Journal of Engineering and Applied Sciences 13 (18): 7607-7611, 2018
ISSN: 1816-949X

Discussion of the Basic Types of Attacks on Stegosystems Digital Watermarking

[1]Sameer Saud Dakhel, [2]Sadiq Sahip Majeed and [1]Rasha Muhseen Hadi
[1]College of Agriculture,
[2]College of Science, Al-Muthana University, Baghdad, Iraq

Abstract: Digital watermarking is a style through which we can achieve the authentication for images,videos and texts. The functions of watermarking are not only for authentication purpose but we have another major issue is a security of the documents this can also be achieved by watermarking, basically, we have secondary image that covers the original image to make it secure. Today there are many people need a safe and protected method to expand information. To achieve the authentication of data as well as the data security, we presents a comparative study on the different types of attacks for digital watermarking method.

Key words: Digital watermark system, digital watermarking, stegosystem, authentication, watermarking method, comparative study

INTRODUCTION

The process of imbedding data within digital media is known as digital watermarking process, it used to hide the data and extracted it during a digital media, the watermark is a piece of information that we can spread it which can't be known by others, the benefit of water marking teqnique is ownership in evidence such as (copyright and ip preservation), hide the data, authentication and coping preclude. The study of watermark attacks has affirmative influence that is supports people doing a study about digital watermarking techniques to expansion anew method that are able to resist attacks. In watermarking, the penetration and strength of attacks requires accurate effort to make watermarked data forge resistant.

Watermark concept

Watermark: Adigital watermark is a message that embedded within adigital media like-image, vedio, audio, text and holds information to the authenticated user, copyright owner. There are various categories of the watermark.

Visible watermark: Visible watermarks are created to be easily recognized by the spectator and determine the owner clearly. This pattern must not reduce from the content of image.

Invisible watermark: This it is designed to be very precise and difficult for a user to identify an added lable if he is not familiar with its format.

CLASSIFICATION OF ATTACKS ON STEGOSYSTEMS DIGITAL WATERMARKING

There is a different classification of attacks on the stegosystem. Now consider the attacks specific to digital watermark systems. It is possible to single out the following categories of attacks against such stegosystems.

Attacks against the built-in message-aimed at removing or corrupting the digital watermark by manipulating stego. The methods of attack included in this category do not attempt to evaluate and isolate the watermark. Examples of such attacks may be linear filtering, image compression, noise addition, histogram alignment, contrast change and so on.

Attacks against the stegodetector-are aimed at making it difficult or impossible to properly operate the detector. In this case, the watermark in the image remains but the possibility of its reception is lost. This category includes such attacks as affine transformations (i.e., scaling, shifts, rotations), truncating an image, permuting pixels and so on.

Attacks against the use of the protocol digital watermark-mainly associated with the creation of false digital watermark, false stego, inversion digital watermark, adding multiple digital watermark.

Attacks against the digital watermark itsel-aimed at evaluating and extracting the digital watermark from the steg message, if possible without distortion container. This group includes attacks such as collusion attacks, statistical averaging, methods of clearing signals from noise, some types of nonlinear filtering (Langelaar *et al*., 1998) and others.

Corresponding Author: Sameer Saud Dakhel, College of Agriculture, Al-Muthana University, Baghdad, Iraq

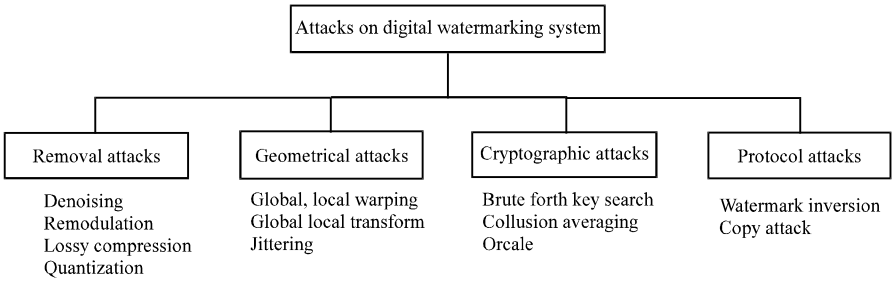

Fig. 1: Different watermarking system attacks

It should be noted that, the classification of attacks under consideration is not the only possible and complete. In addition, some attacks (for example, noise removal) can be categorized into several categories. In (Kutter *et al*., 2000) another classification of attacks was proposed which also has its advantages and disadvantages.

Can be categorized into several categories. By Kutter *et al*. (2000) another classification of attacks was proposed which also has its advantages and disadvantages.

In accordance with this classification, all attacks on the embedding systems of the digital watermark can be divided into four groups (Fig. 1).

Attacks aimed at removing the digital watermark. Geometric attacks aimed at distorting the container. cryptographic attacks. Attacks against the protocol used for embedding and checking the digital watermark.

Attacks aimed at removing the digital watermark: This group includes attacks such as cleaning of container signals from noise, overmodulation, averaging and collisions. These attacks are based on the assumption that the digital watermark is a statistically described noise. Purification from noise consists in filtering the signal using the maximum likelihood criteria or the maximum a posteriori probability. As a filter that realizes the maximum likelihood criterion, a median (for a digital watermark having a Laplace distribution) or averaging (for a Gaussian distribution) filter that is applied in the StirMark Software package can be used. By the criterion of maximum a posteriori probability, the adaptive Wiener filter will be the best (in the case when a nonstationary Gaussian process is used as the container model) as well as threshold methods of noise cleaning (soft and hard thresholds) (the model is a generalized Gaussian process) that have much in common with lossy compression methods.

Lossy compression and noise cleaning significantly reduce the throughput of the stego channel, especially, for smooth areas of the image whose conversion factors can be "Zeroed" without noticeably reducing the quality of the restored image.

Overmodulation is a relatively new method which is specific for the attacks on the digital watermark. The overmodulation attack was first proposed by Kutter *et al*. (2000). Currently, its various variants are known, depending on the decoder used in the stegosystem. In the construction of the attack, there are nuances for the stegosystem of M-ary modulation, a stegosystem using noise-immune codes using a correlation decoder. In either case, it is considered that the digital watermark is embedded in the image using broadband signals and is multiplied by the entire image. Since, the digital watermark evaluated by the decoder is correlated with the true one, it becomes possible to deceive the decoder. The attack is constructed as follows. First, the digital watermark is "predicted" by subtracting the filtered version of the image from the protected image (a median filter is used). "Predicted" The digital watermark is exposed to HF filtering, truncated, multiplied by two and subtracted from the original image. In addition, if it is known that during the implementation of the digital watermark multiplied by some mask to increase the invisibility of embedding, then the attacker evaluates this mask and multiplies the digital watermark on it. As an additional measure to "deceive" the decoder, it seems effective to incorporate images into the high-frequency regions (where distortions are imperceptible) of patterns having a non-Gaussian distribution. Thus, the optimality of the linear correlation detector will be violated.

Such an attack will be effective only against the high-frequency digital watermark, so, the real digital watermark is constructed, so that, their spectrum matches the spectrum of the original image. The fact is that a reliable estimate is obtained only for the high-frequency components of the digital watermark. After subtracting it, the low-frequency component of the digital watermark remains unchanged and gives the detector a positive correlation response. The high-frequency component will

give a negative response which in the sum will give zero and the digital watermark will not be detected. As another counteraction to this attack, it was suggested to perform preliminary low-frequency filtering.

By Su and Girod (1999), a modification of this algorithm is presented which consists in applying the Wiener filter instead of the median and more intelligent way of finding the multiplication factor. It is selected, so as to minimize the cross-correlation coefficient between the digital watermark and stego. In addition, one more step is added: the imposition of random noise. This attack does not work against the adaptively built-in digital watermark because it assumes that the digital watermark and stego are a stationary Gaussian process with zero mean. It is clear that this assumption is not satisfied also for real images. Therefore, Voloshinovsky and others proposed an attack in which the signals are modeled as a nonstationary Gaussian or generalized stationary Gaussian process (Voloshynovskiy *et al*., 1999). Multiplication factor a digital watermark is selected based on the local properties of the image. Instead of imposing random noise, it is suggested to add counts with a sign opposite to the reference sign digital watermark (assuming that the digital watermark is a sequence of bipolar symbols). This makes the operation of the correlation detector even more difficult. Of course, the signs need not be changed at all but only for a part of the counts of the estimated digital watermark, for example, accidentally.

Other attacks of this group include averaging attack and collusion attack. If you have a large number of copies of stego with different digital watermark or with different embedding keys, you can perform their averaging. For example, video frames can have different Digital watermarks. If the digital watermark had zero mean, then after averaging it will be absent in the image.

Geometric attacks: Unlike the deletion attacks, geometric attacks tend not to remove the digital watermark but to change it by introducing spatial or temporal distortion. Geometric attacks are mathematically modeled as affine transformations with an unknown decoder parameter. There are six affine transformations: scaling, changing proportions, rotations, shifting and truncation. These attacks result in the loss of synchronization in the digital watermark detector and can be local or global (that is, applied to the entire signal). It is possible to cut out individual pixels or rows, rearrange them in places, apply some transformations, etc.

There are more "intelligent" attacks on the applied synchronization method digital watermark. The main idea of these attacks is to recognize the synchronization method and destroy it by smoothing out the peaks in the amplitude spectrum of the digital watermark. Attacks are effective under the assumption that periodic templates are used as a synchronization mechanism. In this case, two approaches can be used to ensure synchronization: the integration of peaks in the spectral region or the periodic introduction of the digital watermark sequence. In both cases, peaks form in the spectrum which are destroyed in the attack under consideration. After the destruction, you can apply other geometric attacks: no synchronization.

Modern methods of embedding the digital watermark are robust to global attacks. They use special methods of synchronization recovery which have much in common with those used in communication technology. Robustness is achieved through the use of shift-invariant domains (Lin, 2000) the use of the reference digital watermark (Wu, 2001) the calculation of the autocorrelation function digital watermark.

If the provision of robustness to global geometric attacks is a more or less solved task, then ensuring the resistance to local image changes is an open question. These attacks are based on the fact that the human eye is not sensitive to small local changes in the picture.

Cryptographic attacks: Cryptographic attacks are named, so because they have analogs in cryptography. These include attacks using the oracle as well as hacking with the help of "brute force".

Attack using the oracle allows you to create an unprotected digital watermark image if you have an intruder detector. In research by Hartung *et al*. (1999), the stability of the digital watermark is studied on the basis of the spreading of the spectrum to the attack in the presence of a detector in the form of a "black box". The method consists in an experimental study of the behavior of the detector to determine which images it responds to, to which it does not. For example, if the detector makes "soft" decisions that is shows the probability of having a stego in the signal, then the attacker can figure out how small changes in the image affect the behavior of the detector. Modifying an image of a pixel by a pixel, it can generally figure out which algorithm the detector uses. In the case of a detector with a "hard" solution, an attack is carried out near the border where the detector changes its solution from "present" to "absent".

An example of an attack on a detector with a hard decision: based on the existing image containing the stego message, a test image is created. The test image can be created in different ways, modifying the original image until the detector shows the absence of the digital watermark. For example, you can gradually reduce the contrast of an image or a pixel by a pixel, to replace the actual values with some other ones.

The attacker increases or decreases the value of any pixel until the detector detects the digital watermark again. Thus, it is found whether the value of this pixel is increased or decreased by the digital watermark.

Step 2 is repeated for each pixel in the image. Knowing how sensitive the detector is to the modification of each pixel, the attacker determines the pixels, the modification of which will not lead to a significant deterioration of the image but will disrupt the operation of the detector.

These pixels are subtracted from the original image. Is it possible to build a stegoalgorithm, resistant to such an attack, is still unknown.

Attacks against the protocol used: By Craver *et al.* (1996, 1998) many stegosystems of digital watermarks are sensitive to the so-called inverse attack. This attack is as follows: the intruder states that in a protected image a piece of data is his watermark. After that, he creates ambiguity, subtracting this piece of data. In the ambiguity, there is a real digital watermark. On the other hand, in the protected image there is a false digital watermark proclaimed by the intruder. There is an insoluble situation. Of course, if the detector has an original image, then the owner can be identified. But as shown by Craver *et al.* (1997) not always. The methods of protection against such an attack are presented by Craver *et al.* (1996, 1998). They show that, a watermark resistant to such an attack must be irreversible. For this, it is made dependent on the image using a unidirectional function.

METHODS FOR COUNTERING ATTACKS ON SYSTEMS DIGITAL WATERMARK

In the simplest stegosystems, the digital watermarks is embedded with a pseudo-random sequence 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 described above. To improve the robustness of stegosystems, a number of improvements can be proposed.

In a robust stegosystem, the correct choice of pseudo-random sequence parameters is necessary. It is known that in this case systems with spreading can be very robust with respect to attacks such as the addition of noise, compression, etc. It is considered that, the digital watermark should be detected with a sufficiently strong low-frequency filtering (7×7 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 attack of "collusion" and possible methods of defense against it was considered by Deguillaume *et al.* (2000). The reason for the instability of systems the digital watermark with spreading to such attacks is explained by the fact that the sequence used for nesting usually has zero mean. After averaging over a fairly large number of implementations, the digital watermark is deleted. A special method for constructing a watermark is known, directed against such an attack. In this case, the codes are designed in such a way that for any averaging, the part of the sequence that is not equal to zero (the static component) always remains. 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 growth of the number of distributed protected copies. A possible way out of this situation is the use of hierarchical coding, that is assigning codes to a group of users. Some analogies here are available with the systems of cellular communication with code division of users (CDMA).

Various methods of counteraction were proposed to address the problem of property rights. The first way is to build an irreversible algorithm for the digital watermark. The digital watermark must be adaptive to the signal and be integrated with a one-way function, for example, a hash function (Schneier, 1996). The hash function converts the 1000 bits of the original image V into a bit sequence $b_i i=\overline{1,\ldots 1,000}$. Further, depending on the value b_i the two functions of embedding the digital watermark are used. If $b_i = 0$ then used the function $v_i(1+aw_i)$. If $b_i = 1$ then used the function $v_i(1-aw_i)$ when v_i, i-aspect ratio, w_i, i-bit inline message. It is assumed that such an algorithm for the formation of the digital watermark will prevent falsification. By Craver *et al.* (1996, 1997, 1998) the example showed that in order for this algorithm to be irreversible, all the elements of must be positive.

The second way to solve the problem of property rights is to integrate into the digital watermark some time mark provided by a third, trusted party. In the event of a conflict, a person who has an earlier time stamp on the image is considered to be the real owner.

One of the principles of constructing a robust digital watermark is to adapt its spectrum. In a number of works it was shown that, the envelope of the ideal spectrum. The digital watermark must repeat the envelope of the spectrum of the container. Spectral power density The digital watermark, of course, is much smaller. With such a spectral envelope, the Wiener filter gives the worst estimate of the digital watermark from the possible: the variance of the error values reaches the variance of the values of the filled container. In practice, spectrum adaptation a digital watermark is possible by locally

evaluating the spectrum of the container. On the other hand, the methods of embedding the digital watermark in the transformation area achieve this goal by adapting in the transformant domain.

To protect against affine transformation type attacks, an additional (reference) digital watermark can be used. This digital watermark does not carry information but it is used to "register" the transformations performed by the intruder. In the digital watermark detector, there is a predistortion scheme that performs the inverse transformation. Here, there is an analogy with the test sequences used in the connection. However, in this case, the attack can be directed against the reference digital watermark. Another alternative is to embed the digital watermark in visually meaningful areas of the image that can not be removed from it without significant degradation. Finally, we can place the stego in coefficients that are invariant to the transformation. For example, the Fourier transform amplitude is invariant to the image shift (only the phase changes here).

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

CONCLUSION

We have represented the essential information of digital watermarking technique and different attacks on it. Firstly, we viewed into watermarking and presented general watermarking system with watermarks and watermark protection also represented. We then studied various attacks of watermark includes removal attack, geometric attacks, protocol attack , etc. So, it's important for the beginner can start his work on it.

REFERENCES

Craver, S., N. Memon and B. Yeo, 1998. Resolving rightful ownerships with invisible watermarking techniques: Limitation, attacks and implementation. IEEE J. Selected Area Commun., 16: 573-586.

Craver, S., N. Memon, B. Yeo and M. Yeung, 1996. Can invisible watermarks resolve rightful ownerships?. Master Thesis, IBM Research, Africa.

Craver, S., N. Memon, B.L. Yeo and M.M. Yeung, 1997. On the invertibility of invisible watermarking techniques. Proceedings of the International Conference on Image Processing Vol. 1, October 26-29, 1997, IEEE, Santa Barbara, California, USA., pp: 540-543.

Deguillaume, F., G. Csurka and T. Pun, 2000. Countermeasures for unintentional and intentional video watermarking attacks. Proceedings of the 2000 Conference on Security and Watermarking of Multimedia Contents II Vol. 3971, May 9, 2000, International Society for Optics and Photonics, Bellingham, Washington, USA., pp: 346-358.

Hartung, F., J. Su and B. Girod, 1999. Spread spectrum watermarking: Malicious attacks and counterattacks. Signal Process., 3657: 147-158.

Kutter, M., S.V. Voloshynovskiy and A. Herrigel, 2000. Watermark copy attack. Proceedings of the 2000 Conference on Security and Watermarking of Multimedia Contents II Vol. 3971, May 9, 2000, International Society for Optics and Photonics, Bellingham, Washington, USA., pp: 346-358.

Langelaar, G.C., R.L. Lagendijk and J. Biemond, 1998. Removing spatial spread spectrum watermarks by non-linear filtering. Proceedings of the 9th European Conference on Signal Processing (EUSIPCO'98), September 8-11, 1998, IEEE, Rhodes, Greece, ISBN:978-960-7620-06-4, pp: 1-4.

Lin, C.Y., 2000. Watermarking and digital signature techniques for multimedia authentication and copyright protection. Ph.D Thesis, Columbia University, New York, USA.

Schneier B., 1996. Applied Cryptography: Protocols, Algorithms and Source Code in C. 2nd Edn., John Wiley & Sons, Hoboken, New Jersey, USA., ISBN:9780471128458, Pages: 758.

Su, J.K. and B. Girod, 1999. On the robustness and imperceptibility of digital fingerprints. Proceedings of the International Conference on Multimedia Computing and Systems Vol. 2, June 7-11, 1999, IEEE, Florence, Italy, pp: 530-535.

Voloshynovskiy, S., A. Herrigel, N. Baumgaertner and T. Pun, 1999. A stochastic approach to content adaptive digital image watermarking. Proceedings of the 3rd International Workshop on Information Hiding, September 29-October 1, 1999, Springer, Dresden, Germany, ISBN:978-3-540-67182-4, pp: 211-236.

Wu, M., 2001. Multimedia data hiding. Ph.D Thesis, Princeton University, Princeton, New Jersey, USA.