Discrete Wavelet Transform based Watermarking for Image Content Authentication

Obaid Ur-Rehman and Natasa Zivic
Chair for Data Communications Systems, University of Siegen, Hoelderlinstrasse 3, 57076 Siegen, Germany
{obaid.ur-rehman, natasa.zivic}@uni-siegen.de

Keywords: Watermarking, Content Authentication, Wavelet Decomposition, Security Analysis.

Abstract: A watermarking scheme based on discrete wavelet transform for content based image authentication is proposed in this paper. The proposed scheme is tolerant to minor modifications which could be due to legitimate image processing operations. The tolerance is obtained by protecting the low frequency data of the wavelet transform using approximate message authentication codes. Major modifications in the image content are identified as forgery attacks. Simulation results are given for unintentional modifications, such as channel noise, and for intentional modifications such as the object insertion and deletion. Security analysis is given at the end to analyze the security strength of the proposed image authentication scheme.

1 INTRODUCTION

With the rapid growth of Internet and communications technologies, and the widespread availability of multimedia generation and editing tools, images can be easily generated and shared over the Internet. However, due to this advancement, image content can be conveniently edited and reconstructed. As a consequence, the significance of the techniques for image integrity verification and content authentication is ever increasing. A digital watermark is typically a visible or invisible signature inserted inside the image to proof its authenticity or ownership at a later stage. Digital watermarking (Cox, et al., 2007), as opposed to digital signatures, do not require extra bandwidth for transmission and are designed to be prone to minor modifications in the image data. However, they should be sensitive to modifications in the image content. Digital signatures, on the other hand, are very sensitive to any modifications in the image data (or content). With the standard authentication mechanisms, a well known phenomenon called Avalanche Effect (Fiestel, 1973) will result in failed authentication even in the presence of a single bit error. There is a new class of authentication mechanisms emerging recently called the soft authentication mechanisms (Ur-Rehman, 2013) or noise tolerant authentication mechanisms. These mechanisms are designed to be tolerant to minor modifications in the data, i.e., the authentication will succeed even if the data protected by these mechanisms is a little different than the data on which the authentication tag was computed. A watermarking scheme for image authentication is proposed in this paper which is based on the approximate message authentication code (AMAC) (Graveman and Fu, 1999). AMAC is tolerant to minor changes in data, whereas the standard message authentication code (MAC) does not tolerate any modification of the data.

This paper is organized as follows. Section II discusses some related work. Section III discusses the building blocks of the proposed watermarking scheme. Section IV presents the watermark generation, embedding and the watermark extraction mechanisms. Simulation results are given in Section V. Security analysis of the proposed scheme is presented in Section VI. Finally, the paper is concluded in Section VII.

2 RELATED WORK

Amongst the many methods for noise tolerant data authentication, approximate message authentication code (AMAC) is used in this work. Other techniques for noise tolerant data authentication include, noise tolerant message authentication code (NTMAC) (Boncelet, 2006) and soft input decryption (Zivic, 2008). AMAC is based on majority logic, in which

Ur-Rehman, O. and Zivic, N.
Discrete Wavelet Transform based Watermarking for Image Content Authentication.
DOI: 10.5220/0006232306200625
In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, pages 620-625
ISBN: 978-989-758-222-6

the authenticator tag is generated by arranging the data in rows and columns. Then after XORing with pseudorandom bits, the majority logic is used to obtain the authentication tag. In AIMAC (Graveman, Xe and Arce, 2000), which is a variation of AMAC, the AMAC is adapted to image data, such that it is tolerant to minor changes in the image data but still able to differentiate intentional forgeries. The results in the presence of image modification scenarios including JPEG compression, image forgery and additive Gaussian noise are given in (Graveman, Xe and Arce, 2000).

The NTMAC algorithm is also tolerant to slight modifications in data. The idea is based on splitting the data into blocks, calculating standard MAC on each block and retaining a portion of the whole MAC for each block. This portion is used to detect changes in the block. The concept of partitions is used to introduce tolerance. Again certain variations and improvements on NTMAC have been proposed in literature. These include weighted noise tolerant message authentication code (WNTMAC) (Ur-Rehman, et al., 2011). WNTMAC is based on NTMAC but introduces the concept of weights to differentiate the relatively more important parts of data from the lesser important parts. EC-WNTMAC is an extension of WNTMAC, where the error localization and correction capability is introduced aside from error tolerance. However, all of the above mentioned approaches are based on image data authentication. They need to be used together with image features for authentication of image content.

