Toward Safe and Secure Electronic Documents of E-Governments: Generating Authentic Documents using Image Processing Techniques

Abdullah AL- Shraideh
Technical investigator
Jordan Preventive security,
Amman, Jordan

Suliman Bni Ahmad
Assistant Professor
Al-balqa University, salt,
Jordan

Audeh Bni Ahmad
Assistant Professor
Al-Albayt University, Mafraq,
Jordan

ABSTRACT
E-governmental applications have to find a solution for efficient and effective mechanism for authenticating and signing official documents. We propose and implement a mechanism for issuing secure electronic authentic governmental documents using watermark technology.

The proposed mechanism constructs a watermark from domain name and MAC address of the governmental institution of interest. The watermark is augmented into the governmental document that is available in the form of image. This new mechanism promises allows the usage of watermarking to guarantee the authenticity and integrity of official documents. It can discover the change in these documents and the place where this change occurred. Proven by our experimental results, the proposed mechanism is secure and robust as the watermark will be embedded in multiple places inside the document where sensitive data is located, this way we can significantly reduce the probability of successful fraud of electronic documents.

Keywords
E-Government, Watermarking, Electronic documents

1. INTRODUCTION
By August 2012, UK expects to replace paper documents with web services in many aspects so that UK citizens will be able to access government web-based services using their bank smart cards [22]. Such. Providing authentic government documents is thus of vital necessity.

Computerized systems are now common place. Nowadays, Internet services make exchanging electronic documents in the form of images to be more easy and risky at the same time. This is because digital images can be fraud more easily than the non electronic documents. Consequently, the need for protecting digital image became more important these days.

Digital watermarking can offer a trusted tool for signing official documents that are in the form of images. This tool, i.e. digital watermarking, is strongly linked to the original paper watermarks that appeared at the end of the 13th century [2].

Nevertheless, copyright protection of electronic images is much more difficult than paper-based documents.

The main idea behind watermarking is hiding information inside an image. Information hiding is a general term includes many problems related in embedding messages in content, here the term hiding refer to one of two things, firstly making information unrealizable or keep existence of the information secret [2].

Steganography is one of information hiding application and the word steganography is derived from the Greek word steganos, which means "covered". In computer science, this means hiding information in other information. In other words, steganography is interested in keeping existence of data secret and no body can view the message expect the sender and the receiver [1].

Watermarking is nothing but another application of data hiding. Watermarking is interested in making the information imperceptible. Our discussion will focus on image watermarking which is interested in protecting the content of image from change. An image watermarking technique has to be robust against possible attacks. Notice that we are not addressing the technical modifications that may occur to documents during transmission and storage, such as format conversion, compression, or digital-to-analogy conversion. Rather, watermarking is always associated to the digital object (image in our context) to be protected [1, 2, 3].

2. Types of Watermarking
Digital image watermarking can be classified into different types depending on its application. Next we will show the most popular types of digital watermarking according to its application [2].

Visible or localized watermarks: can be logos or overlay images in the field of image or video watermarking.

Invisible watermark: Most of the current research focuses on invisible watermark in which added data can't be perceived than visible, Invisible watermark increases the likelihood of prosecution after the theft has occurred. These watermarks should therefore not be detectable to attackers, otherwise they would try to remove it, and however they should be easily detectable by owners.

Public watermark: in this type of watermarking watermark can be read or retrieved by anyone using the specialized algorithm. In this sense, public watermarks are not secure.

Private watermark: this type of watermarking require that the original be used as a basis of encryption so its known as secure watermarks. Because they

Robust watermarks: are designed to resist against heterogeneous manipulations. All applications presupposing security of the watermarking systems require this type of watermark

Fragile watermark: On the contrary of robust watermarks fragile watermarking is deigned with very low level of robustness. Such watermarks can be destroyed by data

manipulation, watermarking in this form looks like steganographic methods which focus in hiding message so that this type of watermarking can be used to check the integrity of objects.

