INVISIBLE WATERMARKING for TEXT FILES

MALA DUTTA

Senior Lecturer,
Dept. Of Computer Engineering
I.E.T.(D.A.V.V.),Indore, India

PRATEEK SAXENA

Master of Engineering
(Information Security)
I.E.T.(D.A.V.V.),Indore, India

ABSTRACT

Digital watermarking is a process of inserting copyright data in the main files. The data inserted is the watermark. This watermarking doesn't give any mark on the original content and it has no effect on their appreciation. The main aim of making this paper is to describe the basics of watermarking and to describe the history of watermarking, but also cover the applications, plus the ways that are used for evaluating watermarking systems. As a result, it inspires user to perform research more on certain watermarking.

I.INTRODUCTION

There are many security principles which are covered under watermarking. They are as follows:

A.Authentication

Authentication is the act of confirming an identity claim. When Joe Root visits a bank for making a withdrawal, he informs the bank teller he is Joe Root which is a claim of identity. The bank teller asks for a photo ID, so he provides the teller his voter-id. He checks it to make sure it has Joe Root printed on it and compares the photograph on the voter-id against the person who claims to be John Doe. If both the photo and name matches the person, then the teller has authenticated that Joe Root is the one who claims to be. Similarly by entering the correct password, the user is providing evidence that they are the person they username belongs to.

B.Non-repudiation

*N*on-repudiation means one's intention to fulfil their commitment to a contract. It states that the first party of a transaction cannot refuse after receiving a transaction nor can the other party refuse. In electronic commerce digital signatures and public key encryption are used to establish authentication and non-repudiation.

C.Availability

In information security, the information must be readily available when required. The main purpose of highly available system is to be available every time, thus preventing interruptions in services because of power failures, hardware failures and system updates. For ensuring In order to ensure availability, we must prevent D-o-s attacks, like a series of incoming messages to the end system which forces it to turn off.

D.Integrity

When talking in terms of data security, integrity means retaining and ensuring that the data is accurate and consistent over its life cycle. This means that data cannot be updated in way which is not authorized or detected. This is unlike referential integrity in databases, although it is a primary case of regularity according to classic ACID model of transaction processing. It violates however, when a message is continuously updated again and again over some period.

E.Confidentiality

ISSN: 2278 – 1323

Confidentiality means restricting the access to information and disclosing to users who are authorized and thus preventing access by or disclosure to unauthorized ones. Authentication methods such as IDs and passwords, that identify data systems' users uniquely and control access to data systems' resources, underpin the goal of confidentiality. Confidentiality is a wider concept of data privacy-limiting access to individual's personal information to be accessed.

II.COPYRIGHT

Copyright is a official right which is made by the law of a nation that allows the maker of an authentic work exclusive right to its use and distribution, with the motive of enabling the maker to get compensation for their intellectual attempt. Many exclusive rights typically adds to the maker of the original document that is listed below-

- For exporting and importing the work
- For creating some derivative works i.e. the work that adapt the original work
- For showing the work publicly
- For selling rights to other people
- For transmitting or showing by radio or video

A. *Copyright Infringement*

Piracy refers to the wrongful use of materials which are held by copyright. In order to consider a work as pirated, its unauthorised use must have happened in a country that has copyright laws to some established international convention like WIPO Copyright Treaty. Wrong use of things outside of the given legislation is termed "unauthorized edition", not piracy.

Piracy firstly aims software, music, and video. However, the unauthorised copying of text works still remains common, especially due to educational reasons. Records concerning the piracy effects are hard to find. Some research have tried to calculate a financial loss for industries that are afflicted through piracy by estimating what part of pirated jobs that had been purchased formally if they were not available freely.

