UDC 004.056.5

A METHOD FOR HIDING INFORMATION IN JPEG IMAGES

R.G. Hakobyan

State Engineering University of Armenia (Polytechnic)

Steganography methods, according to which the container is the image stored in JPEG, are poorly understood and not often used. The reason is that the JPEG image is subjected to compression and restoration of image distortions. These distortions cannot be seen with the naked eye, but it affects the hidden information whose extraction from the container is not possible. Because of this, the JPEG often hides information in a table of coefficients, which leads to strong limitation of hidden information. This paper proposes a method - the concealed information is recorded in audio format, and then hidden in the image. In this case, the audio information is also distorted. But unlike the text, in which the loss of the symbols completely distorts the phrase, sound distortions do not affect the speech content, and it may be possible to fully understand the meaning of the hidden information. The developed software package at tests showed that at 20% compression of JPEG distortions are barely heard, and at 80% compression, the distortion cannot be understood. When compressed to 60%, the security of the hidden information can be guaranteed.

Keywords: steganography, image, container, hidden information.

Introduction. Currently, the general trend of enhanced spread of information technologies and, in particular, of Internet accessibility, is accompanied by an increasing threat of unauthorized use and loss of data, as well as by directed attacks against data domains. This growing intervention of IT into the daily life generates the need for special attention and protection of data within any information system.

Protection of the data from targeted adverse impact/damage has two main approaches [1, 2]:

- Cryptography;
- Steganography.

In case of cryptography, it is evident that the data are meant to be classified and are encrypted, however, it is not accessible without the key code.

Steganography is one of the efficient methods of data protection. Steganography is the science that provides a protected storage and transmission of data thus ensuring its secrecy, access and integrity. The modern means of steganography allow to hide classified information within image, music and other types of files that can be held on various carriers, such as local computers or networks. Collection of graphic, sound and other types of "harmless" files is not seen as suspicious and attract attention,

unlike the presence of encrypted files. In addition, steganography is more convenient in terms of programming, particularly as it is easier and therefore cheaper. However, the major advantage of steganography is that it allows to hide the evidence of protected information. Moreover, data can also undergo encrypting prior to being steganographed.

Basic Concepts and Application of Steganography. The numerous ways in which steganography can be applied can be classified into the following groups

- Embedding information for the purpose of hidden storage of transmission;
- Watermarking;
- Fingerprinting;
- Captioning.

The first type of application is mainly used for hidden storage and transmission of information, and this is the particular focus of the present study, while other types will not be discussed.

Basic concepts:

Container – No secret data that is used for hiding of messages

Messaging - secret/classified data the presence of which needs to be protected

Key – a set of encrypted data that allows access to the initial secret data.

A steganographic algorithm includes two types of modifications: (a) direct steganographic modification and (b) reverse steganographic modification.

Direct steganography allows to receive a full container on the basis of messaging, the empty container and a key, while reverse steganography results in hidden messaging on the basis of the full container and a key.

The steganographic system consists of a container, keys, messaging and modifications.

Steganographic stability is the main feature of the steganographic system that characterizes its ability to avoid detection.

Intruder – an individual or a group of individuals who can try to access the content of the message or damage it.

Attack – targeted action of an intruder against steganographic system for access to the data.

The primary and most important task in acting against steganography is detection of hidden data or of its transmission.

The major features that characterize the steganographic system are the organization of its keys and the algorithm. In terms of the organization of keys, the system can be keyless, symmetric, non-symmetric and interconnected.

For operating keyless steganosystems, the availability of algorithm is sufficient and a special steganographic key is not required. Thus, the stability of keyless

steganosystems is based only on the secrecy of direct and reverse operations, which contradict the principle of A. Kerkhovs which stipulates that reliability of the IT system should be defined only by the secrecy of the key rather than the methods employed.

The definition of symmetric, asymmetric and interconnected steganographic systems corresponds to that of similar cryptographic systems, the only difference being that process of direct data cryptography and of its processing are replaced with the direct and reverse steganographic modification. The symmetric system envisages use of symmetric keys, e.g. when keys of direct and inverse modifications are easily derived one from another. In case of the non-symmetric system, there are separate close and open keys that cannot transform into each other. The interconnected system applies both approaches.

Classification of means of data steganography. There are numerous algorithms for data steganography and the new ones are being continuously developed. At the same time, there are several approaches to classification of these algorithms that are based on specific features. The most appropriate mode of classification of algorithms is based on the method and location of the data hiding. In terms of the structure of the steganographic algorithm, steganosystems can be based on [3]:

- Addition;
- Replacement;
- Generation.

The classification above allows to detect the general features of steganographic algorithms and can be useful while investigating the algorithms and analyzing steganography.

In case of algorithms based on addition, the data are placed in those parts of the file structure that are ignored by the file-reading application program. In this case, the size of the given file can increase, however its reproduction by corresponding programs remains unchanged.

The replacement steganographic algorithms are based on replacing the data to be hidden with certain parts of the meaningful data within the container without changing its size. In other words, they search the file structure to identify parts that can be changed without having a notable impact on the file. These parts of the file are then replaced with the data to be hidden.

In this case, it is necessary to select very carefully and reasonably the changed parts of the container, so that it is not damaged or altered notably.

The two types of steganography mentioned above imply the presence of a message and a container. In case of generation methodology, the container is created/generated directly as an outcome of steganography - on the basis of the hidden message and, potentially, with the use of the key.

The structure of JPEG format. In the JPEG algorithm, the original image is an NxN two-dimensional array with elements represented by brightness of color or pixel. The wrapping array is performed in three phases:

- Discrete cosine transformation,
- Step of quantization,
- Secondary phase of compression.

A high efficiency of compression that is possible to achieve with this algorithm, is based on the following: factors of matrix frequencies that are formed after cosine transformation,

This is important because most parts of the graphic image are formed by the data of low frequency and, therefore, the high frequency components of the matrix can be safely removed. This removal is carried out through rounding of frequency coefficients. After rounding, components of low frequency generally remain in the upper left corner of the matrix. The value of the rounded matrix is encoded by additing zeros. As a result, the graphic image is compressed by more than 90%, losing the quality of the image insignificantly at the stage of matrix rounding.

Investigation and definition of the problem. As it has been mentioned, steganography is considered an effective method of data protection. Unlike cryptography that implies modification of data with a cryptographic system without a key unreadable and undecipherable unit, steganography hides the fact of keeping or transferring secret data. This means that a potential intruder should above all be able to detect the presence of the hidden information and only then attempt to restore it [4].

At the initial stage, various steganosystems allowing to hide the data within JPEG files have been studied. The existing systems that work with JPEG files are few and are alike, although these files are most often used to save images. Most often we find applications (?) that allow steganography within JPEG files in the Internet. However, those are not sufficiently secure as most of them add hidden data at the end of the file thus increasing its volume and creating a potential alert for the intruder.

"Real" steganography envisages hiding the data within the pixels of an image. Similar steganographies have high-level protection because after compression of the image it is hard to detect whether its size has changed due to the presence of hidden data or as a result of compression.

In the majority of programs available on the Internet, the steganography of images is implemented after quantization phase, however in such a process the amount of the hidden data is quite limited (12% of the general image size). Such programs are, for example:

- JSTEGWIN;
- JPHS, uses generator of random number that is based on the BlowFish algorithm;

- JSTEG;
- and others.

Let us consider one of these examples: the JSTEG program.

JSTEG is one of the programs that allows steganography through incorporation. It represents steganography through LSB (Least Significant Bits). It is one of the first steganographic programs. Its advantage is in simplicity while one of the main shortcomings is that it resembles steganography on the basis of LSB within BITMAP files. Within BMP files, it is preferable to hide information through modification of color bites; however it leads to the distortion of the image. In order to avoid this distortion it is necessary to select and modify a bit that does not affect the overall quality of the image visually. Thus, the bit with the least significance (LSB) should be taken.

It is known that creation of a JPEG file results in the reduction of file size (the compression algorithms resulting in data losses within JPEG files are a reason for distortions). Therefore, at the time of decoding/recovery of hidden information, it is not possible to receive the original file, but the recovered information will be close to the data as it had appeared before hiding.

In order to avoid such occurrences, we will consider sound files as the original data to be hidden. In this particular case we will not see a change in the text but only some sound distortions. Noises in the sound information appear, but their share is relatively small and thus our hearing can detect the sounds of the words.

The proposed system allows the users to record information and then to hide it within JPEG image files.

For sound data, files with extention .au will be used. The program that runs such dimension files is Audacity recorder [5], that is included in the presented system.

Steganography of sound data within JPEG images. The developed system allows to:

- save information,
- hide information within JPEG images by using a LSB algorithm,
- restore information from JPEG image,
- process the information with the sound player.

The process of steganography and subsequent decoding is shown on Fig. 1.

This system has a convenient interface that allows to choose the image as a container, its size to map it both before and after steganography as well as to select the level (%) of compression.

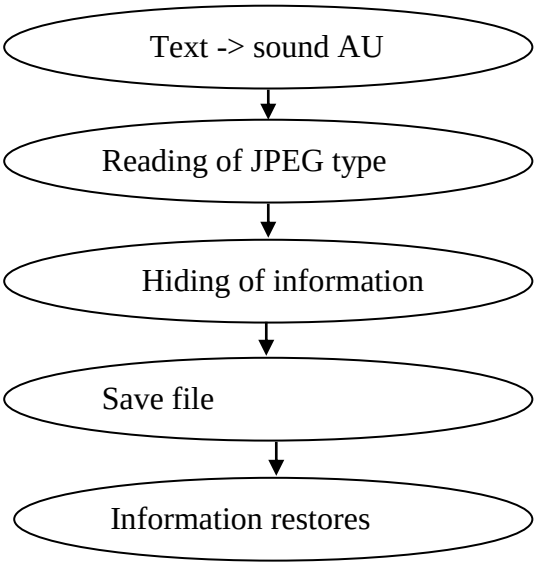

Fig. 1. Steganography and the process of data decoding

Comparative analysis of the results. In order to see if the losses resulting from compression of the image have affected the hidden information, let us consider a small portion of those data – two codes, one from the initial steganography and one - after decoding. In the table, the cases of image compression for 20% and 80% are presented.

Table

Original data for hiding	Image: data after 20% compression	Image: data after 80 % compression
10001100	10001110	10011111
01001100	01111101	01111101
11001100	11101111	11111111
00101100	01101111	01111111
10101100	10101101	10111101
01101100	01111101	11111101
11101100	11111111	11111111
00011100	01011110	01011111
10011100	10011101	10011101
00001100	00101111	01111111

It is seen that in the case of 20% compression, after decoding, the most important bits of the original data generally remain unchanged (compared with the first column), while the change results in drastic sound distortion.

The change of low-value bits results in only minor sound distortion, which does not significantly affect the adequate hearing.

For practical comparison, Figures 2a and 2b present examples of full and empty containers.

Fig. 2a. Full container

Fig. 2b. Empty Container

As it can be seen, the difference between the two figures is not obvious and thus does not cause suspicion.

Conclusion.The presented system with its user-friendly interface allows to carry out steganography of the sound information in any JPEG file. The limitation is that this program does not allow to work with very large JPEG files with the maximal possible size being 512x512. If the image compression value is high, the file suffers significant or total data losses.

References

1. **Peltier Thomas R.** Information Security: Policies and Procedures.- CRC press LLC, 2000.
2. **Stamp Mark.** Information Security principles and practice.- JohnWiley & Sons, 2006.
3. **Eric Cole.** Hiding in Plain Sight: Steganography and art of Covert Communication.- Wiley Publishing Inc., 2003.

4. http://www.guillermito2.net/stegano/jsteg/index.html
5. http://audacity.ru

Received on 26.08.2014.
Accepted for publication on 17.12.2014.

JPEG ՊԱՏԿԵՐՆԵՐՈՒՄ ՏԵՂԵԿՈՒԹՅԱՆ ԹԱՔՆԱԳՐՄԱՆ ՄԻ ԵՂԱՆԱԿ

Ռ.Գ. Հակոբյան

Թաքնագրության մեթոդները, երբ որպես կրիչ հանդես է գալիս JPEG ձևաչափի պատկերը, վատ են ուսումնասիրված և օգտագործվում են ոչ հաճախ : Պատճառն այն է, որ JPEG ձևաչափի պատկերը ենթարկվում է սեղմման և վերականգնելիս աղավաղվում է: Աղավաղումները չեն երևում անզեն աչքով, սակայն ազդում են թաքնված տեղեկույթի վրա, որի պատճառով դրա վերծանումը դառնում է անհնար: Առաջարկվում է մի եղանակ, երբ տեղեկույթը գրանցվում է ձայնային ձևաչափով, ապա թաքցվում պատկերում: Վերծանելիս ձայնային տեղեկույթը նույնպես խեղաթյուրված է լինում: Սակայն, ի տարբերություն տեքստայինից, որտեղ աղավաղումը անհնար է դարձնում բովանդակությունը հասկանալը, ձայնի աղավաղումը գրեթե չի խոչընդոտում բառերը ընկալելուն, և հնարավոր է լինում լիովին հասկանալ թաքնված տեղեկույթի իմաստը: Մշակված ծրագրային փաթեթի փորձարկումները ցույց տվեցին, որ 20% սեղմման դեպքում աղավաղումները հազիվ են զգացվում, իսկ 80% սեղմումը խեղաթյուրում է ամբողջ տեղեկույթը, և հնարավոր չի լինում հասկանալ այն: Մինչև 60% սեղմումը կարող է երաշխավորել թաքնված տեղեկույթի անվտանգությունը:

Առանցքային բառեր. թաքնագրություն, պատկեր, կրիչ, թաքնված տեղեկույթ:

МЕТОД СКРЫТИЯ ИНФОРМАЦИИ В JPEG ИЗОБРАЖЕНИЯХ

Р.Г. Акопян

Методы стеганографии, когда контейнером является изображение формата JPEG, плохо изучены и не часто используются. Причина в том, что изображение в JPEG подвергается сжатию и при восстановлении искажается. Эти искажения не видны невооруженным глазом, но влияют на скрытую информацию, и ее полноценное извлечение из контейнера невозможно. Предлагается метод, когда информация записывается в аудио-формате, а затем скрывается в изображении. При восстановлении звуковая информация также искажается, но в отличие от текста, где потери символов полностью искажают фразу, искажения звука не влияют на содержание речи и есть возможность в полной мере понять смысл скрытой информации. Разработанный пакет программ при испытаниях показал, что при 20% сжатии искажения не чувствуются на слух, при 80% - не позволяют понять текст. При сжатии до 60% можно гарантировать безопасность скрытой информации.

Ключевые слова: стеганография, изображение, контейнер, скрытая информация.