Data Leakage Detection using Cloud Computing

Tushar Aggarwal[1], Narinder Kaur[2]

[1]Student, Department of IT, Maharaja Agrasen Institute of Technology, Delhi, India
[2]Assistant Professor, Department of IT, Maharaja Agrasen Institute of Technology, Delhi, India

---***---

Abstract - *Data leakage is one of the biggest challenges in front of the IT industry's & different institutes worldwide [1]. Data is mainly sent by the distributors which are generally the owner of data to the user which wants the information mainly the trusted third parties. The digital information shared by the distributor should be confidential and must be shared by a secure way. In some scenarios, the data distributed by the organization is copied by different agents who cause a huge damage to the organization and this is termed as data leakage. The data leak must be detected as early as possible in order to prevent the confidential data from being public or for some malicious use. Data leakage is sometimes used by other companies in the competition to downgrade the former company or for some other business tactic. This project deals with protecting the data from being out sourcing by giving a special inscription to the sensitive data so that it cannot be reproduced, known as watermarking.*

Key Words: Data Leakage, confidential data, detection, watermarking.

1. INTRODUCTION

In this high speed century, technology has bloomed in every aspect of life. Countless of data is created in every fraction of a second. With internet, creating, storing and transferring of digital data (images, video and audio files) has grown many fold. But this ease has given rise to issues like conspiracy, fraudulent, copyright issues, piracy & authenticity while handling the digital multimedia. Copying a digital data takes no amount of efforts and is fast [2].

This gives rise to the issues like protection of rights and proving ownership of the content. There is no basis or timing of data leakage, it can happen at any time. The extent of damage done by a data leakage depends solely on the quality of how sensitive (confidential) data is leaked by that person. The leakage could lower the business and may result in the downfall of the company. For these challenges digital watermarking is a viable solution.

In Digital watermarking, information is embedded inside the cover media, which can be later extracted for authentication. Watermark can be visible or invisible. Watermark should be robust to resist manipulation while preserving the image quality. Digital watermarking is a technology for confidential or secret information inside digital content. It can be audio, video, image etc that the creator or owner would like to protect. Information is embedded as a watermark, for protecting copyrights and proving the validity of data. A digital watermark is a digital signal, pattern or another image inserted into digital content. The main purpose of the watermark is to identify who is the owner of the digital data.

2. Introduction to Image Watermarking

Different Organizations have different Image watermarking objectives, which justify different watermarking needs as per usage.

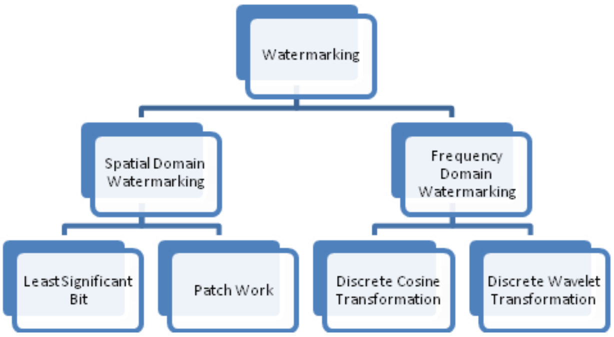
Watermarking
Spatial Domain Watermarking
Frequency Domain Watermarking
Least Significant Bit
Patch Work
Discrete Cosine Transformation
Discrete Wavelet Transformation

Chart-1: Different Watermarking Techniques

Some of the objectives are:

1. Owner identification
2. Copy protection
3. Broadcast monitoring
4. Medical applications
5. Fingerprinting
6. Data Authentication

Digital image watermarking is typically derived from Steganography, a process in which digital content is supposed to be hidden within another digital media for secure transmission of data. In some conditions steganography and watermarking are very similar when the data to be secure is hidden in process of transmission over some carrier. The main difference between these two processes is that, in steganography the hidden data is on the highest priority for sender and receiver but in watermarking, signature or watermarked data is on the highest priority. The aim of a typical watermarking scheme is that whenever there is a malicious attack on the digital media, it must keep the inserted watermark very robust in spectral and real domains.

2.1 Process of Watermarking

The procedure of watermarking is divided into 2 parts:

1) Embedding watermark - Image watermarking is done at the source end. In this, watermark is inserted in the source image by using any watermarking algorithm.
2) Extracting watermark - Extracting watermark from the watermarked image by reversing the embedding algorithm.

In digital watermarking mechanisms, three mutually exclusive parameters are used to evaluate performance.

1) Quality-Minimum distortion of the original image after confidential message has been embedded.
2) Capacity-Maximum size of watermark image embedded on cover image.
3) Robustness-Watermarked should withstand any modification attack.

Image watermarking can be classified on the basis of various criteria: type, robustness, domain, perceptivity, host data and data extraction. In the image watermarking, domain based techniques are generally used. They are spatial domain and frequency domain. Out of these two, frequency domain is more used than spatial domain watermarking.

3. Watermarking Techniques

3.1 Spatial-domain watermarking

In spatial domain, the information can be added by varying the values of the carrier signal. The values of color or the intensities of some selected pixels are directly modified by it. The spatial domain watermarking methods are simple and less robust against various geometric and non geometric attacks. The principle quality of spatial domain techniques is that they have low computational complexity. This technique is less tedious when compared to frequency domain watermarking. Various methods listed in spatial domain are Additive watermarking, Least Significant Bit, Patchwork and Text mapping coding etc [10].

Least Significant Bit (LSB)

LSB technique is the most straightforward method to implement. In this method, watermark bit is embedded to each pixel of lsb. The degradation of quality of image is less and difference is not visible to naked eyes. It has high perceptual transparency. In spite of being easy to implement, it is not robust against attacks, attacker can destroy the watermark. This method can't be utilized for practical purposes as it very sensitive to noise and is vulnerable to sound.

3.2 Frequency Domain Watermarking

Unlike spatial domain technique, frequency domain inserts a message by changing the transform coefficients of the cover image rather than pixel values, and these coefficients are perturbed by a small amount in one of the several ways in order to represent the watermark. The watermark is irregularly distributed over the entire spatial image upon inverse transform, which makes it more difficult for enemies to decode and read the mark. Cropping may be a serious threat to any spatially based watermarking scheme but it is less likely to affect a frequency-based scheme [9].

Applied to digital images, yet there are outstandingly six most commonly used transforms in image watermarking. They are Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Discrete Fourier Transform (DFT), Lifting Wavelet Transform (LWT), Discrete Hadamard Transform (DHT) and Singular Value Decomposition (SVD).

There are present a number of transforms which can be applied to digital images, yet there are outstandingly six most commonly used transforms in image watermarking. They are Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Discrete Fourier Transform (DFT), Lifting Wavelet Transform (LWT), Discrete Hadamard Transform (DHT) and Singular Value Decomposition (SVD).

There are present a number of transforms which can be applied to digital images, yet there are outstandingly six most commonly used transforms in image watermarking. They are Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Discrete Fourier Transform (DFT), Lifting Wavelet Transform (LWT), Discrete Hadamard Transform (DHT) and Singular Value Decomposition (SVD).

There are present a number of transforms which can be applied to digital images, yet there are outstandingly six most commonly used transforms in image watermarking. They are Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Discrete Fourier Transform (DFT), Lifting Wavelet Transform (LWT), Discrete Hadamard Transform (DHT) and Singular Value Decomposition (SVD).

There are present a number of transforms which can be applied to digital images, yet there are outstandingly six most commonly used transforms in image watermarking. They are Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Discrete Fourier Transform (DFT), Lifting Wavelet Transform (LWT), Discrete Hadamard Transform (DHT) and Singular Value Decomposition (SVD).

