DIGITAL WATERMARKING FOR TEXT/IMAGES AND ITS SECURITY INSIGHTS: AN STEGANOGRAPHIC APPROACH

Yashu Pradhan
Asstt. Proff. Dept. Of B.Sc. IT
Science College, Kokrajhar.

Dr. Masih Saikia
Head, Dept. of Computer Science
Pragjyotish College, Guwahati

Abstract

A study of digital watermarking for text / images and its security insights (an steganographic approach) is studied. There are several techniques for information hiding into digital media. They are used for several purposes as well as copyright protection. *Digital Watermarking Technology* is an emerging field in computer science, cryptography, signal processing and communications. A digital watermarking is perceptually invisible to prevent obstruction of the original image. There are numerous watermarks can be produced for use purpose. With the help of the watermark we can able to determine the true owner of the image. The concept of watermarking security is somewhat diffuse and it is still a matter of discussion, it is commonly accepted that if the secret parameters of the embedding/decoding function of a certain data hiding scheme can be estimated, then the scheme cannot be claim to be secure. The current watermarking methods intend to endure from these attacks by the use of templates, invariant domains, image feature dependent methods or self synchronizing watermarks to defeat the geometrical transformations imposed by the attacker.

Introduction :

There are several techniques for information hiding into digital media. They are used for several purposes as well as copyright protection. Two basic methods of information hiding are ***cryptography*** and ***steganography***. The concept of *digital watermarking* is derived from steganography. The term steganography means "cover writing" and cryptography means "secret writing" [7].

Digital Watermarking Technology is an emerging field in computer science, cryptography, signal processing and communications. Digital Watermarking is intended by its developers as the solution to the need to provide value added protection on top of data encryption and scrambling for content protection [9].

The use of steganography for undetected communication, dissidents could create a Web site bursting with politically-correct pictures, such as photographs of the great leader, local sports, movie, and television stars, etc. Of course, the pictures would be riddled with steganographic messages. If the message were first compressed and then encrypted, even someone who suspects their presence would have immense difficulty in distinguishing the message from white noise. Of course, the message should be fresh scans; copying a picture from the Internet and changing some of the bits is a dead giveaway.

Images are no means the only carrier for steganographic messages. *Audio* files also work fine. *Video* files have a huge steganographic bandwidth. Even the layout and ordering of tags in an *HTML* file can carry information.

Although, we have first examined steganography in the context of text. But, it has also many other uses. One common use is for the owners of images to encode secret messages in them stating their ownership rights. If such an image is stolen and placed on a Web site, the lawful owner can reveal the steganographic message in court to prove whose image it is. This technique is called ***watermarking*** [12].

In general, a *digital watermark* is a code that is embedded inside an image. It acts as a *digital signature*, giving the image a sense of ownership or authenticity. Over the past few years, there has been tremendous growth in computer networks and more specifically, the World Wide Web. This phenomenon, coupled with the exponential increase of computer performance, has facilitated the distribution of multimedia data such as images. Publishers, artists, and photographers, however, may be unwilling to distribute pictures over the Internet due to a lack of security; images can be easily duplicated and distributed without the owner's consent. Digital watermarks have been proposed as a way to tackle this tough issue. This digital signature could discourage copyright violation, and may help determine the authenticity and ownership of an image [5].

The Properties of Digital Watermark :

Ideal properties of a digital watermark have been stated in many articles and papers. These properties include the following [5] :

(1) A digital watermark should be perceptually invisible to prevent obstruction of the original image.

(2) A digital watermark should be statistically invisible so it cannot be detected or erased.

(3) Watermark extraction should be fairly simple. Otherwise, the detection process requires too much time or computation.

(4) Watermark detection should be accurate *false positives*, i.e., the detection of a nonmarked image, and *false negatives*, i.e., the non-detection of a marked image, should be few.

(5) Numerous watermarks can be produced. Otherwise, only a limited number of images may be marked.

(6) Watermarks should be robust to filtering, additive noise, compression, and other forms of image manipulation.

(7) The watermark should be able to determine the true owner of the image.

