UDC 003.042

ATTACKS ON THE STEGANOGRAPHIC SYSTEM OF DIGITAL WATERMARKS AND PROTECTION AGAINST THEM

I. ISAKAU, Y. PASTUKHOV
Polotsk State University, Belarus

In this paper, we examine a list of possible attacks on a stegosystem of digital watermarks, and ways to counter these attacks.

Classification of attacks on the DW. DW must satisfy the conflicting requirements of visual (audio) stealth and robustness to the basic signal processing operations. In the future, without loss of generality, we will assume that the image is used as a container.

We turn to the message embedding system by modifying the least significant bit (LSB) of the pixels. Almost any method of image processing can lead to the destruction of a significant part of the embedded message. For example, consider the operation of calculating a moving average over two adjacent pixels $\frac{(a+b)}{2}$, which is the simplest example of low-pass filtering. Let the values of the pixels a and b be even or odd with a probability $p=\frac{1}{2}$. Then the value of the least significant bit will change after averaging in half the cases. A change in the quantization scale, say, from 8 to 7 bits, can also lead to the same effect. A similar effect is exerted by lossy image compression. Moreover, the use of noise purification methods using noise estimation and subtraction will distort the vast majority of bits of a hidden message. [1] There are also much more harmful image processing operations for the DW, for example, scaling, rotation, truncation, pixel permutation . The situation is aggravated by the fact that the conversion of stego messages can be carried out not only by the violator, but also by the legitimate user, or be a consequence of errors during transmission over the communication channel.

A shift of several pixels may result in non-detection of the DW in the detector. Analog video recorders, as a rule, somewhat shift the image due to the uneven rotation of the tape drive engine or the wear of the tape. The shift can be invisible to the eye, but lead to the destruction of the DW.

Now consider the attacks specific to DW systems. The following categories of attacks against such stegosystems can be distinguished [3].

1. Attacks against the built-in message - aimed at removing or damaging the DW by manipulating the stego. Attack methods that fall into this category do not try to evaluate and highlight a watermark. Examples of such attacks include linear filtering, image compression, adding noise, aligning the histogram, changing the contrast, etc.

2. Attacks against the stode detector - aimed at hindering or making impossible the correct operation of the detector. In this case, the watermark in the image remains, but the possibility of its reception is lost. This category includes attacks such as affine transformations (i.e. scaling, shifts, rotations), image truncation, pixel permutation, etc.

3. Attacks against the protocol for the use of DW - mainly associated with the creation of false DW, false stego, inversion of DW, the addition of several DW.

4. Attacks against the DW itself - aimed at evaluating and extracting the DW from the stego message, if possible without distorting the container. This group includes such attacks as collusion attacks, statistical averaging, methods for cleaning signals from noise, some types of nonlinear filtering [4] and others.

It should be noted that the classification of attacks under consideration is not the only possible and complete one. In addition, some attacks (such as noise removal) can be categorized into several categories.

In accordance with this classification, all attacks on DW systems can be divided into four groups:

1) attacks aimed at removing the DW;

2) geometric attacks aimed at distorting the container;

3) cryptographic attacks;

4) attacks against the protocol used to embed and verify the DW.

Methods of counteracting attacks on DW systems. In the simplest DW stegosystems, a pseudo-random sequence is used when embedding, which is a realization of white Gaussian noise and does not take into account the properties of the container. Such systems are practically unstable to most of the attacks discussed above. To increase the robustness of stegosystems, a number of improvements can be proposed [2].

In a robust stegosystem, the correct choice of pseudo-random sequence parameters is necessary. It is known that in this case, systems with spreading the spectrum can be very robust with respect to attacks such as adding noise, compression, etc. It is believed that the DW should be detected with a sufficiently strong low-pass filtering (7x7 filter with a rectangular characteristic). Therefore, the signal base must be large, which reduces the bandwidth of the stego channel. In addition, the memory bandwidth used as a key must be cryptographically secure.

The reason for the instability of the DW systems with the expansion of the spectrum to such attacks is due to the fact that the sequence used for embedding usually has a zero mean. After averaging over a sufficiently large number of implementations, the DW is deleted. A special method of constructing a watermark is known that is directed against such an attack. In this case, the codes are developed in such a way that with any averaging there always remains a non-zero part of the sequence (static component). Moreover, it is possible to restore the rest of the sequence (dynamic component). The disadvantage of the proposed codes is that their length increases exponentially with the increase in the number of distributed protected copies. A possible way out of this situation is the use of hierarchical coding, that is, the assignment of codes for a group of users. Some of the analogies here are with Code Division Multiple Cellular (CDMA) cellular systems.

Various countermeasures have been proposed to solve the problem of property rights. The first way is to build an irreversible DW algorithm. The DW must be adaptive to the signal and be embedded using a unidirectional function, for example, a hash function. The hash function converts 1000 bits of the original image V into the bit sequence b_i, $i = 1 \dots 1000$. Further, depending on the value of b_i two embedding functions of the DW are used. If $b_i = 0$, then the function $v_i(1 + aw_i)$ is used, if $b_i = 1$, then the function $v_i(1 + aw_i)$, where v_i - is the i-th image coefficient,, w_i - is the i-th bit of the embedded message. It is assumed that such an algorithm for the formation of the DW will prevent falsification.

The second way to solve the problem of property rights is to embed in the DW a certain time stamp provided by a third, trusted party. In the event of a conflict, a person with an earlier time stamp on the image is considered the real owner.

One of the principles of building a robust DW is to adapt its spectrum. A number of studies have shown that the envelope of the spectrum of an ideal DW should repeat the envelope of the spectrum of the container. The spectral power density of the DW, of course, is much less. With such a spectral envelope, the Wiener filter gives the worst estimate of the DW possible: the variance of the error values reaches the variance of the filled container values. In practice, adaptation of the DW spectrum is possible by local assessment of the container spectrum. On the other hand, the methods of embedding DW in the field of transformation achieve this goal through adaptation in the field of transformations. [4]

To protect against attacks such as affine transformation, you can use an additional (reference) DW. This DW does not contain information, but is used to "register" the transformations performed by the violator. There is a predistortion circuit in the DW detector that performs the inverse transform. There is an analogy with the test sequences used in communication. However, in this case, the attack can be directed precisely against the supporting DW. Another alternative is to embed the DW in visually significant areas of the image that cannot be removed from it without significant degradation. Finally, we can place the stego in the conversion-invariant coefficients. For example, the amplitude of the Fourier transform is invariant to image shift (in this case, only the phase changes).

Another method of protection against such attacks is a block detector. The modified image is divided into blocks of 12x12 or 16x16 pixels, and all possible distortions are analyzed for each block. That is, the pixels in the block undergo rotations, permutations, etc. For each change, the correlation coefficient of the DW is determined. The transformation, after which the correlation coefficient turned out to be the largest, is considered to be actually performed by the violator. Thus, it becomes possible to reverse the distortions introduced by the intruder. The possibility of this approach is based on the assumption that the intruder will not significantly distort the container (this is not in his interests).

Conclusion. Summing up, we can say with confidence that you can protect yourself from every attack by knowing all the details of this attack, and you should always close the vulnerabilities of your system and conduct attacks against it by testing for resistance to them. Carrying out attacks on our own CEH, we analyze its vulnerability to this attack, which in the end we can close. It is impossible to fully ensure the security of our CEH, because while we close one vulnerability, the attacker is looking for a new one, and this race between the attacker and the computer security specialist will be eternal. The surest way to defend this is an "attack", if we ourselves can identify vulnerabilities, then we can successfully close them, this remains the main aspect of the work of KB experts - the study and practice of attacks, and the development of defense systems.

REFERENCES

1. Sadov, V.S. Computer steganography (Lecture notes) / V.S. Sadov - Minsk: BSU, 2010 .-- 211 p.
2. Schneier, B. Secrets and lies. Data Security in the Digital World / B. Schneier - St. Petersburg: Peter, 2003. - 368 p.
3. Gribunin, V.G. Digital steganography / V.G. Gribinin - Moscow: "Academy", 2009. - 265 p.
4. Shelukhin, O.I. Steganography. Algorithms and software implementation / O.I. Shelukhin / S.D. Kanaev - Moscow: "Hot Line - Telecom", 2017. - 592 p.