NTMAC was used for image content authentication in (Ur-Rehman and Zivic, 2012). Features of the image were generated based on discrete cosine transform (DCT) and they were protected using NTMAC. If error correction is desired in addition to authentication, then error correcting codes have been used together with content authentication. This helps in error localization and correction. In (Lee and Won, 2000), Reed-Solomon (RS) codes are used to calculate parity symbols for each row and column of an image. These parity symbols are embedded as a watermark in the two least significant bit (LSB) planes of the image. RS decoder is used to "correct" the modifications in the watermarked image. In (Tabatabaei, et al., 2015), a two phase authentication scheme is proposed which performs image authentication in two stages. In one stage, the error correcting codes are used to (partially) correct the image and in the second stage, a tolerant authentication is performed. The threshold is totally flexible and can be adjusted to achieve the desired level of flexibility.

Amongst the other interesting techniques, two methods for self-embedding an image in itself were proposed in (Fridrich and Goljan, 1999). This helps in recovering those portions of the image which are somehow damaged, e.g., through cropping, or tampering. In the first method, the 8 × 8 blocks of an image are transformed into frequency domain using discrete cosine transform (DCT) and the coefficients are embedded in the least significant bits of other distant blocks. This method has a good quality of reconstruction but it is very fragile. The second method is based on the principle similar to differential encoding where a circular shift of the original image with decreased colour depth is embedded into the original image.

3 BUILDING BLOCKS OF THE PROPOSED WATERMARKING SCHEME

3.1 Digital Watermarking

Digital watermarking is a technique of covertly embedding digital data with secret information that can be extracted by the recipient (Zivic, 2015). The watermark should be unique, so that it can be later used for authentication. Additionally, the watermark should also be complex making it difficult for an attacker to extract and damage or replace it. An ideal watermark should be such that extracting it damages the cover object. Applications of digital watermarking include owner identification, copyright protection and content authentication, to name a few. Watermarks are typically based on image features. The features of the cover image are extracted at first as,

$$\boldsymbol{f} = \boldsymbol{Feature}(\boldsymbol{Image}) \quad (1)$$

where *Feature(·)* is a feature extraction function, applied on the cover image *Image* to obtain the image feature *f*. The features uniquely identify the cover image and two different images will have completely different features. However, images with the same content as the cover image will have more or less the same features. The image feature, *f*, is then used to generate a watermark, by protecting it using a secret key, *k*, as,

$$\boldsymbol{w} = \boldsymbol{GenerateWatermark}(\boldsymbol{f}, \boldsymbol{k}) \quad (2)$$

where w is the watermark and *GenerateWatermark(·)* is a watermark generation function. Only the intended recipient(s) with the shared key can extract the watermark and authenticate the image. The watermark is then embedded into the cover image. Two methods are typically used for watermark embedding, i.e., either in the spatial domain or in the frequency domain. The watermark can be extracted at any later stage to verify the authenticity of the protected data.

3.2 Discrete Wavelet Transform

Discrete wavelet transform (DWT) is used to decompose an image hierarchically. Wavelet transform decomposes the image into band limited, low and high frequency components, which can be reassembled to reconstruct the original image. A DWT operation decomposes an image into four components represented as LL, LH, HL and HH and as shown in Fig. 1. Where L represents applying a low pass operation and H represents applying a high pass operation. Here LL is the low resolution approximation image and it closely resembles the original image. The other sub bands, LH, HL, and HH represent other details such as edges etc. An example DWT of the Lena image is shown in Fig. 2.

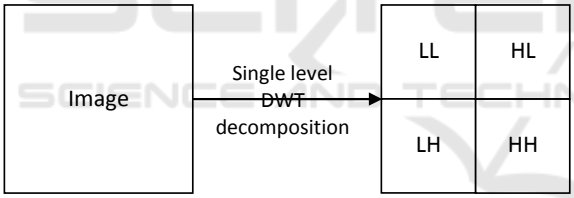

Figure 1: Single level DWT decomposition.

