REPUBLIC OF AZERBAIJAN

On the rights of the manuscript

A B S T R A C T

of the dissertation for the degree of Doctor of Philosophy

DEVELOPMENT OF METHODS AND TECHNOLOGIES FOR COVERT TRANSMISSION OF RESRTICTED-ACCESS INFORMATION OVER INTERNET CHANNELS

Specialty:	3338.01 – System analysis, management and processing of information
Field of science:	Technical Sciences
Applicant:	**Esmira Ali Mustafayeva**

Baku – 2022

The work was performed at the Laboratory of Modeling of Queuing Inventory Systems of the Institute of Control Systems of ANAS.

Dissertation advisor: Professor **Vagif Alijavad Gasimov**, Doctor of Technical Sciences

Official opponents: Corresponding Member of ANAS, Professor **Ismayil Mahmud Ismayilov**, Doctor of Technical Sciences

Professor **Əlakbar Ali Ağa Aliyev**, Doctor of Technical Sciences

Professor **Ramin Rza Rzayev**, Doctor of Technical Sciences

ED 1.20 Dissertation Council of the Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at the Institute of Control Systems of the Azerbaijan National Academy of Sciences.

Chair of the Dissertation Council: Full Member of ANAS, Professor **Ali Mahammad Abbasov**, Doctor of Technical Sciences

Academic Secretary of the Dissertation Council: Professor **Naila Fuad Musayeva**, Doctor of Technical Sciences

Chair of the scientific seminar: Professor **Farhad Heydar Pashayev**, Doctor of Technical Sciences

GENERAL DESCRIPTION OF THE DISSERTATION

The relevance and maturity of the topic. To ensure information security at the necessary level in this modern age, it is essential to create, develop and improve an appropriate infrastructure, as well as a security system, of information resources. However, it is no secret that scientific and technological progress in the field of information technology, along with positive advances, has led to a number of serious problems in the field of information security. These problems can include such dangerous manifestations as illegal interference with computers, computer systems and networks, theft, misappropriation, covert acquisition (interception), transmission (leakage) of computer information.

Nowadays, the Internet and its information services are widely used all over the world for the purpose of information exchange. The Internet provides equal opportunities for all people and organizations using its services, including criminal and terrorist groups and individuals. Telecommunication systems, computer and information networks, including the Internet, are widely used in professional activities by individuals, politicians, businessmen, public, private and religious organizations, criminal terrorist groups and intelligence services of rival (enemy) countries, and act as tools and means of information warfare, confrontation, war, cybercrime and cyberterrorism. In this regard, comprehensive measures must be taken in the field of information security.

In practice, from the point of view of security in the transmission of confidential information via communication channels, there is a serious need to conceal its content and the very fact of transmission. Problems of covert transmission of information are dealt with in the science of steganography. Obviously, steganography does not replace cryptography, but only supplements it.

It should be noted that the introduction of certain restrictions or even bans on the implementation and use of cryptography without a license (or permission) in many countries has given a special impetus to the development of steganography. At the same time, an all-round development and widespread use of the general-purpose Internet, copyright protection, protection of personal privacy, essential in the

exchange of information via the Internet, organization of e-commerce, e-banking operations, prevention of activities of hackers and terrorists, and other similar ongoing challenges require the implementation of new methods and means of information protection. On the other hand, the rapid development of information technology and the Internet creates opportunities for the implementation of proposed new methods, creation and use of covert information exchange channels.

The dissertation investigates the methods of creating covert information exchange channels in the Internet, the use of such channels for the transmission of restricted information, as well as the possibilities, methods and technologies of covert transmission of information of various content.

The object of the research is restricted information and the network in which it is processed and transmitted.

The subject of the research is creation and analysis of channels for covert transmission of restricted information.

The aim of the research is to develop methods and technologies for creating covert channels on the basis of steganographic methods to ensure reliable and secure transmission of restricted information via the Internet.

Research methods. The research carried out in this dissertation work used methods of information theory, systems analysis, mathematical modeling, probability theory, set theory, data compression algorithms and experiments.

Main ideas put forward for defense.

1. A model of a steganographic system and a covert channel;
2. A method for selecting a covert channel and an embedding algorithm depending on the characteristics of the data to be hidden;
3. Methods for creating covert channels via social networks (including WhatsApp) and email service;
4. A method for estimating the bandwidth of covert channels implemented in Internet protocols;
5. Methods for analyzing and evaluating the stability of steganographic systems and the durability of stego-containers;
6. A modified LSB method that provides more reliable concealment of information in graphic files;

7. A character spacing method for hiding information in text-type containers;

8. An algorithm for selecting characters in a pseudorandom sequence using the combination of the linear congruential method and the quadratic Feigenbaum function to improve the reliability of the character spacing method for hiding information in text-type containers.

Scientific novelty of the research.

1. A model of a steganographic system and a covert channel that ensures reliable transmission of restricted information has been proposed.

2. Technologies for creating covert channels in social networks on the Internet have been analyzed, and methods for creating covert channels for transmission via WhatsApp and email services have been developed.

3. Methods for selecting containers, covert channels and embedding algorithms depending on the characteristics of the data to be hidden have been developed.

4. The covert channel bandwidth based on the variation of packet length in Internet protocols has been investigated and the appropriate analytical expression has been obtained.

5. An approach to the evaluation of the durability of stego-containers used for covert information transmission over Internet channels has been proposed.

6. To improve the reliability of hiding information in graphic files, a modified LSB method and the corresponding algorithm of its implementation have been developed.

7. To improve the reliability of the character spacing method for efficient concealment of information in text-type containers, an algorithm for selecting characters in a pseudorandom sequence using the combination of the linear congruential method and the quadratic Feigenbaum function has been developed.

Theoretical and practical significance of the research. The development of technologies for creating covert channels over the Internet allows solving the problems of information security by protecting restricted information, development and practical

application of covert transmission methods. At the same time, the results of the research can be used to identify covert information transmission channels created for malicious purposes.