III.BACKGROUND

A. How does watermarking help in copyright issues

Apart from the security practices, it is very helpful in solving issues related to copyright problems. This is explained in details as seen below:

Digital watermark is a marker type which is embedded in a hidden format in a noise tolerant signal. It is mainly used to recognise the actual holder of the signal's copyright. "Watermarking" is the process of masking digital information in a carrier signal such that the hidden information should be related to the carrier signal. Digital watermarks are also used to validate the authenticity, integrity of the signal or to show the original owners. Digital watermarks are only hidden under some cases, i.e. after using any algorithm, and imperceptible, otherwise. If a digital watermark distorts the carrier signal in a way that it becomes perceivable, it is of no use. Traditional Watermarks may be applied to visible media (like images or video), whereas in digital watermarking, the signal may be audio, pictures, video, texts or 3D models. A signal may carry several different watermarks at the same time. Unlike metadata that is added to the carrier signal, a digital watermark does not change the size of the original carrier signal after the process of addition is performed. The needed properties of a digital watermark depend on the use case in which it is applied. For marking media files with copyright information, a digital watermark has to be rather robust against modifications that can be applied to the carrier signal. Instead, if integrity has to be ensured, a fragile watermark would be applied.

B. Watermarking Life Cycle

The life cycle of watermarking goes through following steps:

1) *Watermark generation*: In this step a new watermark technique is implemented which is performed using the actual text contents. And partial copied contents become recoverable in children.

2) *Watermark embedding*: In this phase the water mark is embedded over the document which is invisible

   and with the clipboard transferable to the children in hidden format.

3) *Attack:* In this stage different operations such as insertion, deletion and editing are performed on original text.

4) *Recovery:* After manipulation of partial copied text, the watermark is recovered using the newly designed recovery system.

IV.CLASSIFICATION OF WATERMARKING TECHNIQUES

Watermarking techniques are divided on different basis.

A. Division Based on Human Perception

It is further sub-divided into visible watermarks and invisible watermarks.

1) *Visible Watermarks:* These watermarks can be seen clearly by the viewer and can also identify the logo or the owner. This technique changes the source signal. The watermarked signal is different from the original signal. Visible watermark embedding algorithms are less computationally complex. The watermarked image cannot with stand the signal processing attacks, like the watermark can be cropped from the watermarked image. Best option is to spread the watermark throughout the image, but this degrades the quality of image which protects the image for being used in medical applications.

2) *Invisible Watermarks*: These watermarks cannot be seen by the viewer. The output signal does not change much when compared to the original signal. The watermarked signal is almost similar to the original signal. As the watermark is invisible, the imposter cannot crop the watermark as in visible watermarking. Invisible watermarking is strong to attacks due to signal processing as compared to visible watermarking. As the quality of the image does not suffer much, it can be used in almost all the applications.

B. Division Based on Applications

On the basis of application, watermarks are sub-divided into following watermarks.

1) *Fragile Watermarks*: These watermarks are very sensitive. They can be destroyed easily with slight modifications in the watermarked signal.

2) *Semi-Fragile Watermarks*: These watermarks are broken if the modifications to the watermarked signal exceed a pre-defined user limit. If the limit or threshold is zero, then it operates as a fragile watermark. This method can be used to ensure data integrity and also data authentication.

3) *Robust Watermarks*: Robust watermarks have the ability to handle attacks due to signal processing, thus cannot be broken easily. It should remain intact permanently in the embedded signal in a way that any effort to abolish it will affect the quality of the image badly. This method can be used to ensure copyright protection of the signal.

C. *Division Based on Level of Information required to detect the embedded data*

Based on the level of required information all watermarks are sub-divided into blind watermarks, semi-blind watermarks and non-blind watermarks.

1) *Blind Watermarks*: These watermarks detect the embedded information without the use of original signal. They are less robust to any attacks on the signal.

2) *Semi-Blind Watermarks*: These watermarks need some kind of sensitive data to discover the embedded data in the watermarked signal.
3) *Non-Blind Watermarks*: These watermarks require the original signal to detect the embedded information in the watermarked signal. They are stronger against any attacks on the signal when compared to blind watermarks.

D. Based on User's Authorization to Detect the Watermark

This is sub-divided into public watermarks and private watermarks.

1) *Public Watermarks*: In this watermarking, the user is authorized to detect the watermark embedded in the original signal.
2) *Private Watermarks*: In this watermarking, the user is not authorized to detect the watermark embedded in the original signal.

E. Division based on knowledge of the user on the presence of the Watermark

This is sub-divided into steganographic watermarking and non-steganographic watermarking.