Digital Image Watermarking Framework [7] :

Image Watermarking, as mentioned earlier, is the process of embedding a secondary signal into an image such that the signal can be detected or extracted later to make an assertion about the image. In general, any watermarking scheme consists of the following three parts:

(1) The watermark signal,

(2) Watermark embedder that embeds the watermark into the media

(3) Watermark detector that verifies the presence of watermark

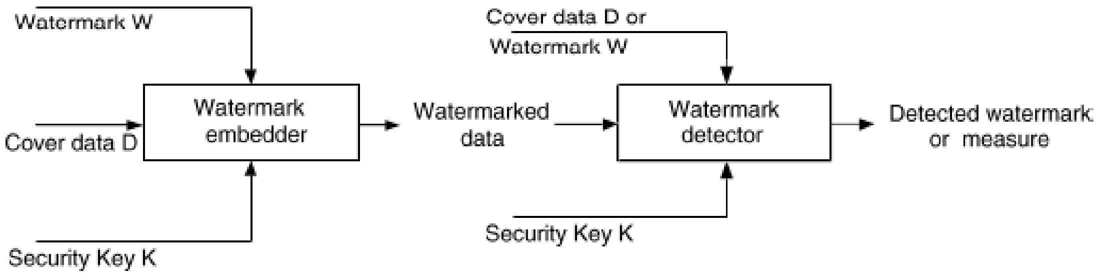

Figure 1: A typical watermarking system

Above Figure 1 is a conventional watermarking system [13] consists of watermark embedder and watermark detector. The inputs to the watermark embedder are the watermark, the cover media data and the embedding security key. The watermark can be a number sequence, a binary bit sequence or may be an image. The key is used to enhance the security of the whole system. The output of the watermark embedder is the watermarked data. The inputs to the watermark detector are the watermarked data, the security key and, depending on the method, the original data and/or the original watermark. According to Cox et al. [1], a watermark detector includes two-step process.

The first step is to extract watermark that requires one or more pre-processes to extract a vector referred to as extracted mark. In this process, original unwatermarked image may be used or may not be used depending on the algorithm. If the detector does not require the original copy, watermarking scheme is called *public watermarking* or *blind watermarking*, if the detector requires the original image, then, it is called *private watermarking* or *non-blind watermarking* [3]. If the original image is used, the watermark can be extracted in its exact form (if the image is not corrupted). If it is a blind detection, we can determine whether a specific given watermarking signal is present in an image.

Then, the second step is to determine whether the extracted mark contains the original watermark or not. The second step usually involves with comparing the extracted mark with the original watermark and the result could be some kind of confidence measurement representing the possibility of the original watermark being present in the document. Correlation method is used for this purpose. The correlation function computes the correlation value and the computed correlation are compared with a detection threshold. If the correlation value exceeds the threshold value, the image is believed to be watermarked. For some watermarking algorithms, the extracted mark can be further decoded to get the embedded message for various purposes such as copyright protection.

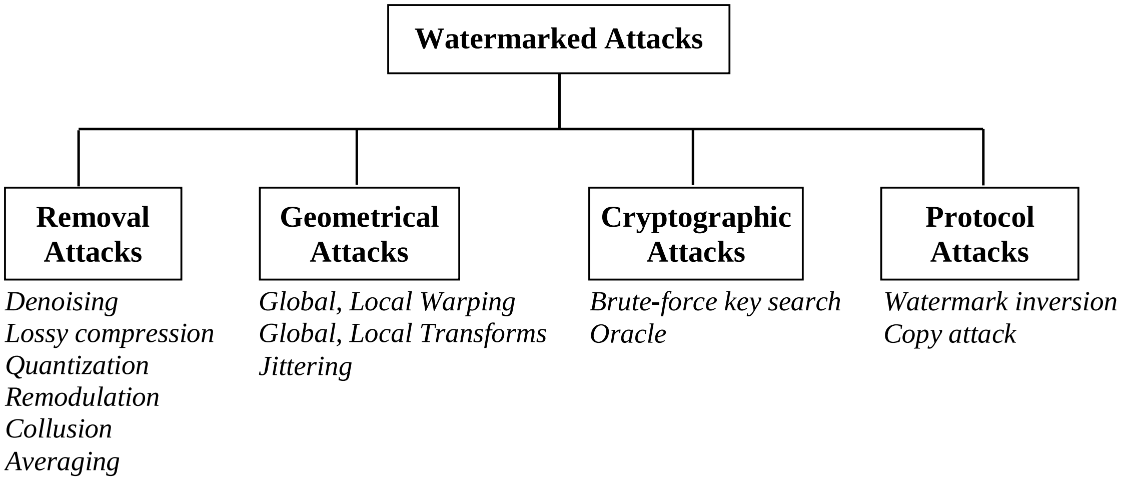

Figure 2: Different attacks on watermark [6]

Attacks on Watermarks [7] :

According to the watermarking jargon (terminology), an *attack* is any processing that may mess up detection of the watermark or communication of the information provided by the watermark. The processed, watermarked data is then called *attacked data*. Robustness against attacks is an important issue for watermarking schemes. The usefulness of an attacked data can be measured by its perceptual quality and the amount of watermark destruction can be measured by criteria such as miss probability, probability of bit error, or channel capacity. An attack may succeed. The wide class of existing attacks can be divided into four main groups: removal attacks, geometrical attacks, cryptographic attacks and protocol attacks [6].

The Figure 2 (as shown on previous page) summarizes the different types of attacks.

(1) Removal attacks

These are the attacks that try to weaken or completely remove a watermark from its associated content, still preserving the content so that it is not useless after the attack is over. This category includes denoising, quantization, remodulation, and collusion attacks. *Denoising* and *quantization* attacks damage the watermark quality as much as possible, while keeping the quality of the attacked data high enough. Lossy compression has the same effect as denoising. The *remodulation* attack intends to predict the watermark. It may be implemented by subtracting the median filtered version of the watermarked image from the watermarked image itself. Then the predicted watermark is removed from the watermarked image, producing the median filtered version of watermarked data. *Collusion* attacks are possible when many copies of a given data set, each signed with a different watermark, are available to an attacker. In this case, a successful attack can be performed by averaging all copies or taking only small parts from each different copy.

(2) Geometric attacks

Geometric attacks consist of the distortions particular to videos and images including operations as rotation, scaling, translation, cropping etc. In contrast to removal attacks, geometric attacks do not actually remove the embedded watermark, but attempt to deform the watermark detector synchronization with the embedded information. The embedded watermark information can be recovered if the perfect synchronization is regained. However, the complexity of the required synchronization process might be too huge to be realistic.

Current watermarking methods intend to endure from these attacks by the use of templates, invariant domains, image feature dependent methods or self synchronizing watermarks to defeat the geometrical transformations imposed by the attacker [10].

(3) Cryptographic attacks

Cryptographic attacks intend to break the security methods in watermarking schemes and thus finding a way to remove the embedded watermark information or to embed deceptive watermarks. *Brute-force search* for the embedded secret information is one such technique. Another attack in this category is the so-called *Oracle attack*, which can be used to generate a non-watermarked signal when a watermark detector device is available. High computational complexity has restricted attackers from applying these attacks on watermarks.

(4) Protocol attacks

Craver *et al*. [2] mentioned an attack, called the *watermark inversion attack* or *IBM attack*, which produces a fake watermarking schemes that can be applied on a watermarked image to create doubt about which watermark was inserted first.

Copy attack is another kind of protocol attack. In this case, the watermark is predicted by using a watermarked data, and this predicted watermark is embedded into another data by adapting the local features to satisfy its imperceptibility.

Watermarking Security [4] :

Although the concept of watermarking security is somewhat diffuse and it is still a matter of discussion, it is commonly accepted that if the secret parameters of the embedding/decoding function of a certain data hiding scheme can be estimated, then the scheme cannot be claimed to be secure. In the remainder of this thesis, we will work with the following definition in mind.

Definition **:** Attacks to security are those aimed at obtaining information from the secret parameters of the embedding and/or decoding functions. From this definition, it follows that a data hiding scheme is secure if it properly conceals the secret parameters. Although this definition of security may appear too restrictive, it poses two clear advantages over other definitions:

(1) It establishes a clear frontier between robustness and security;

(2) It allows to model precisely the problem of watermarking security.

Conclusion and measures

Tools for Measuring Security [4] :

From the definition of security given in above definition, it follows that the security of a watermarking system is directly related to the difficulty in estimating the secret parameters of the embedding function from the observations at hand. Thus, a natural question is that how can we quantify the hardness of such estimation problem. Let us first recall the classical criteria for evaluating the security of cryptosystems [3]:

(1) **Computational security :** This measure is concerned with the computational effort needed to break a given cryptosystem. If the best known algorithm for breaking the system demands a high number of computational resources, the system is said to be computationally secure. Clearly, this measure is not very useful as it is impossible to

prove that a certain cryptosystem is secure (being secure against a certain attack does not imply being secure against a different class of attacks).

(2) **Provable security :** This measure consists in reducing the proof of security to another problem which is well understood and known to be difficult, such as the factorization of integers in prime numbers or many other combinatorial problems, which are known to be *NP-complete*. This definition of security is more useful than the former because it is possible to quantify the minimum effort needed to break the system. However, the drawback of this approach is that it is impossible to prove the security of a given cryptosystem in absolute terms, since the security proof is always relative to some other problem.

(3) **Unconditional security :** A cryptosystem is said to be "unconditionally secure" if it cannot be broken regardless the computational resources employed by the attacker.

The above criteria can be readily translated to the data hiding scenario:

- A *data hiding system* would be computationally secure if the computational complexity of the best known algorithm for extracting the secret parameters of the embedding function is unaffordable.
- A data hiding system would be said to be provably secure if it is proved that disclosing the randomization applied to the embedding function.

References :

[1] Cox, I., Millar, M., and Bloom, J. 2002. "Digital watermarking", Morgan-Kaufmann, San Francisco, CA, ISBN: 1-55860-714-5.

[2] Craver, S., Memon, N., *at.el.*, "Can Invisible Watermarks Solve Rightful Ownerships?" IBM Technical Report RC 20509, IBM Research, July 1996. IBM Cyberjournal: http://www.research.ibm.

[3] December 2004, http//www.watermarkingworld.org/, “Digital Watermarking Frequenly Asked Questions”, last checked on 03 January, 2007.

[4] Freire, Luis P´erez, “Digital Watermarking Security”, Ph.D. Thesis, Submitted for the degree of Doctor in Telecommunications Engineering, Universidade de Vigo, 2008.

[5] Fu, Er-Hsien, “Literature Survey on Digital Image Watermarking”, EE381K-Multidimensional Signal Processing, 1998.

[6] Gokozan, Tolga, “Template Based Image Watermarking in the Fractional Fourier Domain”, MSc thesis, Middle East Technical University, January, 2005.

[7] Haque , S.M. Rafizul, “Singular Value Decomposition and Discrete Cosine Transform Based Image Watermarking”, Master Thesis, Computer Science, Department of Interaction and System Design, School of Engineering at Blekinge Institute of Technology, 2008.

[8] Meerwald, Peter, “Digital Image Watermarking in the Wavelet Transform Domain”, MSc thesis in University of Salzburg, 2001.

[9] Mohanty , Saraju P., “Digital Watermarking : A Tutorial Review”, Dept of Comp Sc and Eng. Unversity of South Florida, Tampa, FL 33620, http://www.csee.usf.edu

[10] Mohanty, Saraju P., “Watermarking of Digital Images”, MSc Thesis, Indian Institue of Science, January 1999.

[11] Stinson, Douglas R., “Cryptography: Theory and Practice”, Chapman and Hall/CRC, Boca Raton, FL, USA, 2nd edition, 2002.

[12] Tanenbaum, Andrew, S., “Computer Networks”, 4th Edition, Prentice-Hall of India Pvt. Ltd., New Delhi, 2003.

[13] Zheng, D., Liu, Y., Zhao, J., and El Saddik, A. “A survey of RST invariant image watermarking algorithms”, ACM Computing Surveys, Volume 39, No. 2, Article 5, June 2007.
