Available online at www.sciencedirect.com

Journal of Applied Research and Technology

Journal of Applied Research and Technology 13 (2015) 435-442

www.jart.ccadet.unam.mx

Original

Script format document authentication scheme based on watermarking techniques

M. González-Lee[a], M. Nakano-Miyatake[b], H. Pérez-Meana[b,*], G. Sánchez-Pérez[b]

[a]*Universidad Veracruzana, Facultad de Ingeniería en Electrónica y Comunicaciones, Poza Rica, Veracruz, México*

[b]*Instituto Politécnico Nacional, Mechanical and Electrical Engineering School, México D.F., México*

Received 21 October 2012; accepted 21 May 2015

Abstract

In almost all watermarking-based document authentication systems, the documents are considered as binary images and then, the watermark is embedded using some image watermarking algorithm. However actually important documents are saved using document file formats, such as Portable Document Format (PDF) or Open Document Format (ODF), among others, because in general the file size is smaller compared with an image file, and also these are considered as more secure than other types of file format. However, the documents with these formats can also be maliciously modified for illegal purposes, making necessary the development of mechanisms that are able to detect such modifications. Considering the situations mentioned above, this paper proposes a document authentication scheme in which a watermark is directly embedded into the document file format as part of the document itself. The experimental results show a desirable performance of the proposed algorithm.

Keywords: Digital watermarking; Document authentication; Document file format; Script format document

1. Introduction

Nowadays almost all documents are generated in a digital form and stored using some file formats such as Portable Document Format (PDF) or Open Document Format (ODF), among others, because of the reduction of storage space and rapid access that these file formats provide. However, the digital documents can be easily modified by unauthorized persons resulting in an altered document with the same quality as the original one. Despite some file formats such as PDF includes some security mechanisms, these can be broken as described in section 2. These tampered documents can be used or distributed illegally causing economical and moral damages to the involved persons. This fact suggests the necessity to develop efficient document authentication systems.

Several schemes have been developed to authenticate digital documents which embed invisible watermark into them; most of these schemes consider the digital documents as binary images. For example, Yang and Kot (2004) proposed a document authentication scheme, in which an authentication code is embedded by changing the spaces size between consecutive words and characters (Yang & Kot, 2004). The main drawback of this scheme is its high computational complexity and vulnerability against noise contamination. Huang et al. (2004) proposed an authentication method for binary images including text documents, in which firstly the binary image is segmented in blocks and then some pixels in each block are rearranged in order to enforce a given relationship between the number of black and white pixels in it. During the authentication process, this relationship is verified for each block in order to authenticate each block. If the determined relationship is satisfied, then the block is considered as authentic, otherwise the block is labelled as tampered. The principal disadvantage of this method is that a degradation introduced in the encoded binary image is noticeable. Wu and Liu (2004) proposed block-wise binary image authentication scheme, in which flippable pixels in each block are manipulated in order to embed a watermark bit in that block (Wu & Liu, 2004). Here the embedded watermark is imperceptible, because the flippable pixels can be flipped without causing any distortion in the binary image. However, in general, the watermark embedding payload is very low compared with the number of flippable pixels into the image. To improve the embedding payload, Gou and Wu (2007) introduced the concept of "super-pixels" and wet paper coding into the Wu and Liu's scheme (Gou & Wu, 2007). The "super-pixels" form a set of individually non-flippable pixels, which can be removed or added together without causing visual distortion. Also Wu and Liu (2004) reported that their authentication scheme is robust to printing and scanning opera-

*Corresponding author.
E-mail address: hmperezm@ipn.mx (H. Pérez-Meana).

tions. However during the scanning process, a rotation, even with angles smaller than 1°, may result in a synchronization loss between watermark embedding and detection process.

Document authentication schemes for document file formats such as PDF or PostScript had received few attention among researchers although many official documents are stored using this type of formats. Adobe Systems Inc. provides two security mechanisms, which are briefly described in section 2. Zhu et al. (2007) proposed a document authentication method using render sequence encoding, in which the encoding process is based on modulate the displayed sequences using a Document Description Language (DDL), such as PostScript, PDF, Printer Control Language, etc. In the render sequence, predefined characters are permuted by a user's secret key, and then, during the authentication process, the document is considered as authentic if the permutation corresponds to the secret key used in encoding stage. This scheme determines correctly if a document is authentic or not, however there are two inconveniences that may limit its practical use. Firstly the size of the encoded document file is considerably increased compared with the original file size, and the second one is the fact that the structure of the encoded render sequence is unnatural, and as a consequence, it can be easily detected by an unauthorized person that the document may be protected, doing it possible the use of reverse engineering to tamper the document. To solve these problems, Gonzalez-Lee et al. (2009) proposed a watermarking-based document authentication scheme, in which the character metrics are used to embed a watermark sequence. The advantage of this scheme is that the watermarked file size is not changed compared with original file size and also the watermarked file conserves its original appearance, enhancing in this form its security because the watermark presence is not evident. However, sometimes the watermarking algorithm proposed by Gonzalez-Lee et al. (2009) causes slight visual artifact in the watermarked document file because the watermark is embedded directly into the character metrics. To overcome the problems present in Zhu et al. (2007) and Gonzalez-Lee et al. (2009), this paper proposes a document authentication method in which the watermarked ASCII values are used to modify the character metrics using more efficient embedding method, to avoid the visual distortion presented in Gonzalez-Lee et al. (2009) as described in "Results" section.

The rest of this paper is organized as follows: in section 2, a brief explanation of DDL and scripts file are given. Also some security mechanisms for PDF documents provided by Adobe Systems Inc. and their vulnerabilities are explained in this section. The proposed document authentication scheme is described in section 3. Section 4 provides the evaluation results and the comparison of the proposed scheme with those proposed in Zhu et al. (2007) and Gonzalez-Lee et al. (2009). Finally the conclusions are provided in section 5.

2. Document Description and Language Security Mechanisms

A DDL is a set of instructions or commands that can be used to describe a document. Some examples of a DDL are the PDF, the PostScript Language, Hyper Text Markup Language (HTML) and Extended Markup Language (XML). On the other hand, a DDS is a structured sequence of commands, derived from the DDL that provides a set of instructions that indicate to the edition software how to interpret the document structure and how to draw it on a screen or printer. The DDS contains a set of low level commands that describe the objects properties in a document. For example, the main properties of a letter are the font type, size, color and position where it must be drawn, while the main properties of a line are the line type (continuous, dotted…), width and color, etc. Thus, the DDS is a subset of DDL instructions structured such that they are able to describe a given document. The DDL is similar to a high level computer language, which includes all available instructions for developing document files, while the DDS corresponds to a source code of a computer program following to the DDL.

Nowadays one of the most widely used DDL is the PDF offered by Adobe Systems Inc. in which some security mechanisms, such as access control and fingerprinting-based authentication, are provided. In access control, a password is assigned to the users which are required for viewing, printing and copy the documents because using this password the file is encrypted. However this password can be obtained using some tools available in Internet such as PDFcrack (Noren, 2012). Once the password is obtained, the document can be modified using some editing tools, such as PDFeditor (Hocko et al., 2012), hexadecimal editors or even, in a few cases, Notepad of Windows. On the other hand, the fingerprinting is also provided to authenticate the document contents. However, because after encryption and/or compression process the PDF file hierarchical structure as well as heading body, cross-reference table and trailer remains as plain text, the fingerprint data can be found and removed.

A common attacker to PDF file firstly tries to obtain the password if the PDF file is encrypted, open the file using the obtained password and then print it in another file. In this time encryption is already broken, next the attacker modifies the content of document for his convenience using some PDF editors mentioned above and finally print it in a new file using the same password. In this form the tampered document can be used for the attacker's purpose. This attack was successfully carried out, with relatively easiness, in our laboratory to several PDF files. Then it is desirable to develop document authentication schemes that complement the security mechanisms provided by the PDF systems.

In this work, we assume that the document under analysis is created using some file standard, in which the metrics are available, such as PDF file standard. This standard describes every row in the document by using either, two vectors where the first one contains the characters to be drawn and the second one consists of the metrics of each character, or using one vector containing both, the characters and their metrics. The metric value is generally a 3 decimal number, for example m_x=3.243. The metrics used here should not be confused with a physical size of character. The metrics are related to the distance between the origins of two glyphs, where a glyph is the graphic shape that represents a given character, usually the origin of a glyph is not the origin of the character itself; because a glyph is a box shape containing a character, as shown in Figure 1.

The character metric indicates the position where the next character is drawn, therefore it controls the properly distribution of the characters in a document.

3. Proposed Document Authentication Algorithm

In the proposed document authentication scheme, the watermark bit sequence is embedded into a DDS, becoming an integral part of the document description. Then in the authentication stage, the document authenticity is determined using the watermarked DDS. Figure 2 shows the block diagram of the proposed scheme that indicates where the embedding process takes place. Here, firstly the input document is assumed to be represented by a set of symbols belonging to the ASCII code, where the description of each character in the document is denoted by a triplet given as: (c, m_x, m_y), where c is the character to be drawn, and m_x and m_y are the horizontal and vertical character metrics, respectively.

3.1. Watermark Embedding Process

The watermark embedding process of the proposed scheme consists on the modification of the metrics values, according to a bipolar zero mean random binary watermark bit sequence (–1 or 1), generated employing a user secret key. Because most tampering attacks intend to replace some words of a given text, while keeping the separation between consecutive rows almost unaltered, in the description and evaluation of the proposed method, only horizontal metrics are watermarked, thus in the rest of the paper it is assumed that $m_y = 0$. However, the proposed scheme can be easily applied to other types of documents in which $m_x = 0$ and $m_y = 0$ $m_y \neq 0$ or even with $m_x \neq 0$ and $m_y \neq 0$.

To embed adequately the watermark sequence into the character metrics, firstly the whole document is interpreted in order to be able to form two vectors. The first vector consists of all characters $\mathbf{T}[t_j]$ in the document, whose j-th element t_j is given by

$$t_j = ASCII(c_j) \tag{1}$$

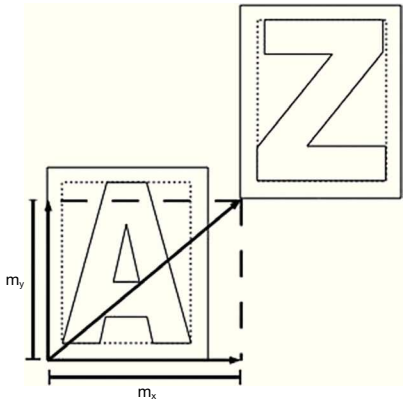

Fig. 1. The character metrics fix the position of the next character, and are used to properly distribute characters in a document.

where c_j is the j-th character in the document and $ASCII(c_j)$ is the ASCII value of c_j. For example if c_j is character 'A', then $ASCII(c_j)$=097. Next the watermark bit is embedded using a multiplicative embedding rule given by

$$t'_j = t_j\left(1 + g w_j\right) \tag{2}$$

where t_j is the watermarked ASCII code of j-th character (note again that $\mathbf{T}'[t'_j]$, $j=1,2,3\ldots,N$, w_j is j-th binary watermark bit $\{-1,1\}$, generated employing the user secret key, and g is the watermark embedding gain which controls the watermark imperceptibility. Experimental evaluations show that g=.05 simultaneously provides a fairly good watermark imperceptibility and tamper detection capability, as shown in the evaluation results provided in section 4. The second vector, $\boldsymbol{M}[m_{x,j}]$, $j=1,2,3,\ldots,N$, contains the metrics in the x-axis, which is combined with the watermarked ASCII code $\mathbf{T}'[t'_j]$, to obtain the watermarked metrics, $\mathbf{M}'[m'_{x,j}]$, whose j-th element is given by

$$m'_{xj} = m_{xj} + \frac{t'_j}{10^6} \tag{3}$$

Then the vector of metrics $\boldsymbol{M}_x[m_{x,j}]$ is replaced by the watermarked vector metrics $\boldsymbol{M'}_x[m'_{x,j}]$, and used together with the vector of the watermarked ASCII code $\mathbf{T}'[t'_{xj}]$ to generate the watermarked document. Here the watermarked metrics is represented using at least 5 decimal digits, in order that the sys-

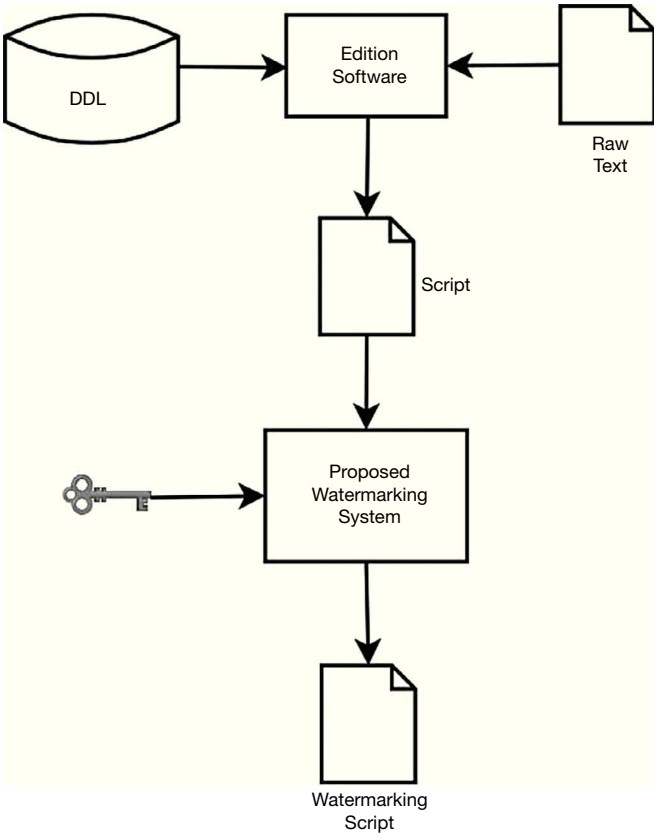

Fig. 2. Block diagram of proposed document watermark embedding system.

tem be able to detect it, as described in section 3.2. The proposed system can be used to authenticate modifications in both *x*-location and *y*-location. To this end, if the system is required to detect modifications in the *y*-locations, that is $m_x = 0$ and $m_y \neq 0$, it is necessary only to replace m_x by m_y in Eq. (3), while if the system is required to detect tampering in both, *x*-location and *y*-location, the system firstly embeds the watermark in the metric m_x using Eq. (3) and then replacing m_x by m_y in the same equation embeds the watermark in m_y. Figure 3 shows the block diagram of the embedding stage of the proposed system.

3.2. Authentication Process

During the authentication process, shown in Figure 4, the watermarked document is inserted into a document interpreter, as that used in the watermark embedding stage, which extracts the ASCII codes and metrics of watermarked document. Next using the ASCII codes and the secrete user key the original watermark is generated using Eq. (2). Then the cross correlation (CC) between the original watermark sequence estimated in the authentication stage, $\hat{t}'_j$, and the extracted one, $\tilde{m}_{x,j}$ is computed in order to determine the presence of the watermark. Thus the CC value *d* between the metrics vector $\widetilde{\mathbf{M}}_x = \left[\tilde{m}_{x,j}\right]$ extracted from the watermarked and possibly tampered document, and the original watermark sequence $\hat{\mathbf{T}} = \left[\hat{T}_j\right]$ is calculated which is given by

$$d = \frac{1}{N}\sum_{j=1}^{N} \hat{m}_{x,j}\hat{t}'_j, \tag{4}$$

where $\tilde{m}_{x,j}$ and t'_j are the extracted watermarked metric and the original watermark of the *j*-th character, respectively and *N* is the watermark sequence length. In order to reduce the influence of the original character metrics, the watermarked ASCII code used to modify the original metrics is estimated using the following equation:

$$\tilde{m}_{x,j} = 10\left(10^3 \bar{m}_{x,j} - 10^3 \bar{m}_{x,j}\right), \tag{5}$$

where $\bar{m}_{x,j}$ is the *j*-th extracted metric from the watermarked DDS, and *x* means integer part of *x*. Then CC value *d* is compared with a predetermined threshold value *Th* to decide if the watermark is present or not. Thus if $d \geq Th$, almost all part of the original watermark is present into the document and then it is considered as authentic, otherwise the document can be tampered. The threshold value is estimated as:

$$Th = 5\sqrt{\frac{2\sigma^2}{N}} \tag{6}$$

where σ^2 is the variance of the extracted metrics, $\tilde{m}_{x,j}$ of all characters of the document under evaluation. Eq. (6) is obtained from the optimal threshold equation for cross correlation-based detectors proposed by Piva et al. (1998), since the proposed algorithm is based on the same assumptions proposed by Piva et al. (1998). To adequate the optimal threshold equation to the proposed scheme, the constant value is modified. The formula

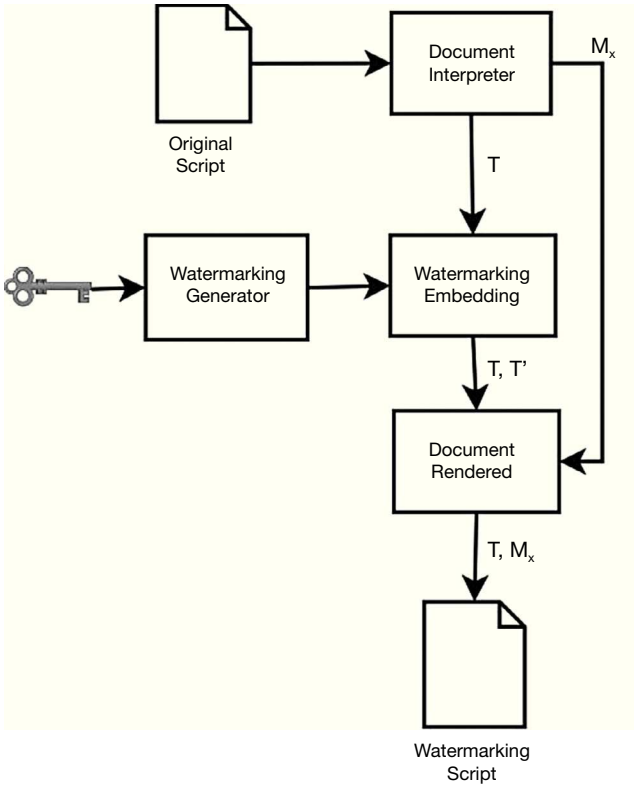

Fig. 3. Embedding stage of the proposed document authentication stage.

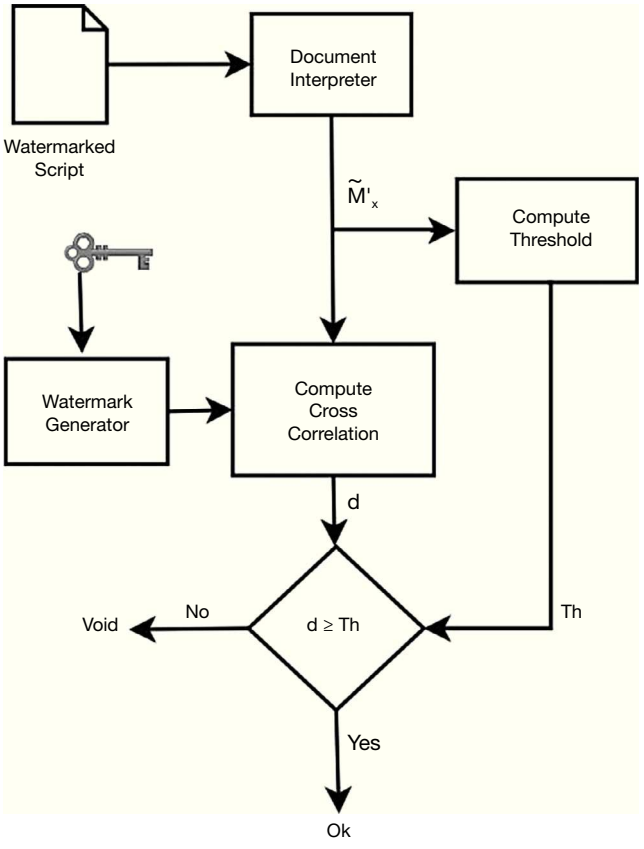

Fig. 4. Authentication stage of the proposed scheme.

proposed by Piva et al. (1998) uses constant value '3.3' instead of '5'; we modified it for the following reason: because the watermarked metrics is highly correlated to the watermark sequence, an increase of the constant value '5' produce a larger value of *Th*. This helps to reduce false negative error rate, avoiding that small alterations cannot be detected. However in this situation, false positive error rate will be increased. The watermark detection process is described in Figure 4.

4. Experimental Results

The proposed document authentication scheme is evaluated from the watermark imperceptibility and tamper detection capability points of view.

4.1. Watermark Imperceptibility

Figures 5A and B show a part of the original document used for evaluation of the proposed system and its watermarked version, respectively. In these figures, both the original and watermarked documents are visually identical. Because the document file is not an image, the watermark imperceptibility cannot be measured by PSNR or other standard image quality assessments. Thus in order to evaluate the watermark imperceptibility, a Mean Opinion Score (MOS) evaluation was used. To this end twenty pairs of different documents containing the original and watermarked versions of each one are shown to 100 observers, which are distributed as shown in Table 1. The observers are then asked to assess the difference between the original and watermarked document, and then according to the criteria shown in Table 2, they assign a score in the range from 1 to 5 for each pair of documents. The MOS value for twenty pairs of documents obtained from 100 observers is 4.8, which suggests that the proposed system guarantees the watermark imperceptibility. The watermark imperceptibility can also be expected by the fact that, according to (2) and (3), the maximum modification introduced in the distance between two consecutive characters is approximately 3.7187×10^{-6} inches, taking into account that a metric equal to 1.0 unit corresponds to 1/72 inches.

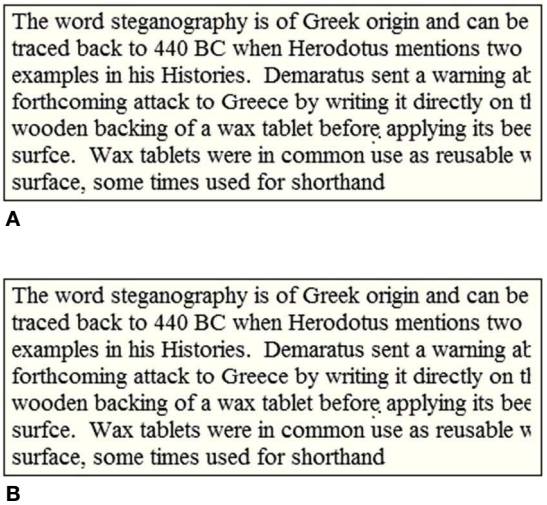

Fig. 5. A part of the original document (A) and its watermarked version (B) used for evaluation of proposed document authentication scheme.

Table 1
Observer's distribution.

Age	Women	Men
20-30	33	32
30-40	4	10
40-50	2	7
>50	3	9

Table 2
Evaluation criteria.

Score	Meaning
5	There is not any perceptible difference
4	There is a slightly perceptible difference that can be ignored
3	There is a slightly perceptible difference that cannot be ignored
2	There is a noticeable distortion
1	There is an unacceptable distortion

4.2. Tamper Detection Capability

If the watermarked document is tampered replacing some words or letters with others, the CC value between the extracted watermarked metric, $\tilde{m}_{x,j}$, and the original watermark sequence decrease according to the number of replaced characters. Figure 6 shows the relationship between the CC value and the percentage of the tampered characters, obtained by averaging 200 different documents with approximately 3500 characters, together with its threshold value. The threshold value also varies, because alteration of characters causes change of the variance σ^2 in (6). The evaluation results show that if more than approximately 0.25% of characters are altered, the proposed scheme determines that the document is tampered. This means that in a document of about 3500 characters, the system determines that the document is tampered if more than 8 characters are altered.

Table 3 shows the CC values, the corresponding threshold values and the minimum percentage of the number of altered characters that the proposed scheme requires for determining that the document is tampered. These 10 documents are selected to show some typical cases including the best and worst cases from the total of 200 documents evaluated. All documents were one page full texts, without special characters, and the number of characters was in the range from 3163 to 3932. In the worst case, if more than approximately 0.7044% of total characters that is approximately 24 characters in a document of 3450 characters are modified, the proposed scheme can detect alteration of the document. However, even if number of altered characters is smaller than 24 in the worst case, the document could be distorted. For example as shown in Figure 7, where some words on the watermarked document shown in Figure 5B were replaced, "Herodotus", "Histories of Herodotus" and "Demaratus" were replaced by "Shannon", "Cover Writing Handbook" and "Somebody". The produced document has many overlapped letters and unnatural spacing between letters of words as shown by Figure 7. So, even if the system cannot detect this alteration, the tampered document cannot be acknowledged as a legitimate one.

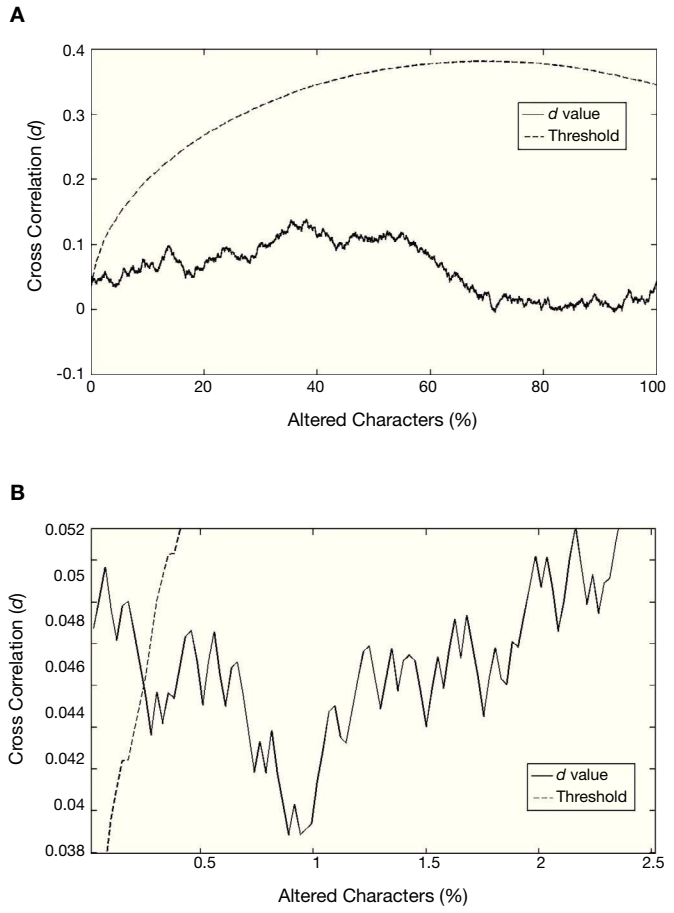

Fig. 6. Average value of d for different percentage of the tampered characters, together with the threshold value Th. A: 0% to 100% tampered characters. B: 0% to 2.5% tampered characters.

Table 3
Percentage of minimum altered characters (AC) that the proposed scheme can detect correctly together with threshold (Th) values and cross-correlation (CC) values.

Doc	(Th)	CC (d)	AC (%)
1	0.0567	0.0565	0.3436
2	0.0496	0.0479	0.2969
3	0.0637	0.0612	0.5691
4	0.0382	0.0379	0.1512
5	0.0498	0.0462	0.4415
6	0.0608	0.0597	0.6677
7	0.0453	0.0428	0.2289
8	0.0479	0.0473	0.2447
9	0.0429	0.0412	0.1597
10	0.0532	0.0524	0.7044

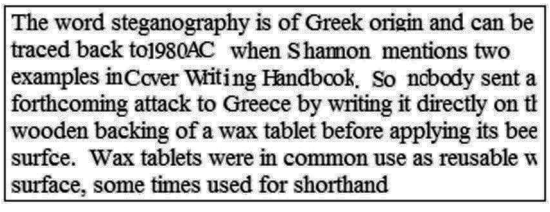
The word steganography is of Greek origin and can be traced back to1980AC when Shannon mentions two examples inCcver Writing Handbcok. So ncbody sent a forthcoming attack to Greece by writing it directly on tl wooden backing of a wax tablet before applying its bee surfce. Wax tablets were in common use as reusable w surface, some times used for shorthand

Fig. 7. Tampered document where some words are replaced by others. The resulting document looks unnatural with many letters overlapped.

The computational complexity is also an important issue for reliability of a document authentication scheme. In the proposed authentication process, the watermark detection is carried out using (4), (5) and (6). Here, when a watermark sequence has N bits, the number of multiplications required in the authentication stages is approximately $5N$. The average execution times of the authentication process measured using several documents with different sizes is shown by Figure 8. This figure shows the execution times including the memory access, as well as the time required by the watermark detection and other several processes, during the authentication task. For a reasonable amount of document, this execution time is short enough, enabling its use in many practical applications.

4.3. Comparison With Previously Works

In the proposed document authentication scheme, the watermark sequence is embedded into the metrics of the Script file; and then in the authentication stage, using these watermarked metrics, it is determined if the document is authentic or not. In the literature, there are few document authentication algorithms; some of the more successful schemes are the proposed by Zhu et al. (2007) and Gonzalez-Lee et al. (2009). In Zhu et al. (2007), the authentication code embedding is carried out into a DDS file; however the size of coded DDS file is increased comparing with its original size, while in the proposed scheme, the Script file size is not changed. Thus the file size before and after watermark embedding is almost same. In Zhu's scheme, the average file size increasing rate is approximately 1.83 (Zhu et al., 2007); that is, if the original file size is 1701 bytes, then the encoded file size becomes to be 3121 bytes, while in the proposed scheme, the average file size increasing rate is approximately 1.05, which means that the file sizes are practically the same before and after the watermarking operation. Here the file size increase is due to the watermarked metrics requires five decimal digits, instead of the three required by the original file.

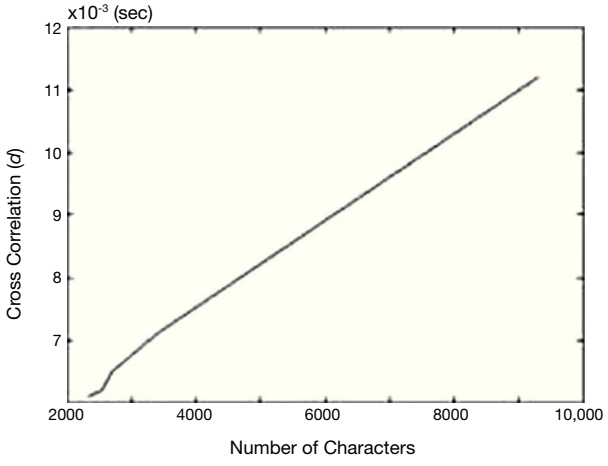

Fig. 8. Execution times of document authentication process with different document sizes.

The structure of the encoded or watermarked file is another important aspect in this type of authentication, because if the encoded file has an abnormal structure compared with a normal file, a suspicion is aroused and using reverse engineering, the encoded file could be counterfeited. Figures 9A and B show an example of original DDS file and encoded DDS file by Zhu's scheme (Zhu et al., 2007). On the other hand, Figures 9C and D show an example of original and Script watermarked files using the proposed scheme.

In the preliminary results reported by the authors in Gonzalez-Lee et al. (2009), proposed a document authentication system based on similar principle to the proposed in this paper, however in Gonzalez-Lee et al. (2009), the watermark is embedded directly into the character metrics, which may produce larger distortion of the watermarked document, because the difference between the watermarked and original metrics may be significant. To avoid this problem, the proposed scheme reduces this difference significantly, as shown in Figure 10. From this figure it follows that the difference between the original metrics and the watermarked one in the proposed scheme is in the order of 10^{-4} in the worst case, which is quite smaller than those of Gonzalez-Lee et al. (2009). Then it can be concluded that the visual distortion of proposed scheme must be negligible. Additionally the proposed scheme introduces an optimal threshold which improves the detection performance compared with Gonzalez-Lee et al. (2009).

5. Conclusions

In almost all watermarking-based document authentication schemes, documents are considered as binary images, and then watermark sequence is embedded modifying binary pixels. However many important documents are saved in document file format in place of image file. Considering this situation, in this paper a watermarking-based document authentication scheme is proposed, in which watermark sequence is embedded directly into the script file format, modifying the character metrics in this format.

The simulation results show the watermark imperceptibility, obtained quite good MOS. The tamper detection capability is evaluated using 200 documents; the simulation results show that, in average, if more than approximately 0.25% of total characters of the document are altered, the proposed scheme can detect alteration of the document, with the best case equal to 0.1512% and worst case is 0.7044%. However, as shown in Figure 7, even if the authentication stage of the proposed scheme considers a document with some few tampered characters as authentic, the tampered watermarked Script file produces (displayed or printed) an unnatural document, which cannot be acknowledged as a legitimate one. Also the comparison results with the previous works by Zhu et al. (2007) and Gonzalez-Lee et al. (2009) show better performance of the proposed scheme. Regarding to Zhu et al. (2007), the proposed algorithm produce a DDL file with same structure and size than the original file, while a file produced by Zhu et al. (2007) looks unnatu-

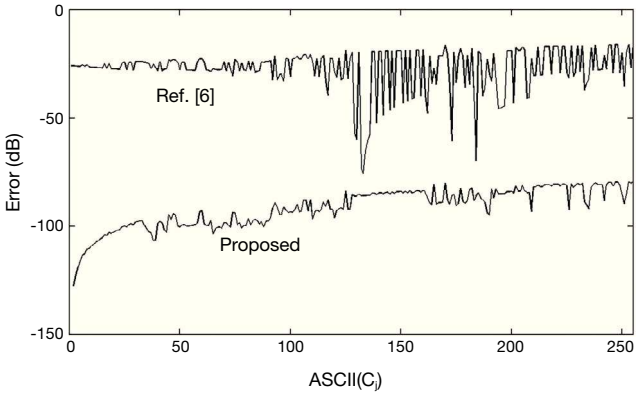

Fig. 10. Comparison of the difference between original metrics and watermarked metrics in the proposed and Gonzalez et al. (2009) schemes.

```
1  100  200  moveto              % positioning
2  (This is a sentence) show     % Characters
3  showpage
```

A

```
1   124  200  moveto      % Positioning
2   (is) show            % Characters
3   144  200  moveto      % Positioning
4   (sentence) show      % Characters
5   100  200  moveto      % Positioning
6   (This) show          % Characters
7   185  200  moveto      % Positioning
8   (.) show             % Characters
9   135  200  moveto      % Positioning
10  (a) show             % Characters
11  showpage
```

B

```
50  742  Moveto
(This is a sentence)
[ 7.33162  6.221628  4.001628  5.111628  3.721778
          4.001628  5.111628  3.721778  5.661938
          3.721778  5.111628  5.661938  6.221628
          4.001638  5.661938  6.221628  5.661628
          5.661938  3.721778  ] x show
```

C

```
50  742  moveto
(This is a sentence)
[ 7.258312  6.159411  4.041644  5.162744  3.684560
          4.041644  5.060512  3.758996  5.605319
          3.758996  5.162744  5.605319  6.283844
          3.961612  5.718557  6.159411  5.718557
          5.605319  3.758996  ] x show
```

D

Fig. 9. Examples. A: original DDL file. B: encoded DDL file produced by Zhu et al. (2007) system. C: Original Script file. D: watermarked Script file by the proposed system.

ral and its size is significantly increased, doing it evident that the original file has been manipulated. On the other hand, in Gonzalez-Lee et al. (2009), the watermark is embedded directly into the character metrics, which may produce visual artifact in the watermarked document, while in the proposed algorithm, the maximum modification introduced in the distance between two consecutive characters is approximately 3.7187×10^{-6} inches, taking into account that a metric equal to 1.0 unit corresponds to 1/72 inches. This fact ensures the watermark imperceptibility.

Finally the proposed algorithm is not intended for replacing the PDF security mechanisms, which presents some vulnerability, but for complementing them such that, even if the PDF security mechanisms are broken, the document can remain still protected using the proposed scheme.

Acknowledgments

The authors would like to thank the Council of Science and Technology (CONACYT) in Mexico, to the Universidad Veracruzana, Poza Rica campus, and to the Instituto Politécnico Nacional for supporting this work.

References

Gonzalez-Lee, M., Santiago-Avila, C., Nakano-Miyatake, M., & Perez-Meana, H. (2009). Watermarking based document authentication in script format. *Proceddings of the 52th IEEE Midwest Symp. on Circuits and Systems, 1,* 837-841.

Gou, H., & Wu, M. (2007). Improving Embedding payload in binary images with super-pixels. *Proceedings of the IEEE International Conference on Image Processing, 3,* 277-280.

Hocko, M., Misurka, J., & Petricek, M. (2002). PDFill. Retrieved April 2012 from: http://www.PDFill.com

Huang, P.M., Wu, D.C., & Tsai, W.H. (2004). A novel block-based authentication technique for binary images by block pixel rearrangements. *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 1,* 903-906.

Noren, H. PDFcrack. Retrieved April 2012 from: http://sourceforge.net/projects/pdfcrack/

Piva, A., Barni, M., Bartolini, F., & Cappellini, V. (1998). Threshold selection for correlation-based watermark detection. *Procedings of COST254 Workshop on Intelligent Communications, 1,* 67-72.

Wu, M., & Liu, B. (2004). Data hiding in binary image for authentication and annotation. *IEEE Transactions on Multimedia, 6,* 528-538.

Yang, H., & Kot, A.C. (2004). Text document authentication by integrating inter character and word spaces watermarking. In: *2004 IEEE International Conference on Multimedia and Expo, 2004. ICME'04* (Vol. 2, pp. 955-958).

Zhu, B., Wu, J., & Kankanhalli, M.S. (2007). Render sequence encoding for document protection. *IEEE Transactions on Multimedia, 9,* 16-24.