3.3 Approximate Message Authentication Code

As already said, AMAC is an algorithm from the class of noise tolerant authentication algorithms, designed to tolerate minor modifications in a message/image. This is different from the standard MAC algorithms, which do not tolerate even a single bit modification. The threshold on the acceptable number of bit modifications is adaptable and the tolerance exhibited by AMAC is due to the majority logic. AMAC tag generation on a message M is shown in Fig. 3, where L is the tag length and R and S are positive integers. R is usually chosen to be equal to S for simplicity (Graveman and Fu, 1999). A pseudorandom number generator (PRNG) is used to generate a stream of pseudorandom bits in the AMAC using a secret key, k_1, shared between the sender and the intended receiver. As long as the bit changes in the data are below the threshold, the data is declared authentic. If the changes exceed the threshold, the data is declared unauthentic.

Figure 2: The single level DWT decomposition of Lena image.

4 WATERMARK GENERATION, EMBEDDING AND EXTRACTION

4.1 Watermark Generation

A source image is taken and DWT is computed on it. The LL sub band of the DWT is taken and passed through the AMAC algorithm. The AMAC tag is taken as the watermark of the image. If there are minor changes in the image, the LL sub band will not change much and thus the AMAC will remain the same. For changes beyond a threshold, such as in case of forgery attacks, e.g., object insertion or object removal, the AMAC will change. The threshold is adjustable as discussed in the section on AMAC. The length of AMAC tag is chosen to be 256 bits. The watermark generation for an example Lena image is shown in Fig. 4.

4.2 Watermark Embedding

The watermark is embedded in the cover image. In this work, the watermark is self embedded in the source image. The source image is split into 8 × 8 pixel non overlapping blocks. The length of the AMAC tag is 256 bits, which is split into 32 sub-AMACs of 8 bits each. One sub-AMAC is taken at a

time and inserted into the LSBs of the first 8 pixel values of the next image block which is obtained via a secret permutation. The next image block is chosen at random, using a secret key, k_2, as the seed value. Thus the AMAC tag is scrambled in the LSB of image blocks. This makes it hard for an attacker, to extract and replace the watermark without the knowledge of the secret key.

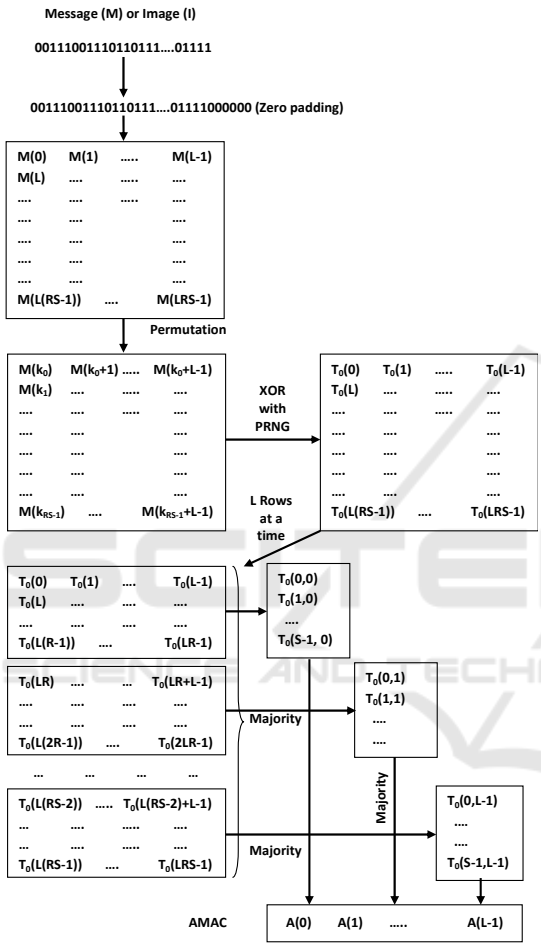

Figure 3: AMAC tag generation (Graveman and Fu, 1999).

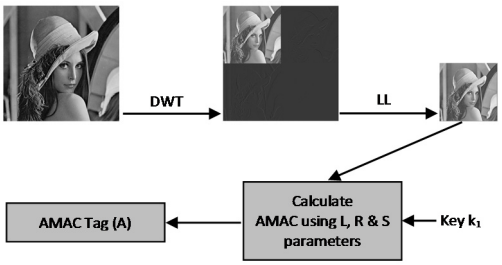

Figure 4: Watermark generation.

