varphi_1, \varphi_2, ..., \varphi_k\}$$

note the set of mapping, here

$$\varphi_i: (M, L) \rightarrow E, i = 1,2, ..., k.$$  

Next, let's define the inverse mapping as follows

$$\varphi_i^{-1}: E \rightarrow (M, L).$$

This mapping opposes each element of the set $E$ to the element of the set $M$ and the element of the set $L$.

Then let us note the set of keys $K = \{K_1, K_2, ..., K_k\}$ so that for any $i = 1,2, ..., k$ the mapping $\varphi_i \epsilon \varphi$ is uniquely determined by the key $K_i$, i.e.

$$\varphi_i: (M, L) \xrightarrow{K_i} E$$

Each specific $\varphi_i$ mapping of the set $\varphi$ corresponds to the method of placing information from the set $M$ in a container of set $L$ using a specific $K_i$ key.

Then we note the set of keys $K^* = \{K_1^*, K_2^*, ..., K_k^*\}$, in general, let $K \neq K^*$.  
\[ \varphi^{-1} = \{\varphi_1^{-1}, \varphi_2^{-1}, ..., \varphi_k^{-1}\} \]

all elements of the inverse mapping set are given with the corresponding key.

\[ \varphi_i^{-1}: E \xrightarrow{k_i^*} (M, L) \]

Each specific \( \varphi_i^{-1} \) of the set \( \varphi^{-1} \) corresponds to the method of extracting (detecting) data from a filled container using the \( K_i^* \) key. If the key \( K_i^* \) is known, then the only response element from the set \( M \) and the set \( L \) is possible as a result of the extraction (detection) operation:

\[ (M_j, L_l) \rightarrow \varphi_i^{-1}(E_w, K_i^*) \]

For a solid steganographic system, the following equation is true:

\[ (M_j, L_l) = \varphi_i^{-1}(E_w + \varepsilon, K_i^*) \]

That is, a small change (up to \( \varepsilon \) quantity) in a full container does not cause the information to be inaccurate.

Weak stegosystem for any sufficiently small \( \varepsilon \) quantity characterized by the performed of inequality.

\[ (M_j, L_l) \neq \varphi_i^{-1}(E_w + \varepsilon, K_i^*) \]

Thus, the abstract definition of a steganographic system includes sets \( M, L, E, \varphi, \varphi^{-1}, K \), and \( K^* \) (open texts, empty containers and steganograms (full containers), sets of direct and inverse mappings, and corresponding keys sets).

The data source generates a flood of \( I_j \) information from the set \( I = \{I_1, I_2, ..., I_m\} \), so that after the initial conversion in the precoder, information is formed from the set \( M \) in the form of \( M_j \). Thus, the precoder performs the function of pre-preparing the placement of information in a container (for example, converts information into a specially formatted digital data array).

Each \( M_j \in M = \{M_1, M_2, ..., M_m\} \) corresponds to the probability \( P(M_j) \). The distribution of probabilities of a random process is given by the probability distribution (axis) of random quantities, ie, given by the set of probabilities:

\[ P_M = \{P(M_1), P(M_2), ..., P(M_m)\} \] \hspace{1cm} (1)

From source the set of the containers \( L = \{L_1, L_2, ..., L_l\} \) spawns a stream of empty containers \( L_u \). The operation of a container source can be described by any random process in which the concrete result is a \( L_u \) container. In this case, we are
dealing with random containers, to which the corresponding probabilities can be assigned:

$$P_L = \{P(L_1), P(L_2), ..., P(L_l)\}$$

In practice, other types of containers are often used - containers whose formation is not described by random processes. In this case, the source of the containers operates in a determined manner provided by either the competent authority or the attacker. In the first case, the container used is not formed randomly, the competent party selects it for any non-stochastic features. In the second case, the source of the containers is under the control of the attacker, and the containers themselves are formed in a manner determined by the attacker. Thus, we have the attached container.

In the simplest case, a set of empty containers contains only one element, which is used to place the data from the transmitting side through the communication channel and their confidential transmission.

The generated $L_u$ container is processed by the container features accounting unit. The main function of the container features accounting unit is to identify the features that will be used when placing $M_j$ data in the incoming $L_u$ container.

In a steganographic system, the key source $K$ and (or) $K^*$ generate a flood of keys. $P(K_i)$ probability for each $K^* \in K = \{K_1, K_2, ..., K_k\}$ key, and for each $K^*_i \in K^* = \{K^*_1, K^*_2, ..., K^*_k\}$ key The probability $P(K^*_k)$ corresponds. The production of keys is given by the probabilities of random processes:

$$P_k = \{P(K_1), P(K_2), ..., P(K_k)\},$$

$$P^*_k = \{P(K^*_1), P(K^*_2), ..., P(K^*_k)\}$$

(1)-(3) the set of a priori probability values constitutes the a priori knowledge of the attacker's source of information and key source. In fact, these sets characterize the attacker's a priori knowledge of the possible "weakness" of the steganographic system.