There are at present a number of transforms which can be applied to digital images, yet there are prominently six most commonly used transforms in image watermarking. They are Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Discrete Fourier Transform (DFT), Lifting Wavelet Transform (LWT), Discrete Hadamard Transform (DHT) and Singular Value Decomposition (SVD).

Discrete Cosine Transform

In DCT, an image is broken up into different frequency bands, making it easy to embed watermarking information into the middle frequency bands of an image, this ensures that watermark will not be removed by any kind of attack. The 2-dimensional Discrete Transform of a matrix gives the transform coefficients in the form of other matrix. The lowest frequency coefficient is represented by the left most corner of the matrix while the highest frequency coefficient is represented by the right bottom most corner of the matrix. DCT is real transform which has better computational efficiency compared to DFT, which is by definition complex [5]. A major drawback observed was that the block wise DCT destroys the invariance properties of the image [8].

Discrete Wavelet Transform

DWT is used more frequently used due to its excellent spatial localization and multi-resolution characteristics, which are similar to the theoretical models of the human visual system. The DWT splits the signal into two parts, i.e., high and low frequency parts. The high frequency part contains information about the edge components, while the low frequency part is split again into high and low frequency parts. The former is usually used for watermarking since the human eye is less sensitive to changes in edges.

In two dimensional applications, for each level of decomposition, DWT is first performed in the vertical direction, followed by in horizontal direction. After the first level of decomposition, there are 4 sub-bands: LL1, LH1, HL1, and HH1. For each successive level of decomposition, the LL sub band of the previous level is used as the input. To perform second level decomposition, the DWT is applied to LL1 to perform third level decomposition, the DWT is applied to LL2 band which decompose this band into the four sub-bands – LL3, LH3, HL3, HH3. This results in 10 sub-bands per component. LH1, HL1, and HH1 contain the highest frequency bands present in the image tile, while LL3 contains the lowest frequency band and the approximate image, and after that the coefficients of frequency are modified depending on the transformed coefficients of the watermark and afterwards a robust watermarked image is obtained. Then the inverse discrete wavelet transform (IDWT) is applied to obtain the reconstructed image [6].

In DWT, we get a very high compression ratio but we lose some minimum amount of information. If we go beyond one level, we may get higher compression ratio but the reconstructed image is not identical to original image [12].

There are disadvantages with DWT as well: The utilization of greater DWT basis functions produces ringing noise and blurring near edge areas in video frames or images. It has very large compression time and is less robust against various geometric attacks [7] [8].

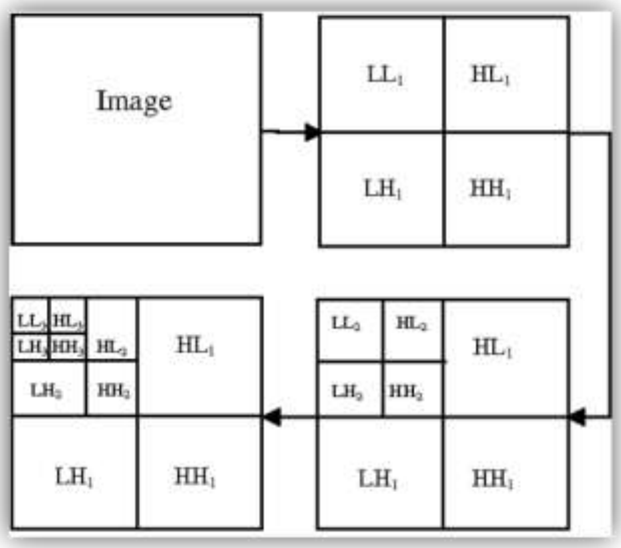

Fig-1: 3-Level discrete wavelet decomposition

4. Singular Value Decomposition

The Singular-Value Decomposition is a matrix decomposition method for reducing a matrix to its constituent parts in order to make certain subsequent matrix calculations simple. SVD is quite an effective method to split the system into a set of linearly independent components [4]. There are many advantages to use SVD in digital image processing. Firstly, the SVD transformation can be applied to an image with arbitrary sizes [11]. Secondly, singular values of the digital image are less affected if general image processing is performed. Lastly singular values contain intrinsic algebraic properties of an image [3].