3. Applications of Digital Image Watermarking

When watermarking was introduced in 1954 by Hembrooke [2], the main application for it was the copyright protection. However, because of the efficiency of watermarking, more applications for watermarking have emerged. In this section we will describe the most common applications of watermarking

3.1 Fingerprinting for Traitor Tracking

Digital, in general, watermarking is concerned with the owner of the image. In this application of watermarking, we are interested in the *customer* of the image rather than the owner. This specific application of watermarking can help in identifying a single distributed copy of the data [1] and in tracing the customer who redistributes contents illegally.

3.2 Covert communication

One of the earliest applications of data hiding is sending secret message so that in general public will not have access to the content of this message. This application implies that the agents who are concerned augment secret messages to their agency. In case that an unauthorized person get the data that contains the message, that person should not be able to retrieve the secret message

3.3 Authentication

It became easier to fraud digital work these days. This can cause many problems especially with government-related applications [3]. A simple example for that is the police evidence that proves that a given person has no criminal history. Such police document can be provided in a digital format in the form of an image. The content of this image can be illegally modified using image processing software. A mechanism to show the tamper in digital working is thus vitally required. One possible mechanism is to use fragile watermarks authentication throw creating digital signature for the image in hand. This signature is essentially an *encrypted summary of the image* [3]. An asymmetric key encryption algorithm can be used so that the key required to encrypt the signature is different from that required to decrypt it. Only the authorized source of messages knows the key required for creating signatures [3].

3.4 Watermarking for Copyright Protection

Copyright protection is the most common known application for watermarking. The objective is to embed information about the source, and thus typically the copyright owner, of the data in order to prevent other parties from claiming the copyright on the data [2].

4. Digital watermarking property

An image watermarking technique must achieve a set of conditions such as: (i) robustness, (ii) tamper resistance, (iii) fidelity, (iv) computational cost, and (v) false positive rate. Next we explain each of these characteristics in a more detailed manner. In general, a good image watermarking technique is the one that adds watermarks that are difficult to notice, resistive to distortions, and resists malicious attacks [4].

Robustness: Robustness refers to the ability to detect the watermark after common signal processing operations [3]. In other words digital watermarking should be strong enough to against modifications and/or malicious attacks. For these reasons robustness is probably the most important requirements that should be taken in mind while we are building digital watermarking [2].

Embedding Effectiveness: The embedding effectiveness of a watermarking system measures the probability of error when the detector is applied immediately after embedding [3].

Security: Security of watermarking refers to its ability to face any attack like unauthorized removal, unauthorized embedding, or unauthorized detection [3]. Hidden information must be exclusively be accessed by only authorized people. This can be achieved through the use of cryptographic keys [5].

Fidelity: Watermarking must be done in a perfectly way which make the added watermark difficult to notice. In other words, the difference between original and the image with the watermark should be minute.

4.1 Watermark extraction and watermark security and Keys

Watermarking system can be categorized into two types: the first type in which we are looking for embedding specific information or pattern and check the existence of the (known) information later on in the watermark recovery process [2]. Second type is embedding of arbitrary information.

Data security can not be classified into levels, the highest level which does not allow user to read or decode an embedded watermark nor can he detect if a given set of data contains a watermark. The second level permits any user to detect if data is watermarked, but the embedded information cannot be read without having the secret key. For that, watermarking must have a high degree of watermarking security to prevent any attempt to change in the watermarking or any trying to remove the watermarking.

5. Watermarking algorithms

In general, the digital image watermarking process can be summarized as follows: (i) generate the mark and (ii) embedding the mark, finally (iii) Creating the key file.

The detail of this process is as follows. The watermark *W* is embedded into a media message, which is defined as the host image *I*. The resulting image is the watermarked image *I'*, in the embedding process, secret key *K* is sometimes involved to generate a more secure watermark. The watermarked image *I'* is then transmitted along a communication channel. The watermark can be detected or extracted later by the receiver now if both the watermarking image and original image identical then the process is considered to be successful; otherwise, the process is considered failed [8].

we can see the process of embedding in more details in Figure 1 shows this process for embedding in detailed.

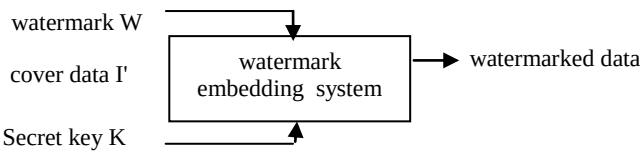

Fig 1: Watermark embedding process

The recovery for watermark system can be described in Figure 2. The input for this process is the watermarked image *I'* and the secret or public key and, depending on the method, the original data the output is the recovered watermark *W*.

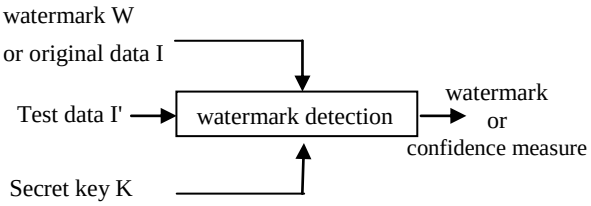

Fig 2: Watermark Extraction process.

6. Watermark attacks

In order to enjoy a strong defense it would be useful to recognize all possible attacking methods and understand what the danger that surrounds watermarking is. This should help building strong and robust watermarking. Next we summarize a number of image watermarking attacks.

In general, watermark attack can be classified into four categories:

Removal attack or active attack which aims at removing the watermark signal from image without attempting to break the security of watermark algorithm [9, 10], This is a big issue in copyright protection, fingerprinting or copy control.

Geometry attack: If the offender forgets to wear gloves, then he should be able to hide his fingerprint throw making some distortion or adding noise to his fingerprint. This form of attacks can work when attacking images, adding noise to images can cancel the watermark [9, 10].

Cryptographic attacks: in this type, the attacker tries to crack the security algorithm and looks for methods to remove the embedded watermarking information [9, 10].

Protocol attacks: in this type, the attacker uses defense as technique of attack so the attacker adds his watermark instead of the original to look as if he is the owner of the image. Protocol attacks aim at attacking the entire concept of the watermarking application [10].

Forgery attacks: in this type of attack the hacker aims at embedding a new, and a valid watermark rather than removing the already existing one.

7. Watermarking Techniques*:*

One prominent criterion to classify watermarking techniques is the domain in which the watermark signal is embedded. Based upon this criterion, watermarking techniques can be classified into two major classes [20].

- *Domain Spatial Domain Watermarking*: in which we insert watermark into a host image by changing the gray levels of some pixels in the host image. This approach has the advantages of low complexity and easy implementation, but the inserted information may be easily detected using computer analysis or could be easily attacked. The advantage of this approach is that it has very low computational complexity and usually can survive the cropping attack, which usually the frequency-domain watermarking techniques fail (explained next)[5].
- *Frequency Domain Watermarking*: In frequency domain watermarking, we can insert a watermark into frequency coefficients of the transformed image by DFT, DCT, or DWT. Because the frequency transforms usually decor relate spatial relationship of pixels, majority of the energy concentrates on the low-frequency components. When we embed the watermark into the low or middle frequencies, these changes will be distributed around the entire image [5].

7.1 DCT Domain Image Watermarking

The DCT is a mathematical transformation that takes a signal and transforms it from spatial domain into frequency domain [6]. Many digital image and video compression schemes use a block-based DCT, because this algorithm minimizes the amount of data needed to recreate a digitized image. In particular, JPEG and MPEG use the DCT to concentrate image information by removing spatial data redundancies in two-dimensional images [19], the general steps for any block based DCT algorithm can be shown below [6]:

1. partition the image into non-overlapping blocks of 8x8

2. after we partition the image Apply forward DCT to each of these blocks

3. some block selection criteria most be applied at the image

4. coefficient selection criteria

5. finally embed watermark by modifying the selected

7.2 Wavelet Transform (DWT)

DWT locally split the content of the image into low frequency and high frequency sub bands; most of the energy is concentrated in the low frequency sub band. In DWT, the image is decomposed into a set of band-limited components which can be reassembled to reconstruct the original image without error. The DWT separates an image into a lower resolution approximation image (LL) as well as horizontal (HL), vertical (LH) and diagonal (HH) detail components.

7.3 Least Significant Bit Modification

The basic idea of LSB is based on the substitution of the LSB of the carrier signal with the bit pattern from the watermark noise. The bits are embedded in certain representation values, such as pixels. The decoder in turn is able to retrieve the watermark if it has the knowledge of the representation values used for embedding the individual bits [17].

8. Related work – Protecting Original Images

Wei Lua and Co [11] presented a new robust method for protecting digital image. In this method the image is sub sampled in four images then the binary watermark image is permutated using chaotic map. Finally, the watermark is embedded into the DCT coefficients of two of the sub images which are selected according to a random coordinates sequence generated by a secret key. Watermark extraction is done by comparing DCT coefficients of the watermarked sub images and does not need original image.

Ying-Hualui and Co [13] also proposed extracting four sub images from the original image. After that, wavelet they applied transformation on the four sub images. The basic idea of this algorithm is based on comparing the sub image features to make a little damage to the original image; the resulting watermarking is a binary image. The key will be formed by comparing the wavelet middle-frequent coefficients of the sub images with the watermark. The power

of this scheme comes from that it unites the character of the host image and slightly modifies the host image. This technique is suitable in medical images and military images.

Kasmani, S. A. and Naghsh-Nilchi [14] combined DWT-DCT transformation to produce robust digital image watermarking algorithm. The algorithm works within the following procedure: embedding the binary watermarked image in certain sub-bands of a 3-level DWT transformed of a host image, then DCT transform of each selected DWT sub-band is computed and the PN-sequences of the watermark bits are embedded in the coefficients of the corresponding DCT middle frequencies. The extraction process is first pre-processed by sharpening and Laplassian of Gaussian filters. Then, the same approach of the embedding process is used to extract the DCT middle frequencies of each sub-band. Finally, correlation between mid-band coefficients and PN-sequences is calculated to determine watermarked bits.

Many algorithms and techniques have been developed to protect the digital image watermarking. These algorithms can be also attacked. Changjiang Zhang and co [15] presented a new approach to increase the security of image watermarking based on stationary wavelet transform. This algorithm encrypts the digital watermark, the encrypted watermarking is then transformed into one-dimensional row vector, and then the pixel value was sorted. The coefficient of the one-dimensional of primitive image of stationary wavelet transformation is sorted too. The sorted watermarking is then inserted to the sorted low frequency to turn it to two dimension data. The image is reconstructed with coefficients of high-frequency.

9. Electronic government

Because of the benefits offered by information technology governments started seriously in planning how to use these technologies in a way which reduces the daily routine and offer high quality of services to citizens electronically. With e-governmental electronic services, citizen will not be forced to review multiple government institutions, so that the only thing citizen should do is to communicate and interact with the

services offered by the government anytime anywhere through online services [21].

E-government also has many risks resulting from the use of the Internet [21]. These are:

Information Intercepting: this means that the e-information of e-governments or users information can be stolen.

- *Information Tampering:* the Internet is a rich environment for hackers who may try to illegally modify the data of governmental web sites
- *Services Denying*: in many cases any system may be hanged due to (i) the presence of a large number of users or (ii) the large requests placed by hackers. In all cases, this leads to deny a number of citizens to benefit from the services provided by the e-Government.
- *Information Faking*: It means hacker could pretend legal users or make false information to cheat other users. The main forms include pretending users to get illegal certifications or forging e-mails.

E-governments naturally lacks for confidentiality because it will be targeted by hackers and vandals. Consequently, citizens will trust e-government services especially when we are talking about issuing electronic documents. Thus, the need for checking the reality and validity of governmental documents is vitally necessary.

An example of such sensitive e-governmental documents is the *No-criminal record certificate*. This certificate is required to be issued when applying to a new job, to get vehicle driving license, or when getting visa. In Jordan, getting such a certificate using the regular paper-based approach is time and effort consuming. Two governmental institutions participate in issuing this certificate; the first one is *ministry of Judaic* and the second one is the *criminal information department* which belongs to the public security directorate.

Figure 3 shows the scenario of issuing the No-criminal record certificate while in figure 4 we can see how this task be done electronically.

10. The proposed method

In this proposed method we are looking to build a meaningful watermark by constructing it from the following institute-specific information: the MAC address of the computer issuing the watermark and domain name of the governmental organization that will issue the electronic document. We choose The MAC address as it is a unique identifier of the computer issuing the watermark. The domain name, on the other hand, should tell us which organization issued that electronic document.

The proposed image watermarking mechanism is composed of two basic processes which are: embedding the watermark and extract the watermark. Next, the two processes are explained.

10.1 Embedding process

Step 1: Read the image

Step 2: Get the *MAC address* of the computer that will issue the image as string

Step 2: Get the *domain name* of the computer that will issue the image as string

Step 3: Combine the MAC address and domain name

Step 4: Convert the combined string into image, the result is the watermark that we need to embed

Step 5: Embed the watermark in the original image in the most sensitive places inside the document to be protected

10.2 Extraction process

Step 1: Read watermarked image

Step 2: Extract the watermark from the original

Step 3: Run letter recognition function to convert extracted watermark in to text

Step 4: Convert coded MAC address to its original.

Step 5: If the extracted MAC address is legal and domain name mach the criteria then the document in hand is original, otherwise, the image has not been altered. The areas where the watermark is corrupted indicates the places they have been altered inside the image.

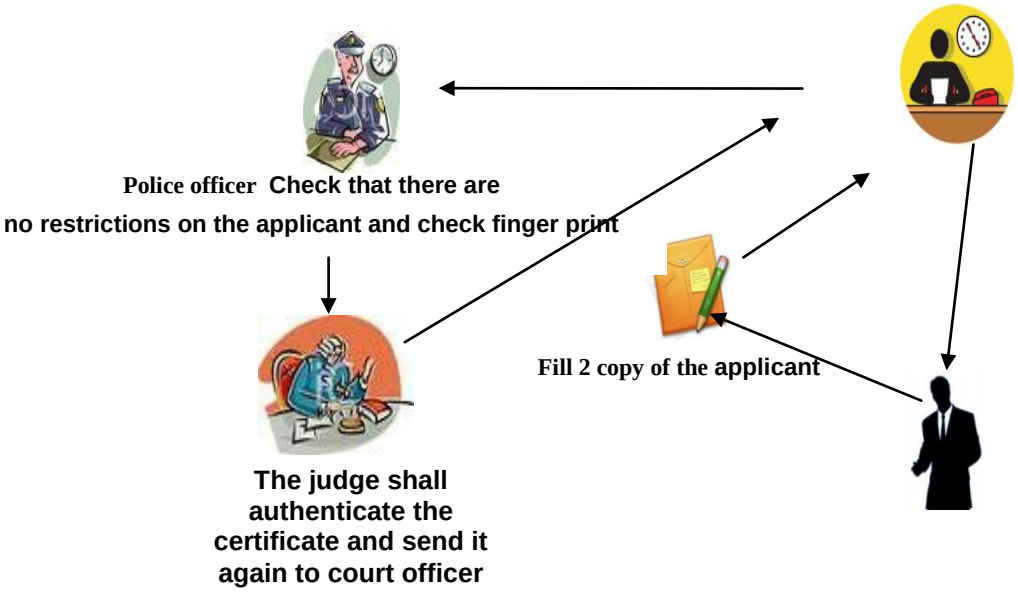

Fig 3:The process of issuing non criminal rescored certificate traditional

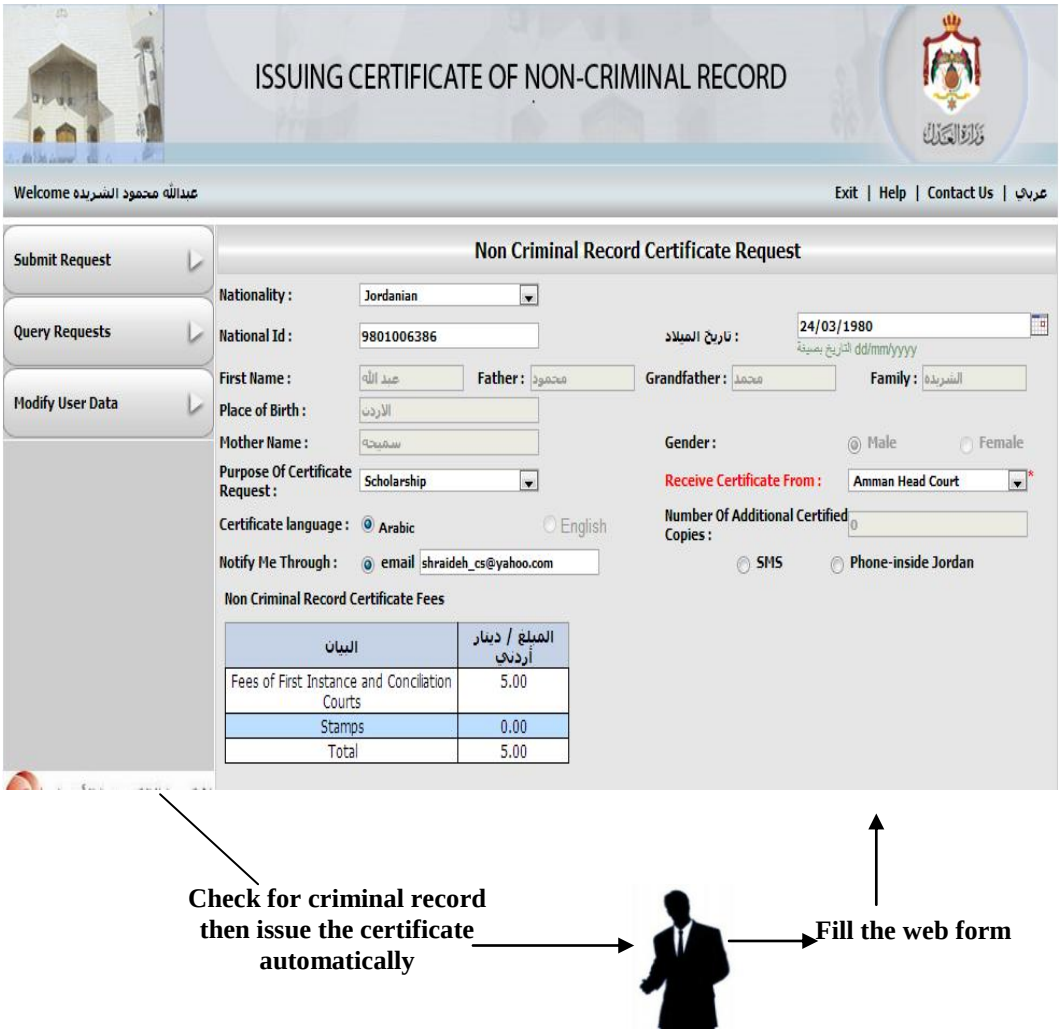

Fig4:The process of issuing non criminal rescored certificate electronically

11. Experimental results and observations

As mentioned earlier, the aim of this study is to propose a new mechanism or framework for the process of issuing secure and authentic electronic official documents. Next we demonstrate that our proposal gives good results. The proposed mechanism can detect the existence and the place where the change(s) has been made inside the document in hand.

Figure 5 show an original issued certificate along with the constructed watermark that is augmented in it. Notice that we can predicate that the attacker, who will change this official document to produce a fake one, will try to do the following changes in such electronic documents (just sample scenarios):

- The first option is that a person with the same exact name can use this document. In such case, the only change required is in the national ID number.
- Alternatively, the brother or sister of the certificate owner that *does have* criminal record to issue a fake one. In this case, he will replace his first name and his national identification number. Such change is looks minute.
- The last possible change, which would be the hardest, is to make new certificate with different name and ID number and add the judicial ministry stamp and add the judicial signature.

However, even with such minor and minimal change possible, our approach can identify the existence and the place of such change.

This was all the possible ways to modify this document. Our watermarking-based document protection mechanism is able to protect the electronic documents and discover the existence and the place of changes inside them, figure 6 shows the first case in which only the ID number is changed ,figure 7 shows the second case in which the first name and the ID number are changed, figure 7 shows the corrupted watermark that has been observed around the area that faced the change.

Fig 5: Original No-criminal record certificate (above) and the augmented watermark (below).

المملكة الأردنيـــــة الهاشمية
وزارة العدل
محكمة بداية عمان

الرقم : ٣٤٨٨/٤/١٧
الموافق:
التاريخ : ٢٠١١/٤/٢٦

وزارة العدل

شهـــــادة عدم محكومية

الاســـم : Abdullah mahmoud mohammd AL-shraideh
الجنسية : الأردنيـــــة
الرقم الوطني : 9801006366
الغاية من عدم المحكومية : التوظيف

المذكور أعلاه غير محكوم بجنحة أو جناية مخلة بالشرف والأخلاق العامة وبناءا على طلبه أعطي هذه الشهادة مصدقة حسب الأصول.

واقبلوا الاحترام

رئيس كتاب محكمة بداية عمان
أصادق على توقيع رئيس محكمة بداية عمان
رئيس محكمة بداية عمان

Fig 6: The fake No-criminal record certificate

المملكة الأردنيـــــة الهاشمية
وزارة العدل
محكمة بداية عمان

الرقم : ٣٤٨٨/٤/١٧
الموافق:
التاريخ : ٢٠١١/٤/٢٦

وزارة العدل

شهـــــادة عدم محكومية

الاســـم : mohammd mahmoud mohammad Al-Shraideh
الجنسية : الأردنيـــــة
الرقم الوطني : 9811006386
الغاية من عدم المحكومية : التوظيف

المذكور أعلاه غير محكوم بجنحة أو جناية مخلة بالشرف والأخلاق العامة وبناءا على طلبه أعطي هذه الشهادة مصدقة حسب الأصول.

واقبلوا الاحترام

رئيس كتاب محكمة بداية عمان
أصادق على توقيع رئيس محكمة بداية عمان
رئيس محكمة بداية عمان

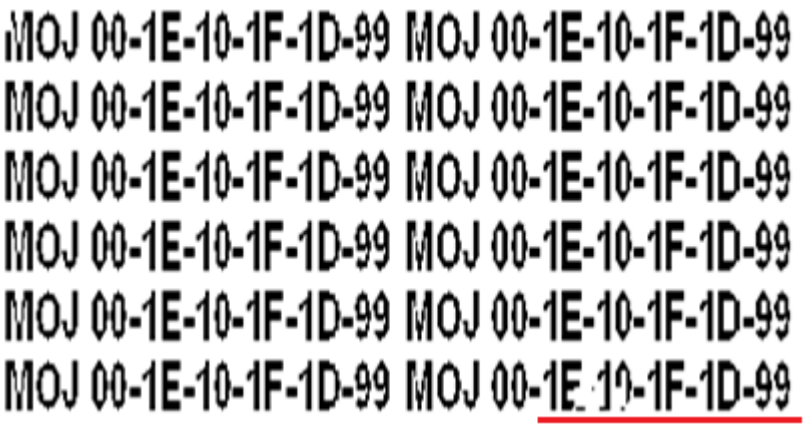

Fig7: The fraud document and the corresponding extracted (corrupted) watermark. The corruption occurred around the area that has been changed inside the image.

12. Conclusion

The rapid revolution in using the Internet-based applications puts us under many challenges. Overtime, it is becoming very difficult to discover crimes in cyber and electronic crime. For these reasons digital image watermarking have recently received high interest by governmental agencies to produce authentic and secure official documents. We proposed and demonstrated a robust watermarking-based document protection mechanism. This new mechanism promises allows the usage of watermarking to guarantee the authenticity and integrity of official documents. It can discover the change in these documents and the place where this change occurred. Proven by our experimental results, the proposed mechanism is secure and robust as the watermark will be embedded in multiple places inside the document where sensitive data is located, this way we can significantly reduce the probability of successful fraud of electronic documents.

or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

13. REFERENCES

Gregory Kipper, 2004 edition, ”Investigator's Guide to Steganograph”,2004

[1] Stefan Katzenbeisser Fabien A. P. Petitcolas,2000,”Information Hiding Techniques for Steganography and Digital Watermarking”

[2] Ingemar J. Cox, Matthew L. Miller, second edition,Digital Watermarking and Steganography,2008

[3] Cox, I.J. Miller, M.L. Bloom, J.A., 2000,”Watermarking applications and their properties”, Coding and Computing, 2000. Proceedings. International Conference, pp 6 – 10, 06 August 2002

[4] Frank Y. Shih 2010,”image processing and pattern recognition fundamental and techniques”,chapter 12

[5] Vidyasagar M. Potdar, Song Han, Elizabeth Chang, International Conference on Industrial Informatics, “a survey of digital image watermarking techniques”,pp709 - 716,2005

[6] Ehab F. Badran, Ahmed Ghobashy, and Khamis El-Shennawy,” DCT-based digital image watermarking VIA image segmentation Techniques ”, Information & Communications Technology, 2006. ICICT '06. ITI 4th International Conference on,pp1-1,2007

[7] Frank Y Shih, 2010,Transactions on Data Hiding,chapter 12

[8] Chunlin Song, Sud Sudirman, Madjid Merabti and David Llewellyn-Jones,” Analysis of Digital Image Watermark Attacks”,, Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE,pp 1-5,2010

[9] S. Voloshynovskiy, S. Pereira, T. Pun, "Attacks on digital watermarks: classification, estimation based attacks, and benchmarks", Communications Magazine, IEEE,, Volume 39,Issue8, pp118 – 126, 2001

[10] Wei Lua,, Hongtao Lu,Fu-Lai Chung, "robust digital image watermarking based on subsampling”, Applied Mathematics and Computation, Volume 181, Issue 2, PP 886-893, 2006

[11] Yonggang Fu,Ruimin Shen,” Robust Image Watermarking Scheme Based on Subsampling”, Information Technology and Applications, 2005. ICITA 2005. Third International Conference on, Volume 2,pp 361-365,2005

[12] Ying-Hualui, Jing Qin, Jun kong, Si-Huili, "A robust digital watermarking scheme based on optimal coefficients selector about sub image", Wavelet Analysis and Pattern Recognition, ICWAPR '07. International Conference on, Volume 4,pp 1865 – 1869, 2007

[13] Kasmani, S.A., Naghsh-Nilchi, "A New Robust Digital Image Watermarking Technique Based on Joint DWT-DCT Transformation", Convergence and Hybrid Information Technology, 2008. ICCIT '08. Third International Conference on, Volume 2,PP 539 - 544,2008

[14] Changjiang Zhang, inshan Wang, Xiaodong Wang," Digital Image Watermarking Algorithm with Double Encryption by Arnold Transform and Logistic,"Networked Computing and Advanced Information Management, 2008. NCM '08. Fourth International Conference on,PP 329 – 334,2008

[15] Ping Wah Wong, Memon," Secret and public key image watermarking schemes for image authentication and ownership verification", Image Processing, IEEE Transactions on, Volume

[16] Michael Arnold, Martin Schmucker and Stephen D. Wolthusen,"Techniques and Applications of Digital Watermarking and Content Protection",2003

[17] Munesh Chandra, Shikha Pandey,"A DWT Domain Visible Watermarking Techniques for Digital Images", 2010 International Conference on Electronics and Information Engineering (ICEIE 2010),2010

[18] Sami Baba, Lala Krekor, Thawar Arif and Zyad Shaaban," Watermarking scheme for copyright of digital images", IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009

[19] Chunlin Song, Sud Sudirman, Madjid Merabti and David Llewellyn-Jones,"Analysis of Digital Image Watermark Attacks", IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2010 proceedings,2010

[20] Zhitian Zhou, Congyang Hu,Study on the E-government Security Risk Management, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.5, May 2008

[21] E-Government Bulletin, ISSUE 336, 01 June 2011, http://www.headstar.com/egb2011/336.html#top