ISSN No. 0976-5697

Volume 4, No. 4, March-April 2013

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

# Attacks on Digital Watermarked Images in the Internet Environment and Their Counter Measures

Tapas Bandyopadhyay*
Scientist F, STQC, Kolkata-91, India
tapas1_banerjee@yahoo.com

B Bandyopadhyay
Professor, C.U. , India
b_bandyopadhyay@yahoo.com

B N Chatterji
Ex Prof IIT,Kharagpur, India
bnchaterji@gmail.com

***Abstract:*** Web application has made it faster and convenient to keep and transact important multimedia contents such as image through the Internet. Digital watermarking to the images can secure the information to a great extend to establish the authenticity of the owner of the images to protect copyright against unauthorized claiming of ownership of the image [1][2][5][6]. Digital watermarking on the image is an effective technique for protection of ownership rights in the untrusted open internet environment. However, the success of a digital watermarking technology depends heavily on its robustness to withstand attacks that are aimed at removing or destroying the watermark from its host data (image artefact)[3]. This paper aims at to analyse a number of digital image watermark attacks that the watermarked image may face and attempt has been made to classify them into different categories [7]. A set of experimental results are also provided to show the effect of these attacks on watermarked images in the internet environment.

***Keywords:*** Digital watermarking, authentication, PSNR, Correlation coefficient, finger print, attacks

## I. INTRODUCTION

Digital information (asset) is very susceptible to having copies made at the same quality as the original easily. The purpose of digital watermarks is to provide copyright protection for intellectual property that is in digital format [4][2]. An artist creates an original image and watermarks it before passing it for distribution in the internet environment. If another malicious person claims the ownership of the image and sells copies to other people, then the original creator can extract the watermark from the image proving the copyright to it. The strength of the system is that the creator will only be able to prove the copyright of the image if the malicious person is not able to modify the image such that the watermark is damaged to a great extent to be undetectable or used his own watermark such that it is difficult or near impossible to discover which watermark was embedded first in the image [13]. The quality of digital watermarks can be assessed in two ways; firstly it must be able to resist intentional (by malicious user) and unintentional (signal processing) attacks and secondly the embedded watermark must not detract from the visual quality of the image. Figure 1 depicts the attack of the watermark in the internet environment channel.

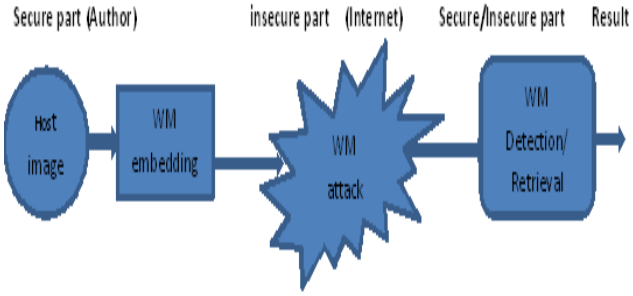


Figure1: Life cycle phases of watermark (WM) attack in the internet channel

## II. APPLICATION OF DIGITAL WATERMARKING AND ASSOCIATED THREATS

The watermarking is being used and has placed its position in the arena of information security because of its different advantages in respect of image authentication, establishing copy right control, checking tampering etc[13][14]. In general, if it is useful to associate some additional information with art work / image, this metadata can be embedded as watermark. Robust watermark is being used in applications such as Copy Control, Evidence of Ownership, and Fingerprinting for official and legal documents like Identification card, passport, where it is important to establish the authentication of the source issuing it[5]. In such applications of digital watermarking there is a possibility of defeating the purpose by tampering or removing the watermark for malicious reason. To initiate any attack it requires three components namely the threat agent, asset and adverse action on the asset. Depending on the type of digital artifact document (its asset value) different attacker will target the attack on it with different attack potential (equipped with expertise, tool, motivation, window of opportunity) to desynchronize, remove or distort the watermark so that original creator may face problem in claiming its ownership.