4.3 Watermark Extraction

When the authenticity of the image has to be proven, the watermark is extracted back from the image. The watermarked image is taken and split into 8 × 8 pixel non overlapping blocks. The LSBs of the first 8 pixel values of each next block is taken and appended to the existing watermark to obtain the complete watermark (the AMAC tag). The next block is chosen again using the pseudorandom permutation based on the shared secret key, k_2, to obtain the same sequence as obtained in the watermark embedding procedure.

4.4 Image Authentication

An image is verified by comparing the extracted watermark with the recomputed watermark. As the watermark is embedded in the spatial domain, a part of the cover image which is the source image as well, is distorted. However, since AMAC is tolerant to modifications below the chosen threshold, the authentication will succeed even if there are other deviations from the original.

5 SIMULATION RESULTS

Resolution of images used in these simulations is 256 × 256 pixels. The input image is first converted to a grayscale image before being process further. Single level DWT transform is applied on the grayscale image. The length of AMAC is chosen to be 256 bits and the length of a sub-AMAC is chosen to be 8-bits. Simulations results are given in this section for authentication in the presence of intentional and unintentional modifications. Results for unintentional modifications are based on "Salt & Pepper" noise of varying magnitudes. Object insertion is performed to test the proposed method in the presence of intentional noise / forgery attacks.

Fig. 5 shows the 4 sub-bands of the single level DWT decomposition of Lena image in the presence of "Salt & Pepper" noise of magnitude 0.001. As it can also be observed from Fig. 5, the LL sub-band of the Lena image in the presence of "Salt & Pepper" noise resembles the LL sub-band of the original Lena image. The Hamming distance between the two is 215. However, their AMAC tags are similar based on the chosen value of threshold to allow for bit differences of up to 300 bits in the data. Therefore the Lena image in the presence of "Salt & Pepper" noise passes the authentication test of the proposed method and is declared authentic.

Figure 5: Authentication in the presence of "Salt & Pepper" noise of magnitude 0.001.

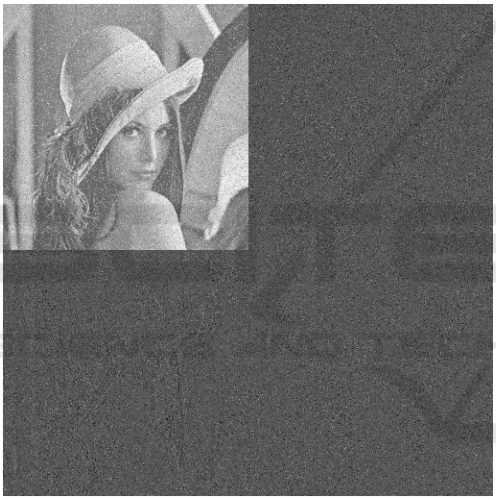

Figure 6: Authentication in the presence of "Salt & Pepper" noise of magnitude 0.1.

However, if the noise level exceeds the threshold, then the image is declared unauthentic. In Fig. 6, the "Salt & Pepper" noise of magnitude 0.1 results in a Hamming distance of 22443 between the LL sub bands of the original and the modified images. It can be noticed from the figure that the other sub bands are also severely affected by the high magnitude of noise, though they are not used in the authentication. Thus the authentication test fails as the AMAC tags are different for both the images. The test case of forged Lena image is shown in Fig. 7, with extra hair on the forehead. The LL band of the forged image has a Hamming distance of 16315 with the LL band of the original Lena image of similar dimension. Thus the AMAC tags are different, resulting in a failed authentication.

Figure 7: Authentication in the presence of forgery attack, with extra hair on the forehead.

6 SECURITY ANALYSIS

The security analysis of AMAC is given in (Onien, Safavi-Naini and Nickolas, 2011), where it is proven that if Hamming distance is used for distance measurement, then it might not be secure for large messages. The general security analysis of the proposed authentication scheme can be done by considering the key recovery and substitution attacks. In key recovery attack, the secret key of the scheme is disclosed using a sufficient number of authenticated image-hash pairs. An attacker can then use the recovered secret key to generate a watermark of his own image to deceive the receiver. In the substitution attack, the attacker tries to substitute a valid image and its tag with another authentic image and its tag. This attack is successful when the substituted image is perceptually different than the original image whereas the difference between their watermarks or tags is below the threshold value.

6.1 Key Recovery Attack

The watermark generation and insertion uses two secret keys in order to generation and embed the watermark. An attacker must recover two secret keys, one for watermark generation and another one for watermark embedding. In AMAC, the reshaped matrix of size $R \times L \times S$ bits is used, which means the attack complexity is about $2^{R \times L \times S}$ function (tag generation and verification) operations, which is

very high complexity even for small images, such as 56 x 56 pixels.