Validation and implementation of the results. The main results of the dissertation were discussed at the following scientific-technical conferences:

- Information Processes and Technologies "Informatics 2014", 7th International Scientific and Practical Conference, Sevastopol, 2014;
- Qafqaz University, 3rd International Scientific Conference of Young Researchers, Baku, 2015;
- National Aviation Academy, "Creative Potential of Youth in Solving Aerospace Problems", 2nd International Scientific and Practical Youth Conference, Baku, 2017;
- 16th Information Technologies and Mathematical Modeling (ITMM-2017) International Conference named after A.F. Terpugov, Tomsk, 2017.

The scientific and practical results obtained in the course of the research carried out for the dissertation have been applied at the Special Design Bureau of Space Instrumentation of the National Aerospace Agency and the Institute of Space Research of Natural Resources for covert transmission of confidential information in their official activities. The relevant acts of application have been obtained.

The author has 8 research articles on the topic of the dissertation published in peer-reviewed journals, including 5 articles in foreign journals. Of them, 1 was published in a journal indexed in the Web of Science database, 1 in a journal indexed in the Ulakbim (TrDizin) database, and 2 in a journal included in the list of the Higher Attestation Commission of Ukraine.

The dissertation work was performed at the Institute of Control Systems of Azerbaijan National Academy of Sciences.

The dissertation consists of an introduction, four chapters, a conclusion, and a list of references. The total character count of the dissertation is 180,596 (204,037), with the introduction containing 7,045 (7,872) characters, Chapter 1 — 46,307 (52,383) characters, Chapter 2 — 36,945 (41,798) characters, Chapter 3 — 45,272 (51,217)

symbols, Chapter 4 — 43,081 (49,159) characters, and the conclusion — 1,446 (1,608) characters.

THE CONTENT OF THE DISSERTATION

The **introduction** substantiates the relevance of the topic, defines the aims and objectives of the research, and describes the novelty of the obtained results, their practical significance, and the implementation of the results of the study.

Chapter 1 highlights the problem of protecting restricted information in information security systems and examines the role of steganography in solving these problems. For this purpose, an analysis of the state of the art in the development of steganographic methods and means is conducted. On the basis of the analysis, the dissertation solves the problems of eliminating the drawbacks of some modern steganographic methods. For instance, the spacing method in textual steganography, the LSB method in digital steganography and network steganography are considered. Further in this chapter, steganographic systems, covert channels and their model are constructed, and methods for creating covert channels are analyzed.

One of the basic concepts of steganography is the concept of steganographic system (stego-system). A steganographic system (stego-system) is a set of methods and tools used to create a covert channel of information transmission. In other words, a steganographic system is the totality of data, containers and transforms that connect them.

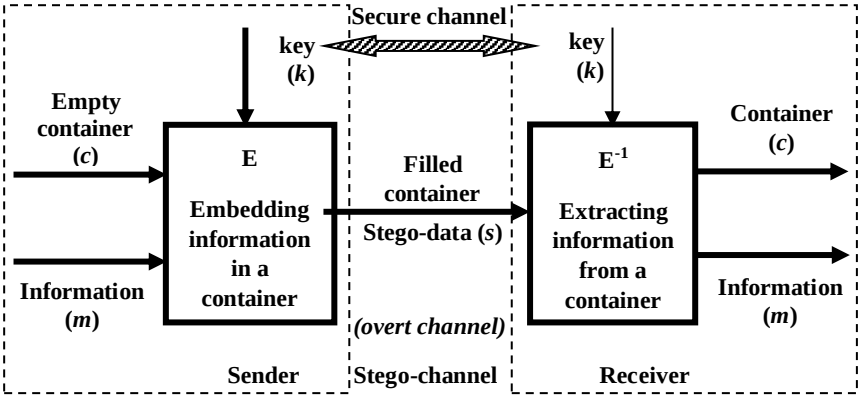

Fig.1. Block diagram of a steganographic system

In the general case, a steganographic system is the totality of data, containers and transforms that connect them together (Fig.1.).

Information (*M*) is the sensitive information, the existence or the fact of transmission of which needs to be hidden. $M=\{m_1,m_2,...,m_n\}$ is the set of sensitive information.

A *container* (C) is a non-confidential object (physical object, data storage medium, information resource, file, etc.) that is used for hiding information. $C=\{c_1,c_2,...,c_q\}$ is the set of containers, where q>>n. Multimedia file formats, such as plain text, can also act as data and containers.

A container without embedded data is called an *empty container* or the original container. A container in which confidential information is entered (embedded) is called a *filled container* (result container), that is, a container (c_m) in which some protected information *m* is stored. A filled container is generally referred to as *stego-data*.

The set of stego-data (filled container) is indicated as follows:

$$S=\{(c_1,m_1), (c_2,m_2) ,..., (c_q,m_q)\} = \{s_1,s_2,...,s_q\}.$$

One of the main requirements for steganographic containers is that the filled container should not be visually different from the original container, or human sensory organs (eyes, ears) should not feel this difference.

A model of a steganographic system can be represented as follows:

$$SteqoSys = (M, C, S, K, E, E^{-1}). \quad (1)$$

Here,

K is the set of steganographic keys.

E and E^{-1} are steganographic transforms (the hiding and extraction processes).

A steganographic key ($k \in K$) is additional information that is used for hiding specific confidential information in a container and extracting it from the container.

The steganographic key is transmitted by the parties to each other through a secured channel in advance.

E is the concealment method, the steganographic transform that performs the embedding of information into the container using the

steganographic key, and E^{-1} is the steganographic transform that performs the extraction of information hidden with the steganographic key from stego-data:

$$E: \ M \times C \times K \rightarrow S, \text{ i.e., } S = E(M,C,K),$$
$$E^{-1}: \ S \times K \rightarrow M, \text{ i.e., } M = E^{-1}(S,K). \qquad (2)$$

In Chapter 2, the possibilities of creating a covert information transmission channel over the Internet are explored, an experiment on creating covert channels on Facebook, Google+, cloud services, and the Web is conducted. Furthermore, channels for covert information transmission via email service are implemented.