### *A. Robust Watermark:*

Applications: Copy Control, Evidence of Ownership, Fingerprinting etc.
Requirement of Robust Watermark: The watermark can still be detected even after severe signal processing.
Attacker's goal against Robust Watermark: To make the detector unable to detect the watermark while keeping the perceptual quality

### *B. Fragile Watermark:*

Applications: multimedia (digital artifact) authentication

Requirement of Fragile Watermark: Determine if the content of the digital artifact (watermarked image) has been changed. It is difficult for an unauthorized person to insert a valid watermark in the artifact.
Attacker's goal against Fragile Watermark: To make the watermark still valid after alteration of Work and generate a valid Work for new data

## III. WATERMARKING ATTACK POTENTIAL AND A ATTACK CATEGORY

The attacks on the watermarking in images are broadly categorized in two types namely: unintentional common signal processing and image manipulation and intentional attack by malicious attacker.

### *A. Unintentional Common Signal Processing And Image Manipulation:*

For the image processing purpose many image manipulations/ modification are commonly used by the artists/author in preparing material for printing and making it suitable for publication in the internet for distribution. Common image manipulations include image rotation, resizing, cropping, sharpening, filtering, compression, printing/scanning, geometric transform, compositing/ mosaicking, colour tablets etc. Again for Internet applications, image segmentation called tiling is used. It is expected that any embedded watermark claiming to be robust should survive any of these image manipulations and remain detectable in the manipulated image [16]. An embedded watermark must survive any image manipulation that does not damage or destroy the image characteristics beyond usability for the purpose; otherwise, its credential as a security of intellectual property right (IPR) is doubtful and questionable.

### *B. intentional attacks by malicious attacker:*

#### *a. Removal and interference attacks:*

The signal processing related attacks such as denoising, remodulations, collusions, averaging types of attacks may be triggered by malicious attacker. Removal attacks are watermarking attack that aims at removing the watermark signal from the watermarked image without attempting to break the security of the algorithm. These types of watermark attack do not attempt to find out the encryption techniques used or how the watermark has been embedded [17]. It results in a damaged watermarked image, hence a damaged watermark signal, where no post processing can recover the watermark from the attacked data. Noising, histogram equalization, blur and sharpen attacks are included in this category.

#### *b. Geometry attacks:*

The geometric attacks are rather different from removal attacks. Instead of aiming to remove or severely damage the watermark signal, this type of attack intends to distort the watermark signal. It is however still theoretically possible for the detector to recover the original watermark if the detail of the geometric attack can be established and countermeasures applied. The process of correcting this type of attack is of often called synchronization. However, the complexity of the required synchronization process might be too expensive and slow. Included in this category of watermark attack are mage rotation, scaling, translation and skewing.

#### *c. Desyncronisation attacks:*

The disabling detector synchronization attacks, all classes of geometrical affine and projection transform, template removal collage etc. comes under this type of attack.

#### *d. Cryptographic:*

The cryptographic attacks (based on crypto analysis) trial and error modification of the protected media. The aim of cryptographic attacks is to attack the methods in watermarking schemes and thus find a way to remove the embedded watermark information or to embed misleading watermarks [12]. One of the techniques is the brute-force search method. This technique attempts to crack the watermark security by a large number of known possible measures to search meaningful secret information. Another technique is called Oracle attack, which is used to create a non-watermarked signal when a watermark detector device is available.

#### *e. Protocol attack:*

The protocol attack introduce ambiguity in the trusted watermarking protocol, copy attack: the main idea of copy attack is to copy a watermark from one image to another image without knowledge of the key used for the watermark embedding to create ambiguity with respect to the ownership of the original image [15].Protocol attacks are a further type of watermark attack. Whereas the other types of attacks aim at destroying, distorting or extracting the watermark signal, protocol attacks adds attacker's own watermark signals onto the work under attack. This results in ambiguities on the true owners in question. Protocol attacks target the entire concept of watermarking techniques as a solution to copy right protection [18].Another protocol attack is the copy attack instead of destroying the watermark, the copy attack estimates watermark from watermarked data and copies to some other data target data.

## IV. ANALYSIS OF WATERMARKING ATTACK SCENARIO

The algorithm for embedding and extraction of is developed in MATLAB version7 [11] watermarking. The biometric fingerprint image used for the experiment purpose is generated using the SfinGe tool. The watermark used for the experiment is a facial image of size 64X64 and fingerprint image used is of siz3 512X512 gray scales 8 bit. The fingerprint image is the host image, which on which invisible watermarking is embedded using the facial image as the watermark. The figure6 shows the result of the experiment, the perceptibility of watermarked image is not lost due to watermarking it resemble the original fingerprint image. The extracted watermark is also is in good shape. The PSNR value of the watermarked image is around 40 dB.

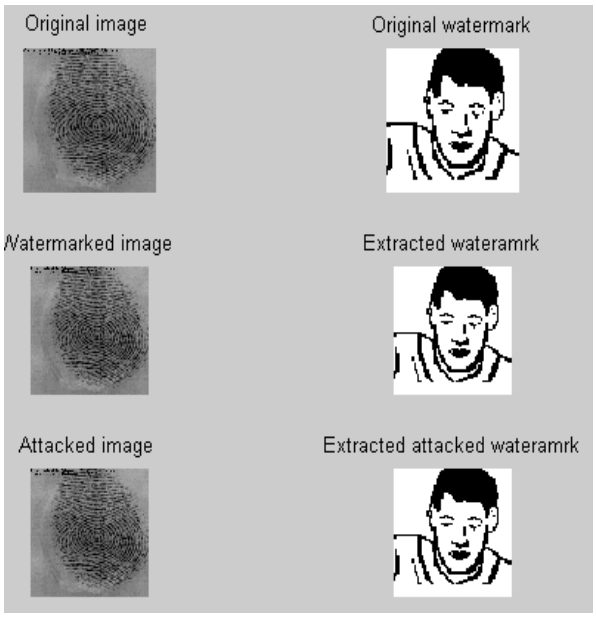


Figure 2: Extraction of the watermark without any attack

### *A. ttacks on the watermarked finger print images:*

To prove the robustness against common signal processing attacks on the watermark embedding process different attacks are initiated on the watermarked fingerprint images. Resistant to rotation is an important factor for watermarked fingerprint images in the real life situation.

#### *a. Rotational attack on the watermarked finger print image:*

The finger print host image watermarked with "logo mark "as watermark is attacked with geometric rotation of 10 degrees in the anti-clockwise direction and then using extraction algorithm the watermark is recovered.

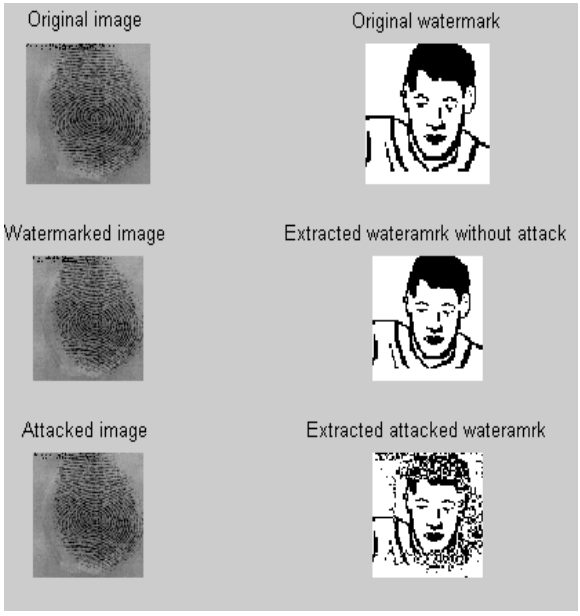


Figure 3: Extraction of the watermark from a rotational attacked finger print image

### *B. Peak signal to noise ratio (PSNR):*

Peak signal to noise ratio (PSNR) [2] is the measure of the image fidelity. The experiment is carried out on fingerprint images with attacks at various degrees of rotation of the images the results shows the graph as depicted in the in figure 8. At lower values of rotation of the watermarked image the image is less corrupted and PSNR value is high and with the addition of more rotation of the watermarked image the quality of the image deteriorates very much and PSNR value decreases. PSNR measures are estimates of the quality of a distorted image compared with an original reference image. PSNR gives a single number which reflects the quality of the distorted image [ ]. Suppose we have a source image w (i, j) that contains M by N pixels and a distorted image w' (i, j) where w' is corrupted with noise or distortion. Error metrics are computed on the luminance signal only so the pixel values range between black (0) and white (255). PSNR in decibels (dB) is computed by using MSE (mean squared error).For two M×N monochrome images w and w' where one of the images is considered a noisy approximation of the other is defined as in the equation (2) and (3):

$$MSE = \frac{\sum\sum [w(i,j) - w'(i,j)]^2}{M*N} \quad \text{---------- (1)}$$

The PSNR in dB value is defined as:

PSNR=20 log10 (Maxi / RMSE) ---------- (2)

Here, Maxi is the maximum possible pixel value of the image i.e. 255. Peak signal to noise ratio (PSNR) is the measure of the image fidelity.

RMSE=Root mean square of the error=√MSE

PSNR=20 log10 (255 / RMSE) ------ (3)

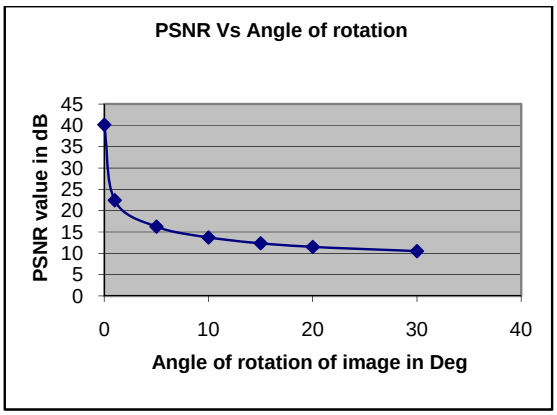


Figure 4: PSNR values vs. degrees of rotation of the fingerprint images

### *C. Correlation coefficient values rotational attack of the fingerprint images:*

Correlation coefficient [1] is the similarity between the original image and the watermarked image. Correlation coefficient is 1.0 for no rotational attack on the images as rotational attacked is added to the finger print images. Correlation coefficient starts deteriorating. The mathematical formula for computing correlation coefficient r is:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \quad \text{----- (4)}$$

Where n is the number of pairs of images (x and y). The value of r is such that -1 < r < +1. The + and – signs are used for positive linear correlations and negative linear correlations, respectively.

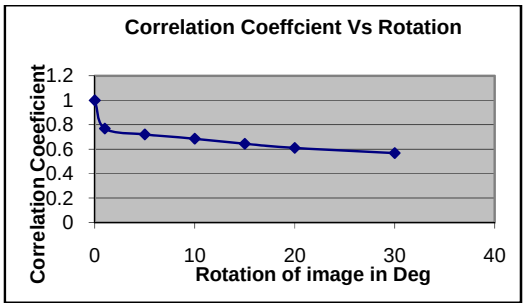


Figure5: Correlation coefficient values vs. degrees of rotation of the fingerprint images

### *D. Common signal processing attacks on the watermarked images:*

To prove the robustness against common signal processing attacks on the watermark embedding process different attacks are initiated on the watermarked fingerprint images [2]. To assess the image quality after different attacks the PSNR and normalized correlation coefficient (r) values are calculated. The high PSNR and r values indicate that the effect of attack on the watermarked image is less.

#### *a. 4.4.1 Median Filtering attack:*

Median Filtering is an image processing technique which is used for reducing the presence of noise in an image, hence enhancing the image quality. The Median Filter is the best-known order-statistic filter, which is a type of non-linear filter. Median filters are based on ordering or ranking the pixels contained in the image area covered within the filter, and then as its name implies, replacing the value center pixel by the median of the intensity values in the neighborhood of that pixel.

Median filters are quite popular because for certain types of random noise, they provide excellent noise-reduction capabilities, resulting in considerably less blurring than linear smoothing filters of a similar size. Median filters are particularly effective in removing the presence of impulse noise, called salt-and-pepper noise because of its appearance as white and black dots superimposed on an image.

#### *b. Low frequency filtering attack:*

Wiener2 low pass-filters an intensity image that has been degraded by constant power additive noise. Wiener2 uses a pixel wise adaptive Wiener method based on statistics estimated from a local neighborhood of each pixel.

#### *c. Gaussian noise attack:*

The host image watermarked with "logo "as watermark is corrupted with Gaussian noise with $\mu$=0.0 and $\sigma$=0.005. Then the watermark is recovered from the noisy watermarked image using the watermark extraction algorithm. The recovered watermark is noisy but still recognizable.

#### *d. Image cropping attack:*

The host image watermarked with "logo "as watermark is cropped [in the portion 100:300,100:300]. Then the watermark is recovered from the cropped watermarked image using the watermark extraction algorithm. The recovered watermark is noisy but still recognizable.

#### *e. Image resize attack:*

The watermarked image is resized to half of its dimension (2:1). Then the watermark is recovered from the attacked watermarked image using the watermark extraction algorithm. The recovered watermark is bit noisy but still recognizable

Table1: Different attacks on Watermarked images and corresponding PSNR and correlation coefficients(r) values

| Host Image | Watermark | Watermarked image | Type of attack | Attacked Watermarked image | Extracted watermark | PSNR value between host image and watermarked image in dB | Correlation coefficient (r) value between original watermark and extracted watermark |
|---|---|---|---|---|---|---|---|
| Lena | | | No attack | | | 36.2849 | 1.000 |
| Lena | | | Median filter attack | | | 34.7635 | 0.9347 |
| Lena | | | Low frequency attack | | | 36.0838 | 0.9645 |
| Lena | | | Gaussian noise attack μ=0.0 σ=0.005 | | | 26.4849 | 0.6726 |
| Lena | | | 'salt & pepper', noise 0.01 | | | 25.0851 | 0.5423 |
| Lena | | | Image croping 200*200 | | | 13.2730 | 0.7663 |
| Lena | | | Image resize 1:0.5 | | | 28.5855 | 0.5715 |

## V. COUNTER MEASURES TO PREVENT WATERMARK ATTACKS

To make the watermarked image secured in the internet environment the scheme should be resistant to common attacks that it may face in the actual situation. Thus the selection of embedding scheme of watermarking, the key etc., are important factors [10][12]. The general countermeasure against the watermark attacks include the following measures

a. Embedding of content information and adjacent block information in the watermark;
b. Embedding of watermark in transform-invariant domain;
c. Use of non-invertible watermark embedding scheme
d. To Prevent unauthorized Embedding (for multimedia authentication Purpose) use cryptographic techniques such as encryption or digital signature.
e. To prevent against copy attack the watermark may be derived from the host data (image).

## VI. CONCLUSIONS

Though much research work has been taken place and progress has been made in this area of technology, still the digital watermarking on host images has some inherent vulnerability in term of security, robustness, capacity and fidelity. The vulnerabilities could be exploited by the attackers in the real life situation and can defeat the main

purpose of the scheme and may initiate collision attacks. The potential attacks on the scheme made it restricted for use by the common people as tool for protecting their Intellectual property right (IPR) and establishing ownership rights in the internet environment. The watermarking scheme is has some inherent limitation as stated below:

a. As yet no standard algorithm for watermark embedding and detection has been established and accepted universally.
b. Fraudulent user may exploit the weakness of the watermarking scheme and can establish as real owner of the digital content.
c. No framework is still established on how the watermarking scheme will operate so that people have much confidence to use scheme.
d. Though Digimark, Checkmark is doing some sort of services in this regard but as a whole, as in the case of cryptographic system no third party certification Agency (CA) which maintain the public key of the users, is not yet formed in the case of watermarking to check and the authenticity of the author/ originator of image and act as arbitrator in cases of any dispute. [8]

The requirement of watermarks is different for different applications; so are attacks on different on the watermarked images. Generally it is experienced and perceived that the attacks on watermarked images include unauthorized embedding, unauthorized detection, and unauthorized removal and system attack. Some representative attacks: Scrambling Attack; Synchronization Attack, Linear Filtering and Noise Removal; Copy Attack; Ambiguity Attack; Sensitivity/Gradient Attack and Collusion Attack. General countermeasures to prevent these attacks are embedding content information and adjacent block information in watermark; embed watermark in transform-invariant domain; use non-invertible watermark embedding etc.

The analysis of attacks on digital watermarks on images reveals that there are several practical problems associated which need to be explored and addressed by the watermark developer before watermarks can become a suitable and popular means for acceptance in the user community as to permanently embed proof of ownership of the image author.

## VII. REFERENCES


[1]. J. Cox, M.L. Miller, J.M.G. Linnartz, T. Kalker, “A Review of Watermarking Principles and Practices” in Digital Signal Processing for Multimedia Systems, K.K. Parhi, T. Nishitani, eds., New York, New York, Marcel Dekker, Inc., 1999, pp 461-482

[2]. J. Dugelay, S. Roche, “A Survey of Current Watermarking Techniques” in Information Techniques for Steganography and Digital Watermarking, S.C. Katzenbeisser et al., Eds. Northwood, MA: Artec House, Dec. 1999, pp 121-145

[3]. Chandra, D. V. S., Digital image watermarking using singular value decomposition, Proc. of the 45th Midwest Symposium on Circuits and Systems, vol.3, pp.264-267, 2002.

[4]. Tapas Bandyopadhyay, B. Bandyopadhyay, B N Chatterji, Image Security Enhancement Through Watermarking and Cryptographic Measures, National conference : INDIACOM-2009 , New Delhi ,February 2009

[5]. Tapas Bandyopadhyay, Robust and secure watermarking for protecting rightful ownership, Recent Trends in Computer Technologies ( RTCT) Seminar organized by B P PodderInstitute of Technology and Management and CSI, 28 March 2009

[6]. F. Mintzer, G. W. Braudaway, M. Y. Yeung, "Effective and Ineffective Digital Watermarks," Proceedings of the IEEE international Conference on Signal Processing (JCIP'97), Vol. III, pp. 9-12, 1997.

[7]. G. W. Braudaway, "Protecting Publicly-Available Images with an Invisible Image Watermark," Proceedings of the IEEE International Conference on Signal Processing (ICIP '97), Vol. 1, pp. 524-527.

[8]. F. C. Mintzer, L. E. Boyle, et. al., "Toward online, worldwide access to Vatican Library materials," lBMJournal ofResearch and Development, Vol. 40, No. 2, pp. 146-149, 1996.

[9]. F. Mintzer, G. W. Braudaway, "A Family of Linear Filters for Image Sharpening," Proceedings of IS&T's Third Technical Symposium on Prepress, Proofing & Printing, pp. 14 1-143, 1993

[10]. Tapas Bandyopadhyay, B. Bandyopadhyay, B N Chatterji, Providing Security of Fingerprint images through Digital watermarking, National conference : INDIACOM-2010 , Jointly Organized by BharatiVidyapeeth’ Inst. of Computer application and Mgt, IETE, CSI, NewDelhi, pp 257-258,262 , February 2010

[11]. Matlab manual, Math work central

[12]. S. Craver, N. Memon, B. Yeo, and M. Yeung, “Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks, and Implications,” IEEE Journal on Selected Areas in Communications, vol. 16, no. 4, pp. 573-586,

[13]. P. Meerwald, A. Uhl, “A Survey of Wavelet-Domain Watermarking Algorithms” EI San Jose, CA, USA, 2001

[14]. Xie, L. and Arce, G.R. (2001) A class of authentication digital watermarks for secure multimedia communication IEEE Transactions on Image Processing, Vol.10, No.11, Pp.1754-1764.

[15]. Xie, Z., Wang, S., Gan, L., Zhang, L. and Shu, Z. (2008) Content Based Image Watermarking in the Ridgelet Domain, International Symposium on Electronic Commerce and Security, Pp.877-881.

[16]. Ye, J. and Tan, G. (2008) An Improved Digital Watermarking Algorithm for Meaningful Image, International Conference on Computer Science and Software Engineering, vol. 2, Pp.822-825

[17]. Wolfgang, R.B. and Delp, E.J. (1996) A watermark for digital images, Proc. 1996 Int. Conference on Image Processing, Lausanne, Switzerland, vol. 3, pp. 219–222

[18]. Qi , X. and Qi, J. (2007) A robust content-based digital image watermarking scheme, Signal Processing, Elsevier, Vol. 87, Issue 6, Pp. 1264-1280.