Selecting the $K_i$ key determines the specific $\varphi_i$ mapping from the set of $\varphi$ mappings. A steganogram (full container) is formed for the incoming $M_j$ information and the received $L_u$ container, taking into account these features, with
the help of the $\varphi_i$ mapping corresponding to the selected $K_i$ key:

$$E_w = \varphi_i(K_i, M_j, L_u),$$

$\forall i \in [1,2, ..., k], j \in [1,2, ..., m], u \in [1,2, ..., l], w \in [1,2, ..., n], n \geq m$

The $E_w$ steganogram is transmitted through a specific channel to the receiving point and can be intercepted by the attacker. At the point of acceptance, the initial data and the empty container are recovered from the $E_w$ steganogram with the help of the inverse reflection $\varphi_i^{-1}$ (given with the key $K_i^*$):

$$(M_j, L_u) = \varphi_i^{-1}(K_i, E_w)$$

When an $E_w$ steganogram is transmitted through a communication channel, the transmitted steganogram may be distorted as a result of the attacker's effect on the $E_w$ steganogram. In this case, the receiving party will add a certain mixture to the resulting full container and to the container when transmitted through the communication channel: $E_w + \varepsilon$. Execution of the inverse reflection operation $\varphi_i^{-1}$ given by the key $K_i^*$, in this case, leads to the formation of a certain value of the transmitted data and the transmitted empty container, ie:

$$(M_j^*, L_u^*) = \varphi_i^{-1}(K_i, E_w + \varepsilon).$$

For a fine steganographic system, the inequality $M_j^* \neq M_j$ should cause the data to fail. That is, the value $M_j^*$ generated during a small ($\varepsilon \neq 0$) distortion of the container should not cause the internal data to be read (when $\varepsilon \neq 0$, the $M_j$ data is destroyed).

Working steganographic systems are resistant to the effects of a full container. This means that when $\varepsilon \neq 0$, the value $M_j^*$ is contrasted with one of the possible data (ideally with the data $M_j$). At the same time, the $E_w$ container received from the communication channel may not store any internal data, ie the $M_j^*$ value extracted from the container may not be opposed to any of the probable data. The internal data detector function decides whether the internal data is in the $E_w$ container based on the value $M_j^*$ obtained. Thus, the value of the $S_j$ detector can be interpreted as a binary solution. The decryption itself is carried out in the message decoder, the main function of which is to compare the extracted $M_j^*$ estimates with one of the possible
messages $M_j$ and convert the latter into a message $I_j$ information given to the recipient.

The attacker can capture the $E_w$ steganogram. In this case, he can try to calculate the aposterior probabilities of different possible data and different possible keys that can be used in the formation of the $E_w$ steganogram:

$$P(M|E_w) = \{P(M_1|E_w), P(M_2|E_w), \ldots, P(M_m|E_w)\}$$  \hspace{1cm} (4)

$$P(K|E_w) = \{P(K_1|E_w), P(K_2|E_w), \ldots, P(K_m|E_w)\}$$  \hspace{1cm} (5)

The aposterior probabilities (4-5) constitute aposterior knowledge of the attacker's keys $K = \{K_1, K_2, \ldots, K_k\}$ and information $M = \{M_1, M_2, \ldots, M_m\}$ after capturing the $E_w$ steganogram. In fact, the sets $P(M|E_w)$ and $P(K|E_w)$ are a set of assumptions that correspond to probabilities.

This section provides a definition of weak and robust steganosystems, as well as probabilistic indicators that characterize the attacker's knowledge of cryptographic keys and incoming messages. A promising direction for future research is the analysis and theoretical substantiation of criteria and indicators of the effectiveness of steganographic data protection systems. It is also planned to study the properties of known steganosystem samples in accordance with the presented indicators and efficiency evaluation criteria.

**Application of steganography in digital images in JPEG format**

Digital images in JPEG (Joint Photographic Experts Group) format are sometimes used as steganographic containers in steganography. To do this, you must first build a mathematical model of a digital image in JPEG format. The mathematical model of a digital image in JPEG format includes the following.

When using a digital image in JPEG format as a steganographic container, the quantized coefficient of discrete cosine transformation (DCT) can be divided into 64 groups depending on the place. One of these groups is the DC-coefficient group, and the other 63 groups are the AC-coefficient group.

Suppose that $n$ the number of DCT coefficients in image, $m$ is the number of groups, $l$ is the number of bits in the entered message, $n_i$ is the number of coefficients in group $i$. 
CONCEPTS FOR THE DEVELOPMENT OF SOCIETY’S SCIENTIFIC POTENTIAL

\[ n = \sum_{i=1}^{m} n_i \]

\( k \) is the number of interchangeable elements of the image, \( k_i \) is the number of coefficients in group \( i \) (the DC coefficient of the image does not change in the hidden transformation: \( k_1 = 0 \)), \( k_i \) is the number of non-zero AC coefficients in group \( i = 2, \ldots, 64 \), \( x_0 \) is the number of DCT coefficients in group used to enter information., \( p_i \) is the probability of changing the non-zero AC coefficient of group \( i \) as a result of data entry:

\[ p_i = \frac{x_i}{2k_i}, \]

\( \delta(a = b), \delta(a = b, c = d) \) let's define the expressions as follows:

\[ \delta(a = b) = \begin{cases} 1, \text{agar} \ a = b \\ 0, \text{agar} \ a \neq b \end{cases} \]

\[ \delta(a = b, c = d) = \begin{cases} 1, \text{agar} \ a = b, c = d \\ 0, \text{agar} \ a \neq b \vee \forall a c \neq d \end{cases} \]

**About algorithms for embedding data in image in JPEG format**

The parameters of this model are mainly F5 and nsF5 placement algorithms. That is, the F5 steganographic algorithm to hide information in digital images in JPEG format uses matrix placement to reduce the absolute value of non-zero AC coefficients to place information in a digital image in JPEG format. The nsF5 algorithm is a modification of the F5 algorithm that uses "wet paper" codes to reduce the external distortion of the histogram of DCT-ratios. The nsF5 algorithm is considered to be the most stable method of placing information on digital images in JPEG format.