It is possible to implement covert channels for exchanging confidential information between users using email. Several methods can be used here:

1. The method of hiding information in the body (text) of an email. A special editor is used to format the text in the email. Besides formatting the text of an email, this editor also offers the feature of inserting special characters, animations and signs in the email. These capabilities of text steganography make it possible to hide information in the open contents of an email message by using spacing, synonyms and conventional signs. Information can also be transmitted confidentially by selecting a font color that matches the background of the email text box.

2. The method of attaching any file (image, graphic, video, document, PDF file, table, etc.) to an email. In this case, the file attached to an email can be a stego-container in which confidential information is hidden by some steganographic method, i.e., it is possible to attach a pre-prepared stego-container to an email.

3. The method based on the principle of drafting an email and storing it in the mailbox as a draft. In this method, the parties wishing to exchange confidential information agree between themselves in advance on the mail server, the email identifier (username) and password, as well as the date and time of the exchange.

The exchange of confidential information is carried out in the following order:

- The sender of the information creates an email address corresponding to the pre-agreed username and password on the

pre-agreed mail server, for example, mailbox "esmira.mustafayeva" on the mail.ru server;

- Using the username and password, the sender of the information logs on to this server at the pre-scheduled time;
- The sender creates an email with contents that corresponds to the confidential information that needs to be sent, saves it in the Draft folder without entering the address to be sent to (click Save), and logs out of the email account;
- The receiver of the confidential information, logs on to the mailbox at the pre-scheduled time;
- The receiver takes (reads or copies) the draft email stored in the Draft folder and deletes it from the folder;
- The receiver logs out of the email service. This is how secret information is exchanged.

It should be noted that this method is implemented within a very short period of time. In other words, the creation of a draft email and its reading and deletion takes place at approximately the same time. The message is not transmitted over the Internet to another user and its server. Therefore, the information (the email) cannot get into the hands of an unauthorized person. As a rule, such mailboxes are intended for one-time use, i.e., after it has been used once, the mailbox is either deleted from the server to ensure confidentiality, or is no longer used.

Chapter 2 then discusses the implementation of a covert channel of information transmission via WhatsApp. The implementation of a covert channel via WhatsApp is investigated, and experimental tests are conducted. In the experiments, image files were used as containers. It is known that during exchange, images sent as a container in such formats as gif, bmp, png, etc., reach the receiver in JPEG format. For this reason, it was deemed appropriate to use images in JPEG format. It was established that when images in other formats are transferred via WhatsApp, they undergo changes and the embedded information is lost. For this reason, only steganographic applications supporting JPEG files were used in the experiments. It was also found through the analysis of these steganographic applications that most of the applications used the LSB method.

The experiments were conducted at different times, for different versions. In applications designed for the Android operating system, such as Stegais and Stegos, during the exchange of information, there is no change in the size of images in stego-containers implemented via WhatsApp, and the embedded information successfully reaches the receiver without any changes or losses. Thus, it is possible to create a covert channel in the WhatsApp channel by embedding the information in image files used as containers.

It is impossible to create a covert channel on the WhatsApp Web social network using steganographic software developed for the Windows operating system.

Further in Chapter 2, methods of steganographic concealment of information in Internet protocols are investigated, and covert channels via IP, TCP, UDP, ICMP protocols are created practically. Wireshark analyzer and PlayCap programs were used in the experiment to send packets, as well as the ping utility in the command line to pass ICMP packets through the network.

Chapter 3 analyzes the performance of steganographic systems implemented over the Internet. The security of stegosystems is described and evaluated on the basis of their stability. Analyzing the stability of stego-systems, we study their resistance to such attempts as hiding the fact of information transmission from intruders, destruction, distortion, breaking of transmitted hidden information by an intruder, as well as the features of confirmation or refutation of the transmitted hidden information. For this purpose, the dissertation examines the issues of selecting containers, covert channels and embedding algorithm based on the characteristics of the hidden information, evaluates the stability of stego-containers used for covert information transmission over the Internet channels, and estimates the bandwidth of covert channels implemented over the Internet.

The bandwidth of a covert information transmission channel is the maximum amount of information that can be embedded in a single element of the container.

As a rule, when a covert channel has a small bandwidth, it does not pose a serious risk. Therefore, a minimum bandwidth limit that is considered safe is set for covert channels.

Estimating the bandwidth is necessary in order to select a covert channel, container and embedding algorithm, depending on the characteristics of the information to be hidden.

Using methods of information theory, the bandwidth C of the covert channel can be calculated from the expression

$$C = \max_{n}\left\{\frac{I(X,Y)}{\tau}\right\}.$$

Here x is average packet transmission time, $I(X,Y) = H(Y) - H(Y/X)$ is mutual information of random variables X and Y describing the input and output characteristics of the covert channel, respectively.

Given that the average packet transmission time is determined as

$$\tau = \frac{2l_{fiks} + n - 1}{2\beta}$$

on the basis of the parameter n of the covert channel, the bandwidth C of the covert channel is calculated as follows:

$$C \approx \frac{2\left(log_2(2l_{fiks} - 1) - log_2\left(W\left(\frac{2l_{fiks} - 1}{e}\right)\right)\right)}{2l_{fiks} + \frac{2l_{fiks} - 1}{W\left(\frac{2l_{fiks} - 1}{e}\right)} - 1}\beta \quad (3)$$

Here β is the bandwidth of the communication channel, and $W(y)$ is the Lambert function defined as the roots of the equation $y = xe^x$, i.e., the inverse function of the function $f(w) = we^w$. When the expression

$$2l_{fiks} + \frac{2l_{fiks} - 1}{W\left(\frac{2l_{fiks} - 1}{e}\right)} - 1$$

in formula (3) takes a minimum value, the bandwidth C of the covert channel is maximum. This means that the parameter n of the covert channel should be taken as follows:

$$n \approx \frac{2l_{fiks} - 1}{W\left(\frac{2l_{fiks} - 1}{e}\right)} \quad (4)$$

Here, when selecting the quantity l_{fiks}, it must be taken into account that if the network layer protocol uses the IPv4 network protocol, the length of network and data link layer headers with Ethernet data link layer technology should be not less than 34 bytes, and not less than 54 bytes if the IPv6 protocol is used.

Taking the above into account, the relative bandwidth (C/β) of the covert channel at different values of the sum of header lengths (l_{fiks}) at network and data link layers of the OSI model is calculated, and the result is shown as a graph in Fig. 2. As can be seen from the graph, at the value of 34 of the parameter l_{fiks}, the relative bandwidth reaches its maximum value and equals approximately 0.021. Note that in this case, when using the IPv4 protocol, the parameter n of the covert channel takes the value 138.

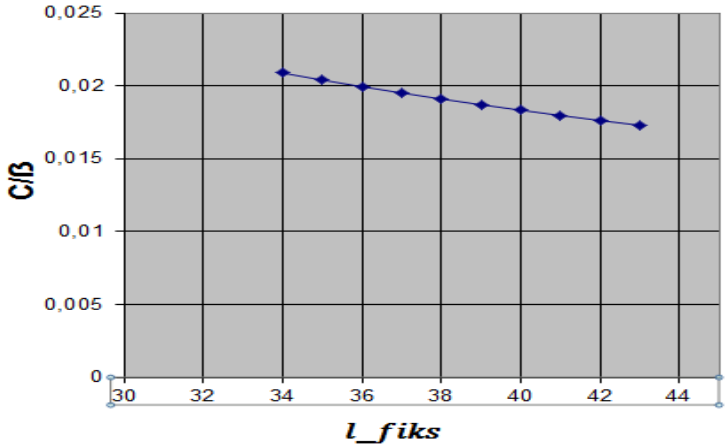

Fig. 2. Graph of the relative bandwidth (C/β) of the OSI model versus the sum of header lengths (l_{fiks}) at the network and data link layers

Thus, it is possible to create on a $5\,Gbps$ communication channel covert channels with a maximum bandwidth of $5\,Gbps$ ·

$0.021 = 105\ Mbps$ when using IPv4 protocols, and $5\ Gbps \cdot 0.014 = 70\ Mbps$ when using Ipv6 protocols.

Chapter 4 reviews software solutions for covert information transmission channels on the Internet and steganographic methods. Existing steganographic software tools are subjected to a comparative analysis based on the use of containers and algorithms.

A modified LSB method is proposed for reliable concealment of information in graphic files, and the character spacing method is proposed for concealment of information in texts.

To improve the reliability of concealment of information in texts, a method of pseudorandom character selection has been proposed. When analyzing steganographic applications, it becomes clear that the LSB (Least Significant Bit) method is used in most of these applications. This method is based on the limitations of human sensory organs. For example, the human eye cannot make out small changes in color made in the container as a result of concealment of information. The drawback of these methods is that the stego-container can fall into a rival's hands when sent through the communication channel and the hidden information is detected (extracted).

To eliminate this drawback, a modified version of the LSB method has been proposed. The proposed method uses two graphical files, one of which acts as the container, and the other as the steganographic key. The distinguishing feature of the method is that the data embedded into the container does not contain the data itself, but information about it. That is, the data bits hidden with the last bits of RGB pixel channels of the second graphic file (the steganographic key) are assembled modulo 2. Based on the result, the last bits of the RGB channels of the corresponding pixels of the container are changed. Since the key file is not transmitted through the communication channel, the hidden information cannot get into a rival's hands, as it cannot be opened using the container, which does not contain the modified and hidden data.

Suppose a 24-bit RGB bitmap image is used as a container. In such an image, each point (pixel) is encoded by three bytes. Here, the first byte determines the intensity of the color Red, the second byte the

intensity of Green, and the third byte the intensity of Blue, whose values together determine the color shade of that pixel. In these bytes, the small (rightmost) bits have much less effect on the hue of the image than the large (leftmost) bits. Replacing one or two of the small bits in these bytes results in a change that will not be perceptible to the human eye.

For visual purposes, the 24-bit RGB bitmap image hides information about the name of steganographic method — "LSB". The description of "LSB" information in binary codes will be "01001100 01010011 01000010". It takes four pixels, two bits per byte, to embed the 24-bit sequence in an image. Images before and after embedding these bit sequences into the pixels are shown in Fig. 3:

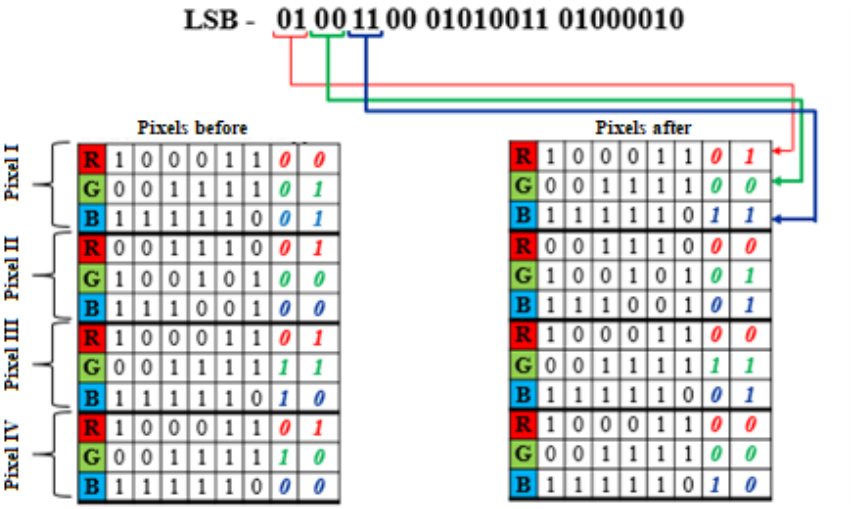

Fig. 3. Embedding a hidden bit sequence by the LSB method

The least significant bit method has a certain drawback, such as there are also applications based on the inverse algorithm, which can detect the hidden information. To eliminate this drawback, a modified version of the LSB method is proposed.

The proposed method uses two graphical files PIC1 and PIC2 (e.g., 24-bit RGB bitmap images) of identical size and graphical parameters. The PIC1 graphic file acts as the steganographic key and is used in the process of hiding confidential information in the

container and extracting it from the container. This file is pre-agreed upon by the parties sending and receiving the information. The parties transmit this file to each other in advance via a covert channel or take it from any agreed-upon public source (e.g., a web site).

The PIC2 graphic file is used as a container by the person sending the confidential information. It should be noted that the container does not contain transmitted data, but rather the information about that data. The last bits of the RGB channels of the container pixels are changed in accordance with the transmitted information and the PIC1 graphic file. The receiver of the modified PIC2 file recovers the confidential information using the previously received PIC1 file and container. Since the PIC1 file is not transmitted via a communication channel, it cannot fall into a rival's hands. Since the modified PIC2 container transmitted via the communication channel does not include the confidential data itself, it cannot be recovered from this file.

The secret information to be transmitted is first converted to binary code for embedding in the container. In this case, the information can be pre-encrypted to enhance security.

Under the proposed method, since the container contains 3 bits per image pixel, i.e., 1 bit per color channel, the maximum amount of information hidden in the container will be $L \leq 3*I$. Here, L is the size of the hidden information in bits, i.e., the number of bits embedded in the container, and I is the number of pixels in the container, i.e., the PIC2 image.

As already mentioned, it is also possible to hide two bits in each color channel of the pixels of the container, i.e., in each byte. In this case, the limit on the amount of hidden information is $L \leq 6*I$.

The algorithm for hiding information in the container can be described as follows. The first (second and third) bit of the hidden information is taken and assembled modulo 2 ($\oplus$) with the last bit of the first (second and third) byte of the first pixel of the PIC1 image. The result is placed respectively in the last bit of the first (second and third) byte of the first pixel of the PIC2 image (Fig. 4.).

All the bits of the information are assembled modulo 2 with the last bits of the color channels of the corresponding pixels of the PIC1

image and, depending on the result, are changed with the last bits of the color channels of the corresponding pixels of the PIC2 image. As a result, the graphic file of PIC2 is changed. However, as can be seen, this file does not contain the data to be transmitted. It only contains the information about whether the bits of the information being sent and the last bits of the color channels of the PIC1 file are identical or different. A person who does not have a PIC1 file cannot find the hidden information.

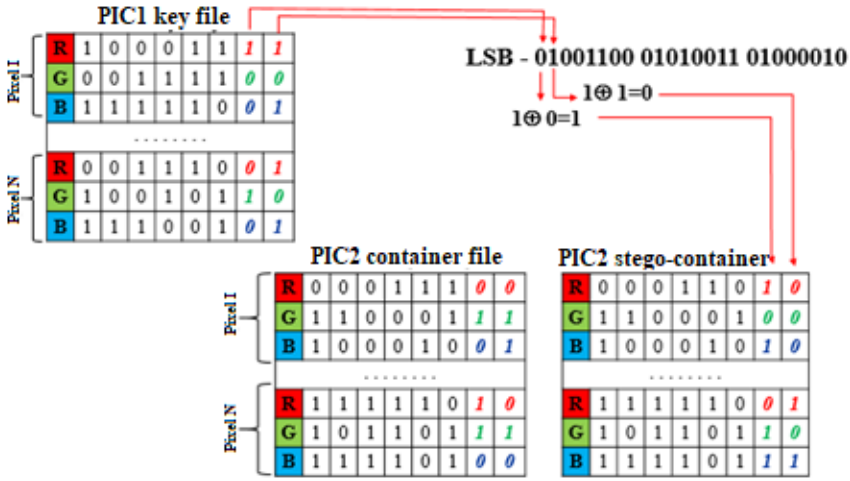

Fig. 4. Description of the modified LSB method

The modified PIC2 file is sent to the receiver over an open communication line without any risk. If the sent file falls into a rival's hands and it is suspected that the file contains hidden information, it will be impossible to retrieve the hidden information from the stego-container, because the PIC1 file is unknown.

The receiver of the PIC2 stego-container sequentially assembles the last bits of all channels of its pixels with the last bits of the PIC1 key file pixels modulo 2 to extract the hidden information. The result shows the bits of the hidden information.

The proposed modification of the LSB method with two graphic files eliminates the drawback of the regular LSB method. In this method, it is not the hidden data itself that is embedded into the graphic file used as a container, but the information about it. This information

is compiled from a second graphical file, which acts as the steganographic key with the hidden information. Since the steganographic key is not transmitted via a communication channel, the hidden information is protected more securely.

It should be noted that a bmp, jpeg, gif, etc. graphic file can be used as a container. The only requirement here is that both the container and the steganographic key must be of the same format and size.

Further in Chapter 4, the character spacing method for hiding information in Word documents is proposed.

It is well known that the majority of text files used in practice today are documents made in the MS Word processor. These documents offer many opportunities to hide information. One of them is using character spacing of the text. For this purpose, the information to be hidden is first converted to binary code (e.g., using the ASCII standard). After that, this information is hidden in the text documents by changing the values of the *Expanded* and *Condensed* parameters in the *Spacing* drop-down list in the *Character Spacing* section of the *Font* window of the MS Word processor (Fig. 5.).

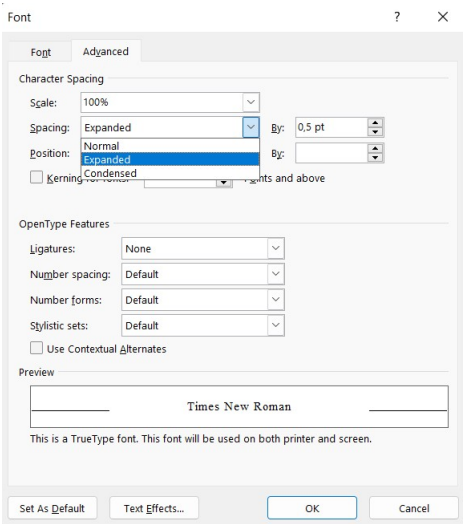

Fig. 5. Changing spacing parameters to hide information

It should be noted that if the value of the character spacing parameters *Expanded* (and *Condensed*) is set to 0-0.3 pt in a Word document, then the changes made in the text do not exceed the threshold visible to the human eye. Therefore, by changing the character spacing up to 0.3 pt, it is possible to encode information presented in binary format and to embed it in the document.