1) *Steganographic Watermarking*: The user is not aware of the presence of the watermark.
2) *Non-Steganographic Watermarking*: The user is aware of the presence of the watermark.

V. RECENT DEVELOPMENT

A. Text Line, Word or Character Shifting

This class of methods is used for shifting a text line, a group of words, or a group of characters by a minor amount for embedding data. They are applicable to documents with formatted text.

B. Fixed Partitioning of Images

This class of methods partitions an image into fixed blocks of size m x n, and computes some pixel statistics or invariants from the blocks for embedding data. They can be applied to binary document images in general; e.g. documents with formatted text or engineering drawings.

VI. OVERVIEW

In the current state of Internet's global network, we often like to save our digital data. In recent years, due to the sudden rise of Internet, all forms of media are available significantly. This can be seen in case of song recordings, in which, computer-based systems have created unbelievable sound of song content which are widely available and that too without the permission of the song's creators. The case is not special to the song industry because all content-originators like photographers, filmmakers etc have the same problem to different degrees. The solution to this problem of piracy is a current problem, some technology are created for help.

For this reason we need a security system. Digital watermarking is a solution for protecting content copyright in the global network. It imposes extra robustness on embedded information Digital watermarks are hard to abolish without affecting badly the original content in cases where cryptography is unable to provide robustness. The content watermarking is done by converting copyright information into some random digital noise through special algorithm that is noticed only by the creator. Watermarks are robust against filtering and remain with the content till the original has not been purposely destroyed. Digital watermarking is the method of obscuring a message that is associated with a digital signal (i.e. an image, song, video) within the signal itself. It is a concept which is closely related with steganography, in that they both obscure a message inside a digital signal. However, what separates them is their goal. Watermarking attempt to obscure any message related to the original data of the digital signal, while in steganography the digital signal has no association with the message, and it is just used in the form of cover to obscure its existence.

Recently, many digital watermarking techniques are seen which are based on discrete cosine transform, discrete wavelets transform and discrete fourier transforms. Here, an algorithm for digital image watermarking technique is proposed which is based on singular value decomposition; both of the U and L components are investigated for watermarking algorithm. This technique refers to the watermark embedding and watermark extracting algorithm. The experimental results show that the quality of the watermarked image is very good and robust to many geometrical attacks.

VII. OUTCOME

The large need of networked multimedia system has created the need of "COPYRIGHT PROTECTION". It is very important to protect intellectual properties of digital media. Internet playing an important role of digital data transfer. Digital watermarking is the important solution of the copyright protection problem. Digital watermarking is the solution for the protecting legal rights of digital content owner and the customer.

In this paper we discussed about digital watermarking technique. There are two types of digital watermarking techniques known as visible and invisible watermarking. Watermarking provides owner authentication. If we will use digital watermarking technique properly, we can save data from unauthorized duplicity.

This data, we hope, has helped the user to prepare for upcoming leanings about a certain watermarking and created some interest. This information can be supportive in covering the necessary issues that were discussed about watermarking. Watermarking is an active research field with many applications. Despite being a relatively new field, it has made some important algorithms for obscuring messages into digital signals. These can be described by many different models. Two broad categories for these models were described in this essay. These are communication-based models and geometric models. Communication-based models can be further divided into those which use side-information and those that don't. One example system was used to illustrate non-side-information models, and two example systems were used to illustrate side-information models. Each of these systems has its advantages and disadvantages, and each one trades some important watermarking property for another. The choice of which to use relies on the underlying application's requirements.

REFERENCES

[1] Content Based Watermarking of Images. Proceedings of the Sixth ACM

[2]http://en.wikipedia.org

[3] http://www.digitalwatermarkingalliance.org

[4] Cox I, Miller M, Bloom J, Fridrich J, Kalker T (2008) Digital Watermarking and Stagenography

[5] Steganography Second Edition. Elsevier, 2008

[6] W. Bender D. Gruhl N. Moromoto and A. LU Techniques for data hiding.

[7] C. Cachin. AN information- theoretic model for steganography. Proc. Of 2nd Workshop on information hiding, 1996