Performance Assessment of Zero-Watermarking Techniques for Online Arabic Textual-Content

Omar Tayan[1,2], Yasser M. Alginahi[1,3] and Muhammad N. Kabir[1]

[1]IT Research Center for the Holy Quran and Its Sciences (NOOR)
[2]College of Computer Science and Engineering
[3]Deanship of Academic Services
Taibah University, P.O Box 344, Al-Madinah Al-Munawarrah, Saudi Arabia
{otayan, yginahi}@taibahu.edu.sa, dr.nomankabir@gmail.com

Abstract: The advantages of fast and simple digital information exchange over the Internet has attracted problems and threats in the form of digital attacks that could compromise information integrity, protection and authentication. Such issues are more prominent for text-content due to the dominance of textual-data sent online. Any compromise on integrity and security is clearly intolerable for the case of sensitive textual-content being propagated. This paper addresses mechanisms for ensuring intact integrity and authentication of samples of sensitive textual-content disseminated over the Internet through the use of zero-watermarking. In this work, two robust zero-watermarking approaches (Method A and B) are proposed which are capable of detecting any content-modifications while avoiding any embeddings/modifications on the original text to be disseminated. The proposed methods provide a good indication of the relative sensitivities of each approach to third-party modifications during the key-extraction phase. The contribution of this paper includes a comparative analysis of the newly proposed methods against the existing relevant state of the art techniques based on two cost functions specifically applicable to our target application domain with promising results. Finally, the derived system was capable of achieving a pivotal requirement by ensuring that the textual-content could be traced back to its original publisher for authentication purposes, or otherwise, had detected document-tampering, as in the case of third-party modifications.

[Tayan O, Alginahi Y.M, Kabir M.N. **Performance Assessment of Zero-Watermarking Techniques for Online Arabic Textual-Content**. *Life Sci J* 2013;10(4):93-100] (ISSN:1097-8135). http://www.lifesciencesite.com. 13

Keywords: watermarking; evaluation; online Arabic text; sensitivity-analysis.

1. Introduction

Since the invention of the Internet, the last two decades has witnessed a tremendous increase in digital content opening opportunities in many areas of research. This can jeopardize the security and authenticity of the digital content since it is easy to edit and port documents on Internet. One area of research which gained much importance is the security and protection of digital content. The perception of securing text documents began towards the end of the last century with the increase in dissemination of text documents over the Internet (Maxemchuk and Low, 1997). The major content of the Internet is plain text besides other types of digital content such as images, audio and video. Therefore, authentication and protection of text documents is essential during the online dissemination process.

Digital watermarking involves the incorporation of information into any digital content, which is then used for the purpose of copyright protection and authentication. A digital watermark is considered as a unique signature of the owner who owns the copyright of the document/content with the purpose of protecting the digital content from illegal copying, modifications (forgery/distortions) and/or redistribution. Therefore, in order to achieve this, the characteristics of a reliable watermark algorithm must include security, imperceptibility, integrity and authenticity (Adesina *et al*., 2010; Jalil and Mirza, 2009).

The watermarking process involves several stages, starting from the generation of the watermark, followed by the distribution phase that includes exposures to attacks and finally the watermark detection stage. The early days of the Internet raised the need for conducting research in the area of text watermarking; the first work that appeared on text watermarking can be traced back to Maxemchuk and Low (1997). Since then, many techniques have been published in this area. According to the literature, text watermarking techniques are classified into four main approaches: image-based, semantic-based, syntactic-based and structural-based. In image-based methods, the watermark is embedded in the text image. In semantic-based methods, the semantics of the text are utilized in order to embed the watermark. In syntactic methods, the syntactic structure of text is utilized to embed the watermark. In structural methods, the structural properties of characters are explored to embed the watermarks.

Image based approaches include shifting of words, sentences and paragraphs (Brassil *et al.,* 1999), shifting average inter-word distances (Huang and Yan, 2001), and adjusting letter slopes (Davarzani and

Yaghmaie, 2009). Semantic based approaches include: synonym substitution (Jensen, 2001), algorithms based on noun-verbs, algorithms based on typos, acronyms and abbreviations, algorithms based on linguistic approaches of presuppositions, and algorithms based on text-meaning representational strings (Lu, 2009). Syntactic based approaches include: Natural Language (NL) watermarking, syntactic tree and morpho-syntactic alterations (Maxemchuk and Low, 1997). Structural based approaches include: techniques based on occurrences of double letters (Kaur and Babbar, 2013).

It was noted that the above mentioned approaches were not applicable to all types of tampering attacks. Hence, in techniques based on a modified zero-watermarking approach are proposed in this paper to provide authentication, protection and tamper detection measures. The purpose here is to develop techniques which will not cause any modifications to the text within a document. Therefore, the choice is to use zero-watermarking approaches as an appropriate method. The related work on text zero-watermarking techniques is presented here in order to provide an overview on this approach which has not been studied significantly in the literature, as in particular for the case of Arabic-text, and it is one of the most important techniques to be used on sensitive documents such as Holy Quran Scriptures, where no alterations or modifications are allowed on the text.

In all zero-watermarking algorithms found in the literature, the original copyright author/owner generates the watermark key from the text document without performing any modifications to the text. On the contrary, the text attributes/properties are used to generate the watermark key. This author-generated key is then registered with time and date, and stored with the CA. In case of modifications done to the author's copyrights, this key is used to verify the original owner by using the extraction algorithm.

In (Jalil *et al*., 2010a), a zero-watermarking technique which uses the properties of text to generate a watermark is proposed. This is done by first analyzing the text document and locating propositions, in which the average frequencies proposition (AFP) is identified and the AFP partitions of the text are created. Next, the count of alphabetical characters in each partition is calculated and the highest non-vowel ASCII characters are used to populate the most occurring non-vowel list (MONV); which is then used to generate the watermark key. Finally, in the detection (extraction) stage this key is used by the certification authority (CA) to compare with the key obtained from the tampered document upon a request or a claim.

In another work by (Jalil *et al*., 2010b), a proposed a zero-watermarking algorithm to protect against different kinds of tampering, such as passivization, clefting, topicalization or rephrasing is proposed. The objective was that such tampering should not modify the nouns, proverbs or adjectives which contain more than four letters. The attackers here only shuffle the text (words positions) and cannot avoid skipping them. To create the watermark, all the initial letters from all the words containing more than four letters are used to generate the watermark patterns from each sentence before those characters are concatenated to construct the watermark.

In a another work (Jalil *et al*., 2010c), the proposed zero-watermark key is based on generating the key from the maximum occurring first letter (MOFL) list formed after using the propositions in the text as the separator which are used to form groups where each group contains group size (GS) partitions. Next, from each group the occurrences of the first letter in each double letter are analyzed from which the MOFL list is formed. Finally, the watermark key is generated from the list.

The algorithm in (Jalil *et al*., 2010d) depends on author's selection of a keyword from the text. Therefore, the watermark is generated based on the length of preceding and next word lengths to and from the keyword occurrences in the text. In (He *et al*., 2009), the text zero-watermark is created based on extraction of words which correspond to one special part of a speech tag sequence. These words were chosen under the control of a chaotic function. The experimental results show that the watermark key cannot be destroyed by stochastic synonym substitution and sentence transformation. This algorithm is robust, secure and imperceptible. Hence, zero watermarking techniques depend on the characteristics of the language and therefore many of the proposed techniques available may not be applicable to other languages.

In another approach based on sentence entropy (Yingjie *et al.,* 2010), the entropy of the sentence is calculated based on word frequency and crucial selections are based on entropy from which the watermark is constructed with the order of crucial sentences. Zero-watermarking based on space models was also proposed in (Yingjie *et al.,* 2011), whereby the zero-watermark was constructed from a three-dimensional model; constructed from the 2D coordinate of word-level and the sentence weights of the sentence-level. The approaches in (He *et al.*, 2009; Yingjie *et al.,* 2010; Yingjie *et al.,* 2011) were proposed for Chinese text and were based on the characteristics of the Chinese letters.

From the literature review it is shown that extensive research was conducted on watermarking techniques in digital natural language documents, particularly digital text-content in Latin and Chinese languages, with only few techniques presented for other languages which use Arabic scripts (Davarzani and Yaghmaie, 2009; Tayan *et al.,* 2013; Al-Haidari *et al.,* 2009; Aabed *et al.,* 2007; Shirali-Shahreza and Shirali-Shahreza, 2008). However, only approach (Tayan *et al.,* 2013) was based on zero-watermarking for Arabic scripts.

The reminder of the paper is organized as follows: the introduction section is followed by the proposed modified watermarking approaches in section 2. The results and discussion section is given in section 3, and finally, section 4 concludes the paper.

2. Methodology

This paper introduces two new approaches for integrity and authentication protection of online Arabic plain-text documents based on zero-watermarking. The proposed algorithm performs a logical, rather than physical, embedding of watermark-data in the cover-document. The motivation here is to provide a mechanism for the secure dissemination of critical and sensitive documents in which any physical modification can render the document as invalid. Application examples are numerous, and include: Holy Scriptures such as the (Holy Quran), Arabic poetry text which include diacritical symbols… etc. In the proposed approaches, two zero-watermarking frameworks are implemented. Both frameworks are based on the original approach proposed by the authors in (Tayan et al., 2013). In this previously proposed system (Tayan et al., 2013), the original publisher embeds a data-sequence obtained from an image-logo in a copy of the cover-document. This is achieved after processing and classification of the cover-text into its constituent word-sets (groups), with one-watermark bit being inserted per set, with the set-size being a variable parameter. Next, using a key-generation algorithm a unique key is produced which is registered together with the logo at a CA. In the decoding stage, the CA embeds the publisher's logo in the sample text before the process of test-classification and grouping is performed. The key is then extracted based on the characteristics of the input document. Finally, the newly generated key is compared with the CA registered key for confirming authentication of the textual-content. The reminder of this section provides details of the proposed approaches.

Proposed Method A:

During the encoding phase shown in Figure 1, the original publisher embeds a data-sequence, W_{CS}, obtained from the watermark logo W_L, on the words of the copy cover-document (T_C). All consequent operations are only performed on the duplicated text (T_C), whilst the original text (T_O) is untouched and sent for dissemination. Next, using a key-generation algorithm, a unique key, W_{KG}, is produced, in that from the data sequence, W_{CS}, one bit is inserted to the least significant end of the Unicode binary values obtained from each character (Unicode value) in the word, which are then summed up and then concatenated with the results of all the words in the whole document to produce the watermark key, W_{KG}. Following this step, the key is registered with the CA. In the decoding stage, the key W_{KE} is extracted based on the characteristics of the input document using the same encoding procedure, Figure 2. Finally, the newly generated key (W_{KE}) is compared with the CA registered key W_{KG} for authentication of the sample with the original-publisher. The pseudo-code implementation pertaining to the encoding and decoding processes are detailed in Algorithms 1 and 2 below.

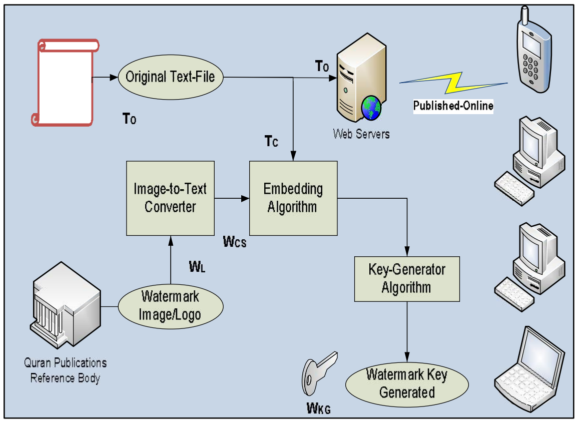

Figure 1. Watermark Encoding Process (Proposed Method A)

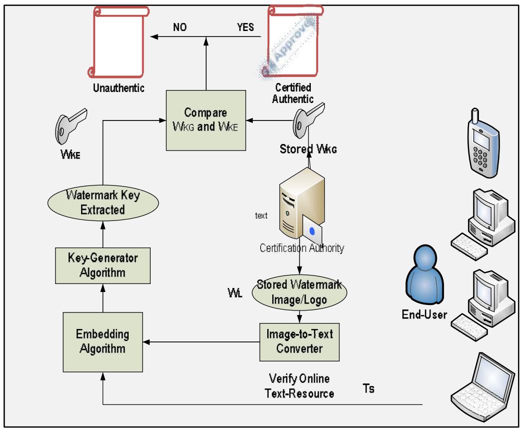

Figure 2. Watermark Decoding Process (Proposed Method A)

Watermark Encoding and Decoding Algorithms for Proposed Method A

```
Algorithm 1: Encoding Algorithm
Input: original cover-document $T_0$, logo-watermark $W_L$
Output: key of watermarked text $W_{KG}$.
1. Convert $W_L$ to a bit-stream, $W_{CS}$.
2. Make a copy-document of the original cover-document, $T_C$.
3. Find the number $n$ of words in $T_C$ .
4. for $i$ = 1 to $n$
        Get $i$-th word from $T_C$.
        Add the Unicode values of all characters of $i$-th
        word and convert the result to binary
    value P.
        Get next embedding-bit $e$ from $W_{CS}$.
        Add $e$ to $P$, and convert the result to decimal $W_{KGi}$
   end for
5. Obtain the key-generated, $W_{KG}$, by combining the result $W_{KGi}$ for
all $i=1$ to $n$ words.
6. Output $W_{KG}$ and the character-stream, $W_{CS}$
```

```
Algorithm 2: Decoding Algorithm
Input: Document under scrutiny $T_S$, Watermark key: $W_{KG,}$ and
original logo $W_L$.
Output: $W_{KE}$, decision-of-authenticity $d$.
1. Convert $W_L$ to a character-bit, $W_{CS}$.
2. Find the number $n$ of words in $T_S$ .
4. for $i$ = 1 to $n$
        Get $i$-th word from $T_S$.
        Add the Unicode values of all characters of $i$-th
        word and convert the result to binary value P.
        Get next embedding-bit $e$ from $W_{CS}$
        Add $e$ to $P$, and convert the result to decimal $W_{KEi}$
   end for
5. Obtain the key-generated, $W_{KE}$, by combining the result $W_{KEi}$ for
all $i=1$ to $n$ words.
6. Compare $W_{KE}$ and $W_{KG}$ for similarity.
7. Output $d$ based on result of comparison.
```

Proposed Method B:

The encoding and decoding processes of Method B are illustrated in Figures 3 and 4, with the pseudo-code implementations detailed in Algorithms 3 and 4, respectively. In this method, no key is used on the cover-document T_0, rather the characteristics of the input document are fully utilized to obtain the watermark key, W_{KG}. Initially, T_0 is divided into n word-sets using a set-size (S). Next, the watermark key for the text document, W_{KG}, is formulated by summing the Unicode values of each character in the set and then concatenating the result of all sets in the whole document. Thereafter, the key (W_{KG}) is registered with the CA. In the decoding stage, the key is extracted based on the characteristics of the input sample document (T_S) using a process largely similar to the encoding process. Finally, the newly generated key (W_{KE}) is compared with the CA registered key (W_{KG}) for authentication of the sample-text with the original publisher.

3. Results

The two proposed methods (Method A and B) were tested on hundreds of documents. From the tested documents, the results for five sample documents were used for comparing between the two proposed methods and three other methods from the literature. The performance results presented in this section are classified into two cost-function metrics before the analysis and assessment of the results were made. Table 1 defines the two cost functions used in this work.

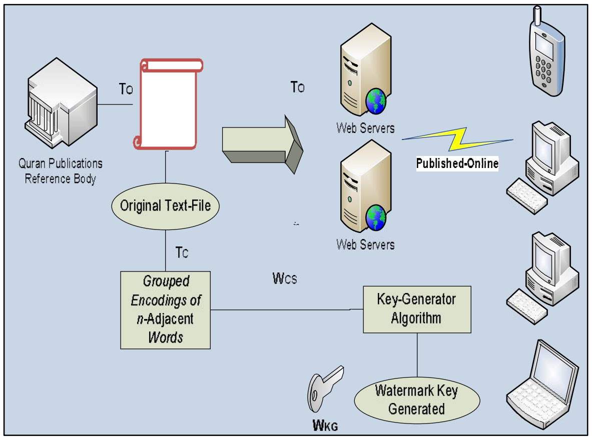

Figure 3. Watermark Encoding Process (Proposed Method B)

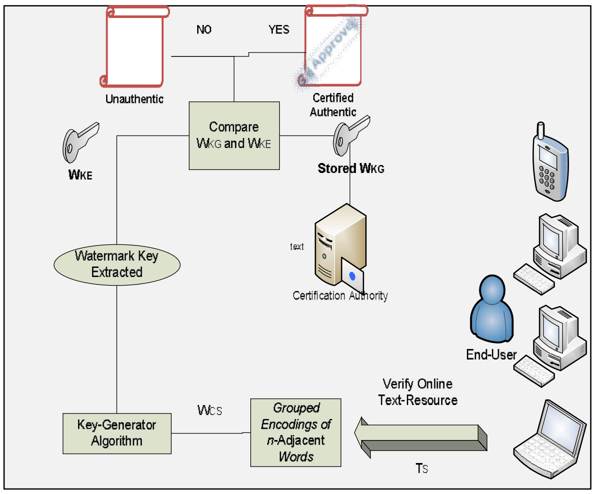

Figure 4. Watermark Decoding Process (Proposed Method B)

Watermark Encoding and Decoding Algorithms for Proposed Method B

```
Algorithm 3: Encoding Algorithm
Input: original cover-document $T_0$, set-size S
Output: key of watermarked text $W_{KG}$.
1. Make a copy-document of the original cover-document, $T_C$.
2. Divide $T_C$ into $n$-word sets according to the set-size $S$.
3. for $i$ = 1 to $n$
        Get $i$-th word-set from $T_C$.
        Add the Unicode values of all characters of all words
  $i$-th set to produce $W_{KGi}$.
   end for
4. Obtain the key-generated, $W_{KG}$, by combining the result $W_{KGi}$ for
all $i=1$ to $n$ words.
5. Output $W_{KG}$
```

Algorithm 4: *Decoding Algorithm*

```
Input: Document under scrutiny $T_S$, Watermark key: $W_{KG}$, set-size S
Output: $W_{KE}$, decision-of-authenticity $d$.
1. Divide $T_C$ into $n$-word sets according to the set-size $S$.
2. for $i = 1$ to $n$
        Get $i$-th word-set from $T_C$.
        Add the Unicode values of all characters of all words
    $i$-th set to produce $W_{KEi}$
 end for
3. Obtain the key-generated, $W_{KE}$, by combining the result $W_{KEi}$ for all $i=1$ to $n$ words.
4. Compare $W_{KE}$ and $W_{KG}$ for similarity.
5. Output $d$ based on result of comparison.
```

Table 1. Definitions of the cost-functions used in the analysis of watermarking methods

Cost-Function Name	Description/Comments
Computational Time	This performance metric provides an insight into the processing complexities involved at each phase.
Percentage Watermark Key-Change	An indicator of the sensitivity of the generated watermark following modifications/tampering in the input document. Typically, higher values indicate a more robust approach tailored for sensitive cover-text documents.

Table 2 and Figure 5 show the comparison for the computational time for five sample documents with different input sizes as the first cost function used in this paper.

Table 2. Computational time comparison of five different methods

File Name	No. of Chars	Computational time [encoder (ms)]					Computational time [decoder (ms)]				
		(Tayan et al., 2013)	(Jalil et al., 2010d)	Yingjis et al., 2010)	Proposed Method A	Proposed Method B	(Tayan et al., 2013)	(Jalil et al., 2010d)	(Yingjis et al., 2010)	Proposed Method A	Proposed Method B
Text1	28915	30	180	210	20	40	20	170	224	10	20
Text2	47974	40	160	350	30	60	30	240	371	20	40
Text3	54839	60	195	410	50	100	40	278	460	40	40
Text4	116794	80	1714	1850	70	190	70	1684	1891	60	80
Text5	166166	130	590	620	120	300	100	520	640	100	130
Average/ char		0.00089	0.00627	0.00832	0.00071	0.00158	0.00065	0.0067	0.00878	0.00052	0.00074

It is clear that both proposed methods provide lower computational time as compared with the other three methods tested for the encoding and decoding phases. This enhancement was due to the lower number of processing tasks used in both systems. However, Method A was faster than Method B since it had reduced algorithmic complexity. On the other hand, the method of (Yingjis *et al.*, 2010) provided the highest computational time since more computations were required to compute the statistical features from the sample documents.

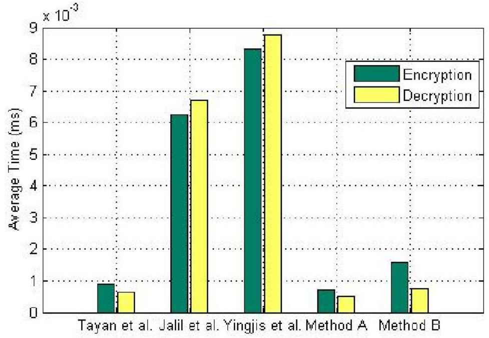

Figure 5: Computational time comparison of five different watermarking methods

During the analysis of a second cost function, three documents of different sizes (small, medium and large) were modified with different percentages and tested to verify the percentage watermark key differences produced as a result of modifications performed on the documents. The results in Tables 3 – 5, provide the percentage changes in the watermark key as a result of modifications made to the text documents. From the results in Tables 3 – 5, it is clear that the statistical based method (Yingjis *et al.*, 2010) provides a closer match in percentage watermark-key modifications corresponding to the percentage modifications made to the text documents for all different document sizes, while the method proposed by (Jalil *et al.*, 2010d) seemed to become unstable as noticed from the large size document. However, with smaller sized documents, the percentage modifications corresponding to the percentage ratio watermark key changed up to an average of 50% for text documents with less than 30000 characters. For documents with 60 – 100 % -modifications, it was noticed that the percentage watermark key change did not show much difference. Therefore, the two proposed methods and the work in (Tayan *et al.,* 2013) had provided consistency with any input sized-sample, while it was noted that with a 100% modified text document, the percentage watermark key modifications was around 55% or less with method A; showing an increase of up to around 75% in the resultant watermark key modifications for large documents only. This shows that these three methods were closely related and that method B seemed to yield the best results for our designated cost function.

Table 3. Percentage Watermark Key Change compared to Percentage modifications in small-size document

Text-1	Original characters	Modified characters	Percentage Modified	Tayan et al.	Jalil et al.	Yingjis et al.	Proposed Method A	Proposed Method B
1	28915	2021	6.98945	2.91	6.486	6.625	3.36	3.22
2	28915	5792	20.0311	8.59	17.29	19.529	9.62	9.85
3	28915	7407	25.6165	11.04	22.7	25.19	12.4	13.09
4	28915	18920	65.4332	28.95	58.37	64.51	32.4	32.51
5	28915	28915	100	48.44	61.08	99.389	49.43	51.89

Table 4. Percentage Watermark Key Change compared to Percentage Modifications in a medium-sized document

Text-2	Original characters	Modified characters	Percentage Modified	Tayan et al.	Jalil et al.	Yingjis et al.	Proposed Method A	Proposed Method B
1	47974	44569	7.09759	11.7477	1.5463	8.984	4.03624	4.227
2	47974	39675	17.299	20.51	94.81	18.99	6.37	9.045
3	47974	23233	51.5717	30.77	95.13	52.26	21.19	25.07
4	47974	11990	75.0073	46.038	96.69	73.73	27.49	35.11
5	47974	0	100	53.08	94.81	96.02	34.89	44.92

Table 5. Percentage Watermark Key Change compared to Percentage Modifications in a large-sized document

Text-5	Original characters	Modified characters	Percentage Modified	Tayan et al.	Jalil et al.	Yingjis et al.	Proposed Method A	Proposed Method B
1	166166	156043	6.0921	3.794	0	7.02	5.71	3.7838
2	166166	119128	28.3078	10.87	3.204	21.41	15.606	11.87
3	166166	85860	48.3288	22.0161	8.06	43	31.32	23.9
4	166166	53450	67.8334	32.995	95.59	64.17	46.49	35.73
5	166166	0	100	50.562	95.28	98.54	71.276	55.06

Experimental results had revealed that the proposed Method-A had achieved superior computational-time performance in the encoding/decoding process, while the proposed Method-B had provided comparatively good performance as compared with the performance from other approaches in the relevant literature. Additionally, both methods proposed in this paper had consistently achieved relatively low "ratios in watermark-key changes" as a consequence of the "modification ratio in the original document" when compared to the related state-of-the-art, providing a good indication of the relative sensitivities of each approach to third-party modifications during the key-extraction phase.

4. Discussion and Conclusion

This paper has proposed and evaluated two new algorithms against other approaches from the state-of-the-art in the text-based zero-watermarking literature. Both proposed algorithms were designed with the goal of protecting online textual-content by providing procedures for tracing back content to their original owners for confirming authenticity, or otherwise to successfully detect forgery and third-party alterations to the digital-content. Additionally, both methods proposed were able to avoid third-party suspicions of publisher-embedded watermarks since no physical watermark-embeddings were performed on the published digital texts. Hence, this work is essential for attempts to avoid/minimize publisher-watermark targeted attacks by any intercepting intermediate-parties involved.

Performance results pertaining to two designated cost functions were obtained and compared with the relevant literature. The proposed methods had achieved improved computational time requirement at the encoder and decoder as compared to the previous approaches. Finally, the evaluation of a second cost-function was considered in order to investigate the effect of any modification on the cover-text on the consequently generated watermark at the encoder. This evaluation was essential for investigating sensitivity-analysis of the different methods and examining the capability of each approach in detecting even slight tampering on the input-text.

Acknowledgements:

The authors would like to thank and acknowledge the IT Research Center for the Holy Quran and Its Sciences (NOOR) at Taibah University for their financial support during the academic year

2012/2013 under research grant reference number NRC1-126.

Corresponding Author:
Dr. Omar Tayan
IT Research Center for the Hoyly Quran and Its Sciences and College of Computer Science and Engineering,
Taibah University
Saudi Arabia
E-mail: otayan@taibahu.edu.sa

References

1. Maxemchuk N. F. and Low S., Marking Text Documents, