XML Digital Watermarking

Khalid Edris[1], Mohammed Adam Ibrahim Fakharaldien[2], Jasni Mohamed Zain[3], Tuty Asmawaty Abdul Kadir[4]

Software Engineering Research Department,
Faculty of Computer Systems & Software Engineering,
Universiti Malaysia Pahang
26300 Gambang, Kuantan, Pahang, Malaysia
Khalod07@yahoo.com, mohammedfakeraldeen@yahoo.com, jasni@ump.edu, tuty@ump.edu

Abstract. Due to the widespread and population of internet, digital data is copied and reproduced easily, which generate the high demand for copyright protection. Digital watermarking is one of the most efficient methods to protect the digital data. Digital watermarking is the process of hiding digital signal into digital document. The digital signal is called watermark. This paper covers the basic concepts of digital and XML watermarking.

Keywords: Digital watermarking, Copyright protection, Attack, XML.

1 Introduction

With the increasing of using internet, people can share and access information without difficulty, also they can sent or receive any type of data such as text, image, audio, and video. Now a day's many educational, banks, organizations, and industries use internet for their working and transactions, so it is very important to protect the data from illegal users, access and copy. To protect data there are many ways like cryptography, steganography and watermarking, the best way is digital watermarking.

Digital watermarking is the process of embedding information into digital signal, which used to verify the authenticity or the integrity of the signal or to show the identity of its owners. Digital watermarking can be used for copyright protection, owner identification to identify the owner, finger printing to identify the customer, broadcast monitoring and authentication to determine whether the data is changed or not.

2 Digital Watermarking Phases

Digital watermarking process composed of three stages:

2.1 Embedding:

In embedding, an algorithm accepts the host and the data to be embedded, and produces a watermarked signal. The watermark is embedded in the original signal; the watermarked digital signal is transmitted to another user.

2.2 Attack:
Normally the attack happened when someone makes a modification. However, modification may not be malicious; the term attack arises from copyright protection application, where third parties may attempt to remove the digital watermark through modification.
2.3 Detection:
Detection or extraction is a method which is applied to the attacked signal to attempt to extract the watermark from it. If the signal cannot modify during transmission process, then the watermark still is present and it can be extracted. In robust digital watermarking applications, the extraction method must produce the watermark correctly, even if the modifications were strong. While in fragile digital watermarking, the extraction method must fail if any change is made to the signal [1].

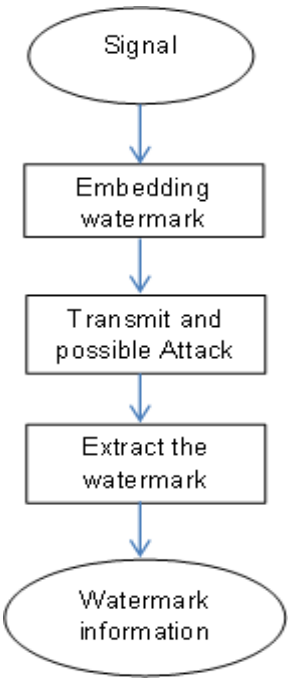

Fig.1. Fundamental Process of digital watermarking

3 Basic Characteristic of Digital Watermarking

There are a number of characteristics, the most important of them are:
3.1 Robustness:
Robustness refers to that the watermark embedded in data has the ability of surviving after many processing operations and attacks. Then, the watermark must be robust for general signal processing operation, geometric transformation and malicious attack. The watermark for copyright protection does need strongest robustness and can resist

malicious attacks, while fragile watermarking; annotation watermarking do not need resist malicious attacks.

3.2 Imperceptibility:
Means the watermark can be detected by special processing or special circuits and undetectable by a human perceptual system [2].

3.3 Security:
The ability of watermarks to prevent illegal users from tampering, then the watermark only can be detected or extracted by authorized users.

3.4 Capacity:
Watermarking capacity is the size or number of bit information that can be hidden into content signal.

3.5 Complexity:
Complexity is concerned with the time needed to embed, detect, and extract the watermark from content [3, 8].

4 Classification of Digital Watermarking:

Digital watermarking can be classified into different categories according to many different features as follow [2]:

4.1 According to the type of documents:
The important types are:

i. Image watermarking: add watermark in the image.
ii. Video watermarking: embed information in the video stream.
iii. Audio watermarking: add watermark to the audio.
iv. Text watermarking: add watermark to PDF, DOC, XML and other text file to prevent changes of text.

4.2 According to the working domain:

i. Spatial domain watermarking:
The watermark embedding is done on image pixels, it is simple and with low complexity, but it is more prone to attack.
ii. Frequency domain watermarking:
The watermark embedding is done on the image after converting the image into frequency domain; it is complicate, but more robust than spatial domain.

4.3 According to the human perception:
There are three types:

i. Visible watermarks:
The watermark is added to the image but it is transparent, the watermark is appeared to the viewer.

ii. Invisible watermarking:
The watermark is embedded in the original image that it cannot be seen by human eye.
iii. Dual watermarking:
Is a combination of visible and invisible watermarks, the invisible watermark is used to as backup for the visible watermark.

4.4 According to the ability of watermark to resist attack:

i. Fragile watermarks:
Fragile watermark is designed in a way that any modification will effect in the watermark, it is very sensitive to the changes of signals and it used for integrity protection.

ii. Robust watermarks:
The watermark is not destroyed after some attack and can be extracted, thus it is used in copyright protection.

4.5 According to the detection process:

i. Blind extracting watermark:
This type of watermarking does not need the original data in the watermark detection and extraction. Watermark detection becomes difficult when the watermarked data is destroyed.
ii. Non-blind extracting watermark:
Non-blind or visual watermarking needs the original data in the watermark detection and extraction. It has stronger robustness but it is used with limited application [1, 2, 5, 6].

5 Digital watermarking attacks:

Attacks on watermarks happened by accidental or intentional. Accidental attacks may cause due to the standard image processing or due to the compression procedures. There are many categorizations of attacks on watermarks. The Following are the methods of attacks according to robustness and Perceptibility:

5.1 Mosaic attack:

This attack usually happened with pictures and displayed it as to confuse watermark-searching program, known as “Web Crawler”.

5.2 Geometric attack:

This type of attack usually attack images, documents and audio files. Geometric attack can be classified as:

5.2.1 Subtractive attack:

Subtractive attack happened in the area of located watermark if imperceptible then removing the mark by cropping or digital editing. It performed on robust watermark.

5.2.2 Distortive attack:

Attacker tries to make some distortive changes in the images such that mark becomes unrecognizable. This attack performed on robust watermark.

5.3 Stirmark attack:

Stirmark attack can be defined as generic tool developed for simple robustness techniques of image marking algorithms and steganographic techniques. Stirmark can simulate resampling process in which it introduces same kind of errors into an image to print it on high quality printer and scanning it again with high quality scanner.

5.4 Forgery attack:

In this attack, the attacker adds watermark overlaying to the original image to marking the content as their own.

5.5 Inversion attack:

In this attack, the attacker when receives watermarked data claim that data contains his watermark by declaring part of data as his watermark. The attacker also can easily generate the original data by subtracting the claimed watermark.

5.6 Cryptanalysis attack:

It is a way in which attacker attempts to find the decryption key for an encrypted pieces of information so that it can be made useful again [8, 9].

6 Applications of digital watermarking:

There are different types of watermarking applications; the most important applications are described ass following:

6.1 Copyright protection:

By identify the ownership; Digital watermarking can be sued for copyright protection to deny the other parties from claiming the copyrights.

6.2 Finger print:

Digital watermarking is used to recognize the content buyers by giving the information about the customers like serial number, which help to trace the source of illegal copies.

6.3 Content authentication:

Authentication is very important to use because it confirm the integrity of watermarked data and to make sure that the data is not being tampered.

6.4 Copy protection:

Copy protection is a set of rules used to deny people from making illegal copies of the content. These rules can be like “this document may not be copied, or this document may be copied but no copies may be made of the copy”.

6.5 Broadcast monitoring:

Digital watermarking can be sued to protect broadcasted content like TV from illegal transmission. Watermark identifies the owner of the content, when it detected with the aid of automated system monitor when and where the content appears [1, 4, 7].

7 XML digital watermarking:

7.1 Brief introduction of the XML documents

XML or Extensible Markup Language is used to establish the markup language. It is put forward by W3C to overcome the limitations of HTML (or Hypertext Markup Language), which is the basis for all websites. Similar to HTML, XML is based on SGML, or Standard Generalized Markup. Though SGML has already exited in the publishing industry for several decades, its complexity has discouraged many people who have the intention to use it.

Since XML was put forward in 1998 by W3C many researches has been put, however, the major of these researches in XML security. The research focus has shifted from XML documents' digital-signature and encryption to its visit control and security protocol, etc. Currently, some new requirements for security have been constantly in the spotlight, such as copyright protection of XML documents. Therefore, XML digital watermarking technology has come into being.

7.2 Research situation of XML digital-watermarking

XML has provided a method to describe data and it is suitable for many applications. With the development of network technology, more data are expressed in XML. Since XML documents are texts, to copy XML documents is very simple. This puts XML documents at the threat of illegal copy. Therefore, it is very important to solve the problem of copyright protection of XML documents.

There are three kinds of message similar to XML documents, namely text, database and website. The research of the digital-watermarking of database and network has achieved certain progress. Due to the fact that XML documents are semi-structured data, it is practicable to imply its structure and data to embed into the watermarking message. XML can establish documents seemingly different while the same data or semantic value. This differs from the structure of the entities, the arrangement of the attributes, encoding of the characters or unnoticeable margins. Currently,, researchers mainly focus on the study of XML digital watermarking technology based on XML logical structure and content. It can generally be classified into four types as follows [10]:

7.2.1 XML digital-watermarking based on the logical structure

Based on research findings of the message hiding method of XML, there are mainly five methods for the message hiding of XML documents. These methods can classify as follows:

(1) Blank elements

When the element doesn't include any markup, it is called blank element, which can be illustrated by the following two examples, that is the line feed of HTML (
) and image (<img>). In the blank elements of the XML documents, the ending slash can be put at the beginning markup. The above mentioned two examples are the same to the XML analyzer.

```
<!-- Two equivalent break elements -->
<br></br>
<br/>
<!-- Two equivalent image elements -->
<img src="xml/img/c.gif"></img>
<img src="xml/img/c.gif" />
```

Thus, different blank elements can be employed to embed into watermarking message. The application of <img></img> calls for that the byte of the hided watermarking message should be zero, and in terms of "<img/>", the byte should be 1.

XML document before the adding of watermark:

```
<img src="f1.jpg"></img>
<img src="f2.jpg"></img>
<img src="f3.jpg"></img>
<img src="f4.jpg"></img>
<img src="f5.jpg"></img>
```

Watermarking message needs to be added: 10110

XML document after the adding of the watermark message:

```
<img src="f1.jpg"/>
<img src="f2.jpg"></img>
<img src="f3.jpg"/>
<img src="f4.jpg"/>
<img src="f5.jpg"></img>
```

(2) Blank space within the markup

Blank space can be used in many places in XML document, like blank character, tab character, carriage return and line feed. Sometimes, the adding of the blank space is to indent the XML documents; while in other conditions, the blank space of XML documents can be added into the watermarking message. Whether to add blank space to the ending of the markup or not is to express the same message. By taking advantage of the attribute, it can be embedded into the watermarking message.

Use <tag>, </tag> and <tag/> to express the byte of the hided watermarking message as 0;

Use <tag>, </tag > and <tag/ > to express the byte of the hided watermarking message as 1.

The XML document before the adding of the watermark:

```
<user><name>Alice</name></user>
<user><name>Bob</name></user>
```

The watermark message to be added: 1010 0101

The document after adding watermark:

<user ><name>Alice</name ></user>

<user><name >Bob</name></user >

(3) To change the order of the element

Watermark message can be embedded by changing the element order of XML documents:

<tag1><tag2>data</tag2></tag1> to express the byte of the hided watermark message as 0

<tag2><tag1>data</tag1></tag2> to express the byte of the hided watermark message as 1

(4) To change the attribute order

Watermark message can be imbedded by changing the attribute order of the XML document:

<tag att=”att1” att=”att2”/> is to express the byte of the hided watermark message as 0.

<tag att=”att2” att=”att1”/> is to express the byte of the hided watermark message as 1.

(5) Element embedding

When element embedding is used for digital-watermarking design, tags should be mutually embedded. For example, <tag1> and <tag2> can be mutually embedded. Thus, they are practicable.

<tag1><tag2>data<tag1><tag2> is to show the byte of the hided watermark message as 0

<tag2><tag1>data<tag2><tag1>is to show the byte of the hided watermark message as 1

7.2.2 XML digital-watermarking combining the logical structures and content

The process of combining the logical structure and its content in XML watermarking normally can be done in two steps:

Firstly, node content of the XML documents can be images, text, data, software, etc. Taking advantage of the watermark algorithms, part of the watermark is embedded into the node content.

Finally, the watermark is assembled into the whole watermark through the logical structure of XML documents or tree structure.

7.2.3 XML digital-watermarking based on statement enquiry

David Gross-Amblard proposed the methods of XML digital-watermarking based on the parameterized statement inquiry [12, 13, 14].

Within the acceptable error access, watermark can be embedded through a serious of statement enquiries methods of a certain language.

7.2.4 XML digital-watermarking based on the content

Wilfred Ng et al. put forward to embed the watermark message into the XML documents through the selective approach and the compression approach [15]. The former is by changing the value-type data of the XML documents; and the latter is through the method of compressing the XML documents.

8 CONCLUSIONS:

This paper includes the concepts of digital watermarking and XML digital watermarking technology; it gives review information about the classification of digital watermarking according to the various criteria, the basic characteristics of digital watermarking and digital watermarking attacks. The paper also includes the applications of digital watermarking and classifications of XML digital watermarking.

References

1. J. Jaspreet and K. Karmjeet, "Digital Watermark: A study", International Journal of Advanced Research in Computer Science and Software Engineering, August 2012, Volume 2, Issue 8.
2. S. Shraddha, "Digital watermarking: Review", International Journal of Engineering and Innovative Technology (IJEIT), February 2012, Volume 1, Issue 2.
3. Z. Yanquan, "Digital Watermarking Technology: A review", International Conference on Future Computer and Communication, 2009.
4. P. Manjunatha and K. Shivaprakash, "A comprehensive Survey of Contemporary Researches in Watermarking for Copyright Protection of Digital Images", International Journal of Computer Science and Network Security, April 2009 Vol.9 No.4.
5. L. Robert, "A study on Digital Watermarking Techniques", International Journal of Recent Trends in Engineering, May 2009, Vol. 1, No. 2.
6. J. Zunera and M. Anwar, "A review of Digital Watermarking Techniques for Text Documents", International Conference on Information and Multimedia Technology, 2009.
7. H. Faroog, "A survey of Digital Watermarking Techniques for Multimedia Data", International Journal of Electronics and Communication Engineering, Jan 2012,Vol 2, No. 1, pp. (37-43).
8. R. Nidhi, "Digital Watermarking", Global Journal of Computer Science and Technology, 2012, Volume 12 Issue 13 Version 1.0.
9. R. Ridzoň, , D. Levický, and Z. Klenovičová, "Attacks on watermarks and adjusting PSNR for watermarks application",14th international Czech - Slovak scientific conference, Bratislava, s, April 2004, 374-377,.
10. Y. Jie, "Algorithm of XML document information hiding based on equal element". In Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on. 3:250-253.

11. A. Castiglione, B. D'Alessio, A. D. Santis, and F. Palmieri, "Hiding Information into OOXML Documents: New Steganographic Perspectives", Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 2011, 2(4), 59-83.
12. D. Gross-Amblard, "Query-preserving watermarking of relational databases and XML documents", Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. ACM, 2003, 191-201.
13. C. Constantin, D. Gross-Amblard, and M. Guerrouani, "Watermill: an optimized fingerprinting system for highly constrained data". Proceedings of the 7th workshop on Multimedia and security, 2005, 143-155.
14. J. Lafaye, and D. Gross-Amblard, "XML streams watermarking Data and Applications Security", 2006, XX. 74-88.
15. W. Ng, and H. L. Lau, "Effective approaches for watermarking xml data", Database Systems for Advanced Applications. Springer Berlin/Heidelberg, 2005, pp. 68-80.