As we know, the initial value for these parameters in MS Word is 0, with an increment/decrement step of 0.05. Thus, by changing the character spacing from 0 to 0.25 with a step of 0.05 one can generate 10 encoding variants. This allows hiding more information in the document (Table 1).

Table 1. Character spacing values for encoding bit sequences

Encoded bits	Character spacing	
	Expanded	Condensed
000	0.05	-
001	-	0.05
010	0.1	-
011	-	0.1
100	0.15	-
101	-	0.15
110	0.2	-
111	-	0.2
0	0.25	-
1	-	0.25

It should be noted that in the proposed method, the size of a container needed to hide confidential information in the text can be much smaller than that of containers used in other existing methods of text steganography. For example, in the inter-word spacing algorithm one bit can be hidden in each word, in the inter-line spacing method 1-3 bits are hidden in a line, while in the proposed method 3 bits can be hidden in each character. In other words, to hide information consisting of 300 characters, the first method requires a text of 2,400 words, the second method requires about 18,000 words and the proposed method requires 100 words.

To more clearly demonstrate the principle of the proposed method, consider the following example. Suppose one wants to hide the information "*The workshop will be held tomorrow*" in an MS Word document. To do this, the information is first converted to binary code by the ASCII standard:

11010001 11100101 11101100 11101000 11101101 11100000
11110000 11110001 11100000 11100001 11100000 11111001
11101010 11100101 11110111 11101000 11110000 11101000
11101011 11111111 11100110 11111111 11101010

As can be seen, after converting the information consisting of 23 letters to binary code, we get a sequence of 184 binary characters. The resulting sequence is split into groups of 3 bits each, yielding a total of 62 groups. A container document containing at least 62 words is selected (Fig.6).

To hide the information, the character spacing values in the words of the container document are set in accordance with the values of the corresponding bit groups (Fig. 7).

To further improve the resistance of the character spacing method of hiding information in texts to steganalysis, it is advisable to introduce a process of entering hidden information into the container randomly, rather than sequentially. This process can be implemented using a pseudorandom number sequence, that is, it determines a chaotic selection of characters in the text in the container for embedding hidden information grouped by bits according to Table 1. There are many methods for processing a pseudorandom number sequence, but the effectiveness of methods based on different algorithms depends largely on the fields in which they are applied. In more serious issues, such as information protection, a combination of several methods is considered advisable.

In our study, in order to embed hidden information into a random sequence in a text-type container, we used a pseudorandom number sequence obtained by a combined use of the linear congruential method and the quadratic Feigenbaum function.

Fiziki proseslərdə bir neçə enerji forması mövcuddur. Onları bir neçə qrup daxilində birləşdirmək mümkündür. Enerji forması onun daxil olduğu qrupdan asılı olmayaraq sistemin halını göstərən xarakterik kəmiyyətdir. Burada enerjinin saxlanması qanunu hökm sürür, yəni qapalı sistemin ümumi enerjisi həmişə sabitdir. Yalnız sistemə kənardan təsir etdikdə (əlavə enerji verdikdə) onun ümumi enerjisi dəyişir. Mexaniki sistemin enerjisini kibernetik və potensial enerjinin cəmi kimi təsvir etmək olar.

Şək.6. Orjinal konteyner

Fiziki proseslərdə bir neçə enerji forması mövcuddur. Onları bir neçə qrup daxilində birləşdirmək mümkündür. Enerji forması onun daxil olduğu qrupdan asılı olmayaraq sistemin halını göstərən xarakterik kəmiyyətdir. Burada enerjinin saxlanması qanunu hökm sürür, yəni qapalı sistemin ümumi enerjisi həmişə sabitdir. Yalnız sistemə kənardan təsir etdikdə (əlavə enerji verdikdə) onun ümumi enerjisi dəyişir. Mexaniki sistemin enerjisini kibernetik və potensial enerjinin cəmi kimi təsvir etmək olar.

Şək.7. İnformasiya daxil edilmiş konteyner

The linear congruential algorithm was proposed by D. H. Lehmer in 1948; it is one of the algorithms created for generating uniformly distributed random numbers. The algorithm is based on the expression

$$x_{n+1} = (ax_n + c)\ mod\ m, n \geq 0 \qquad (5)$$

Here x_0 is the initial value $(x_0 \geq 0)$, a is the multiplier $(a \geq 0)$, c is a constant number $(c \geq 0)$, and m is the module $(m > x_0, m > a, m > c)$. The sequence obtained as a result of the algorithm implementation depends on the choice of the initial value of x_0, and with different values of x_0 we get a different sequence of random numbers. There are occasional repetitions in the sequence of numbers obtained with certain values of the quantities x_0, a, c and m, that is, these values cannot be chosen arbitrarily. The period of a linear sequence is equal to m only when it satisfies the following conditions:

1) c and m are coprime integers;

2) for each prime divisor p of m the number $b = a - 1$ is a multiple of p;

3) if m is a multiple of 4, then b is a multiple of 4.

The selection of the constant a with the above conditions satisfied guarantees a reasonably good result; of course, here it is also necessary to take into account that the condition $m>a$ set for the numbers a and m will also be satisfied. The constant c does not play an essential role in determining the value of a, and the main requirement here is that c must be an odd number. For example, if we take $m = 1024, a = 397$, and $c = 11$, then using the expression

$$x_{i+1} = (397x_i + 11)\ mod\ 1024, \quad (6)$$

we can process a sequence of random numbers up to 1024. A visual representation of a pseudorandom number sequence starting with $X_0 = 497$ is shown in Fig. 8, and as can be seen, the numbers have a uniform distribution.