6.2 Substitution Attack

An attacker can execute the substitution attack in two steps, a forgery attack on the AMAC and then an attack on the watermark embedding. The possibility for an attacker to pass the first step can be calculated as follows. Let T be the threshold value below which the difference between the AMAC tags is acceptable and let t indicates the threshold for difference between DWT's LL sub-band tolerance. The probability (P_t) of changes in the "majority" selection round of the AMAC is calculated in (Onien, Safavi-Naini and Nickolas, 2011) as,

$$P_t = \frac{1}{2^L \sum_{i=0}^{t} \binom{L}{i}} \sum_{i=0}^{t} \sum_{j=\left\lceil \frac{t+1-2i}{2} \right\rceil}^{\left\lfloor \frac{t-1}{2} \right\rfloor} \sum_{k=0}^{\left\lfloor \frac{2i-2j-t-1}{4} \right\rfloor} \binom{t}{j} \binom{j}{k} \binom{t-j}{i-k} \quad (3)$$

Based on Pt, the probability of deceiving the attacker (P_D) is calculated as,

$$P_D = \sum_{i=0}^{T} \binom{L}{i} P_t^i (1-P_t)^{L-i} \quad (4)$$

P_D can be decreased by increasing the length of AMAC tag.

7 CONCLUSIONS

The paper proposes a watermarking scheme for content based image authentication. The scheme consists of generating the watermark based on image features using discrete wavelet transform and protecting them using the noise tolerant AMAC algorithm. Simulation results show the noise tolerant authentication capability for unintentional modifications, such as through channel noise. However, intentional modifications, such as forgery attacks can be recognized using the proposed watermarking scheme.

REFERENCES

Cox, I, Miller, M, Bloom, J, Fridrich, J and Kalker, T, 2007, *Digital Watermarking and Steganography,* Morgan Kaufmann.

Fiestel, H 1973, *'Cryptography and Computer Privacy', Scientific American,* vol. 228, no. 5, pp. 15-23.

Ur-Rehman, O 2013, *Applications of iterative soft decision decoding,* Aachen, Shaker Verlag.

Graveman, R and Fu, K 1999, *'Approximate message authentication codes',* Proceedings of 3rd Fed. lab Symposium on Advanced Telecommunications / Information Distribution.

Boncelet, C 2006, *'The NTMAC for authentication of noisy messages', IEEE Transactions on Information Forensics and Security,* vol. 1, no. 1, pp. 35-42.

Zivic, N 2008, *Joint Channel Coding and Cryptography,* Aachen, Shaker Verlag.

Gravemen, R, Xie, L and Arce, GR 2000, *'Approximate image message authentication codes', Proceedings of 4th Annual Symposium on Advanced Telecommunications and Information Distribution Research Program*.

Ur-Rehman, O, Zivic, N, Tabatabaei, AE and Ruland, C 2011, *'Error Correcting and Weighted Noise Tolerant Message Authentication Codes', 5th International Conference on Signal Processing and Communication Systems,* Hawaii, 12-14 December.

Ur-Rehman, O and Zivic, N 2012, *'Noise Tolerant Image Authentication with Error Localization and Correction', 50th Annual Allerton Conference on Communication, Control and Computing,* Monticello, Illinois, 1-5 October.

Lee, J and Won, CS 2000, *'A Watermarking sequence using parities of error control coding for image authentication and correction', IEEE Transactions on Consumer Electronics,* vol. 46, no. 2, pp. 313-317.

Tabatabaei, AE, Ur-Rehman, O, Zivic, N and Ruland, C 2015, *'Secure and Robust Two-phase Image Authentication', IEEE Transactions on Multimedia,* vol. 17, no. 7, pp. 945-956.

Fridrich, J and Goljan, M 1999, *'Images with self-correcting capabilities', Proceedings of the IEEE International Conference on Image Processing,* pp. Kobe, 25-28 October.

Zivic, N ed. 2015, *Robust Image Authentication in the Presence of Noise,* New York, Springer.

Onien, D, Safavi-Naini, R and Nickolas, P 2011, *'Breaking and repairing an approximate message authentication scheme', Discrete Mathematics, Algorithms and Applications,* vol. 3, no. 3, pp. 393-412.