5. Watermarking by Embedding and Extraction

In this method, the insignificant part of the fractional part of the pixel intensity value of the cover image is encoded to provide watermark. The watermark in the insignificant maintains the accuracy of the image. A large amount of watermark can be easily embedded and extracted using this method which will help companies and firms involved in digital information security products, this is an added advantage of this method. Various algorithms for embedding and extraction are used in this technique.

6. Spread Spectrum Watermarking

Most watermarking algorithms add watermark in the significant component of the signal to make it robust against common signal distortions and malicious attacks. This technique can be used to protect audio, video and image. But this kind of modification can lead to degradation of the data signal. So, watermark must be added to the spectral components of the data using techniques similar to spread spectrum communications, hiding a narrow band of the signal in a wide band signal. This watermark is very difficult for an outsider to remove; even many outsiders collude with different copies of the watermarked data.

7. CONCLUSIONS

Frequency domain methods are more useful than spatial domain as they deal with the rate of pixel change instead of image plane.

Out of all methods in frequency DWT is more effective watermarking algorithm because it has peak sound to noise ratio(PSNR) against different attacks whereas other algorithms are not so robust against attacks like salt and pepper, rotation, enhancement, Gaussian noise and speckle attack.

In DCT, reconstructed image is not as the original image but is seemingly identical to original because of block artefacts. For this choose small size of blocks.

Simulation results show that DWT is somewhat better than DCT but, the combination of these two will give much better result than individual [8].

These algorithms can be used to embed watermark in audio and video and a combined modification of DCT & DWT will give much better result in MATLAB. This can be done in the next step.

REFERENCES

[1] Abhijeet Singh, Abhineet Anand, "Data Leakage Detection Using Cloud Computing", International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 6 Issue 4 April 2017, Page No. 21141-21144.

[2] Sandilya Pemmaraju, V. Sushma & Dr. K. V. Daya.Sagar," Data Leakage Detection using Cloud Computing", Global Journal of Computer Science and Technology: B Cloud and Distributed Volume 14 Issue 3 Version 1.0 Year 2014.

[3] Anumol Joseph, K. Anusudha," Robust watermarking based on DWT SVD", International Journal of Signal & Image Processing, Issue. 1, Vol. 1, October 2013.

[4] Rowayda A. Sadek," SVD Based Image Processing Applications: State of The Art, Contributions and Research Challenges", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 7, 2012.

[5] Chetna," Digital Image Watermarking using DCT", IJCSMC, Vol. 3, Issue. 9, September 2014, pg.586 – 591.

[6] Sumedh P. Ingale, Prof. C. A. Dhote," Digital Watermarking Algorithm using DWT Technique", IJCSMC, Vol. 5, Issue. 5, May 2016, pg.01 – 09.

[7] Lalit Kumar Saini, Vishal Shrivastava," A Survey of Digital Watermarking Techniques and its Applications", (IJCST) – Volume 2 Issue 3, May-Jun 2014.

[8] Mandeep Singh Saini, Venkata Kranthi B, Gursharanjeet Singh Kalra, "Comparative Analysis of Digital Image Watermarking Techniques in Frequency Domain using MATLAB SIMULINK", (IJERA)-Vol. 2, Issue 4, May-Jun 2012

[9] https://www.researchgate.net/publication/220413592_Frequency_Domain_Watermarking_An_Overview.

[10] https://www.researchgate.net/publication/318361111_Spatial_and_Frequency_Domain_Digital_Image_Watermarking_Techniques_for_Copyright_Protection.

[11] https://en.wikipedia.org/wiki/Singular_value_decomposition.

[12] Anilkumar Katharotiya, Swati Patel, Mahesh Goyani, "Comparative Analysis between DCT and DWT Techniques of Image Compression",(IISTE)- Vol 1, No.2, 2011