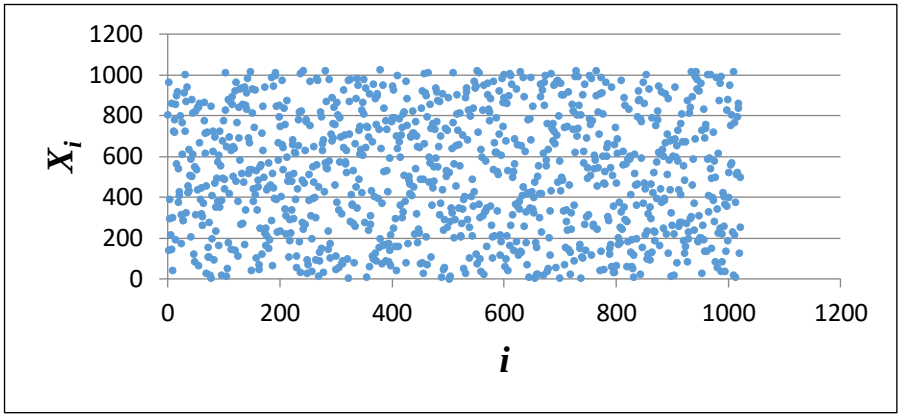

Fig.8. A sequence of pseudorandom numbers obtained by the linear congruential method

However, it should be noted that, as with most PRNGs, there is a certain correlation between consecutive elements in this method, and when several consecutive numbers from this correlation are known, it can be used to calculate others. Given this, it is possible to significantly complicate the process of calculating other elements from the known

elements of the sequence by weakening the relationship in certain points of the number sequence. To this end, in our study we propose to use the quadratic Feigenbaum function.

The quadratic Feigenbaum function is one of the functions reflecting the processes of deterministic chaos and is expressed by the following iterative formula:

$$y_{n+1} = r y_n (1 - y_n) \qquad (7).$$

Quadratic function (7) allows generating a sequence of numbers varying chaotically from 0 to 1 of y_i for values of the parameter r from 3.57 to 4. It is this chaotic property that generates a new sequence by selectively removing a portion of the sequence of numbers derived from expression (6). In the new sequence obtained, the relationship between the numbers is sufficiently weakened, which makes it difficult to calculate other elements using the known elements of the sequence.

A visual representation of a pseudorandom number sequence, obtained on the basis of the linear congruential method and the quadratic Feigenbaum function ($r = 3{,}9612345678$, $y_0 = 0{,}5678987654$), is shown in Fig. 9. From comparison with Fig. 8, it is clear that the numbers belong to a pseudorandom sequence that does not have a uniform distribution.

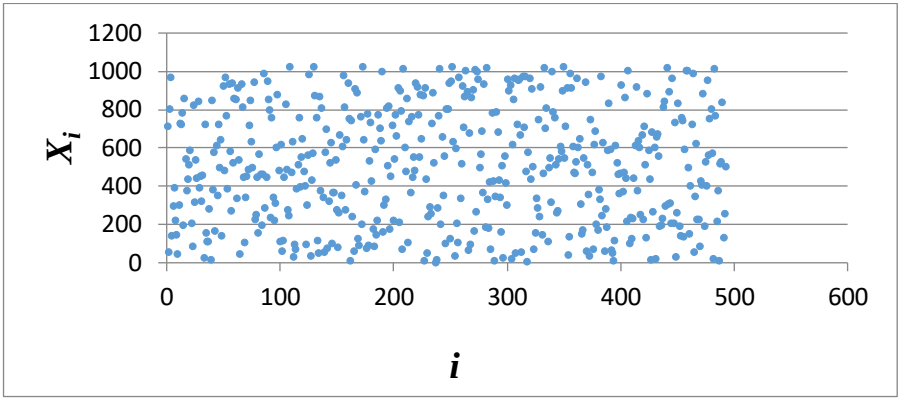

Fig.9. A pseudorandom number sequence obtained by the linear congruential method and the quadratic Feigenbaum function

The process of obtaining a new sequence is carried out in the following steps:

a) Based on expression (6), a sequence A consisting of n pseudorandom numbers is generated;

b) The generated number sequence is divided into $n/10$ groups q_i $(i = 1..n/10)$, each group consisting of 10 numbers;

c) Based on expression (7), a sequence B consisting of $n/100$ elements is generated (where the length of the fractional part of each element in fractional form must be at least 10 digits);

d) The first elements of the 1st group of element A are taken as the first elements of the new sequence C in the quantity corresponding to the 1st digit of the 1st element of the sequence B after the decimal point;

e) This process continues until the end of sequence A, and the results are added successively to sequence C.

The resulting sequence C is a pseudorandom number sequence obtained as a result of a combined use of the linear congruential method and the quadratic Feigenbaum function. The sequence of characters in the container is selected in accordance with this pseudorandom number sequence, and the information to be hidden is placed in the spaces between the characters in the selected sequence (Fig. 10).

The process of extracting hidden information from the stego-container is the reverse of the process of embedding information in the container, i.e., to extract hidden information from the stego-container, the receiver must execute this process in the reverse sequence. Along with Table 1, which is pre-agreed upon between the sender and the receiver, a pseudorandom sequence table is also sent to determine the position and sequence of characters in the text. The receiver retrieves the hidden information from the spaces between these characters in accordance with Table 1, determining the position and sequence of characters in the stego-container on the basis of the pseudorandom sequence table.

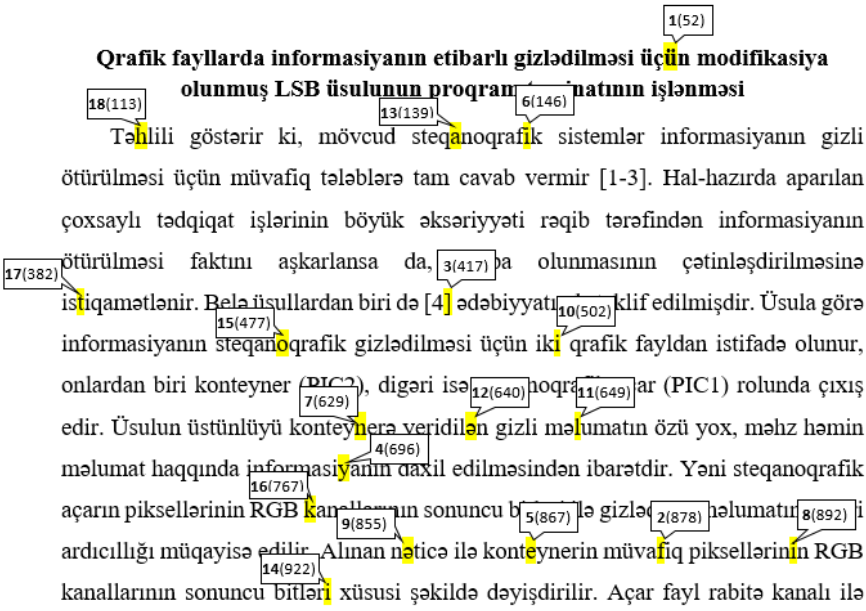

Fig.10. Selecting characters in a container based on a pseudorandom sequence

The confidentiality of information embedded in the spaces between characters of text containers on the basis of a pseudorandom number sequence obtained by the proposed combination of the linear congruential method and the quadratic Feigenbaum function is considered to be sufficiently enhanced. The aim of complicating the calculation of a pseudorandom number sequence used for chaotic selection of characters in containers in order to hide information is to prevent certain relationships between sequentially positioned elements.

MAIN RESULTS

1. A model of a steganographic system and a covert channel that ensures reliable transmission of restricted information has been proposed.
2. Technologies for creating covert channels in social networks on the Internet with the purpose of restricted information transmission have been analyzed, and technologies for implementing covert channels via WhatsApp and email services have been proposed.
3. Methods for selecting containers, covert channels and embedding algorithms depending on the characteristics of the data to be hidden have been developed.
4. On the basis of theoretical and information methods for calculating the covert channel bandwidth, based on the variation of the packet length in the Internet, the corresponding analytical expression has been obtained.
5. The problem of analysis and evaluation of the stability of steganographic systems and the durability of stego-containers has been considered, and appropriate recommendations and proposals have been developed.
6. A more efficient method of hiding information in text-type containers has been proposed, a new character spacing method has been developed, and appropriate software has been implemented.
7. To improve the reliability of the character spacing method for hiding information in texts, an algorithm for selecting characters in a pseudorandom number sequence obtained by a combined use of the linear congruential method and the quadratic Feigenbaum function has been proposed.

The main results of the dissertation appeared in the following publications:

1. Qasımov V.Ə., Məmmədov S.Z., Mustafayeva E.Ə. Korporativ kompüter şəbəkələrində informasiya təhlükəsizliyinin təmin olunmasında elektron imza texnologiyasının tətbiqi haqqında [Gasimov V.A., Mammadov S.Z., Mustafayeva E.A. On the application of the electronic signature technology to ensure information security in corporate computer networks] // BDU-nun xəbərləri. Fizika-riyazıyyat elmləri seriyası. No 2, 2007, pp.102-108
2. Qasımov V.Ə., Mustafayeva E.Ə. Word sənədlərində informasiyanın gizlədilməsinin simvollararası intervallar üsulu [Gasimov V.A., Mustafayeva E.A. The character spacing method of hiding information in Word documents] // Azərbaycan Milli Elmlər Akademiyasının xəbərləri. Fiziki-texnika və riyaziyyat elmləri seriyası. İnformatikanın və idarəetmənin problemləri. 2014, No 6, pp.124-129
3. Qasımov V.Ə., Mustafayeva E.Ə. Internetdə informasiyanın gizli ötürülmə kanallarının yaradılması üsulları [Gasimov V.A., Mustafayeva E.A. Methods of creating covert information transmission channels on the Internet] // "Milli təhlükəsizlik və hərb elmləri" elmi-praktik jurnalı. Azərbaycan Respublikası Silahlı Qüvvələrinin Hərbi Akademiyası. Baku. 2016. No 3. pp.122-128.
4. Gasimov V.A., Mustafayeva E.A., Hüseynova G. Implementing covert channels to transfer hidden information over WhatsAPP on mobile phones // International Journal of Engineering and Applied Sciences (IJEAS) ISSN: 2394-3661, Volume-6, Issue-2, February 2019.
5. Gasimov V.A., Amashov Y.A., Aliyeva F.P., Mustafayeva E.A., Mutin D.I. Bolnokin V.E. Development of the information security system effective structure for the distributed computer networks. // IOP Conf. Series: Materials Science and Engineering. Vol. 537, IOP Publishing. 2019 (***Web of Sciences***)

6. Мустафаева Е. Принципы выбора контейнеров для стеганографических систем. [Mustafayeva E. Principles of selecting containers for steganographic systems] // International Journal of 3D Printing Technologies and Digital Industry, Volume: 4 - Issue: 3, pages: 264-269, 04.11.2020 (***Ulakbim, TrDizin***)
7. Мустафаева Е.А. Исследование зависимости стеганографических систем от характеристик скрываемой информации. [Mustafayeva E.A. A study of the relationship between steganographic systems and the characteristics of hidden information] // Проблеми інформатизації та управління, Vol. 3 No 67 (2021), pp. 46-53 (***Higher Attestation Commission of Ukraine***)
8. Gasimov V.A., Mammadov J.I., Mustaphayeva E.A. Intersymbol interval method for hiding secret information based on pseudo-random number sequences. Проблеми інформатизації та управління, Vol. 1 No 69 (2022), pp. 18-23 (***Higher Attestation Commission of Ukraine***)

The defense of the dissertation will be held at 16^{00} on 26.10.22 at the meeting of the ED 1.20 Dissertation Council operating at the Institute of Control Systems of the Azerbaijan National Academy of Sciences.

Address: Institute of Control Systems of the Azerbaijan National Academy of Sciences, 68 B. Vahabzade Str., Baku AZ1141

Dissertation is accessible at the library of the Institute of Control Systems of Azerbaijan National Academy of Sciences.

Electronic versions of the dissertation and its abstract are available on the official website of the Institute of Control Systems of the Azerbaijan National Academy of Sciences.

Abstract was sent to the required addresses on 23 sept. 2022.

Signed for print:

Paper format: A5

Page count:39042

Number of copies: 30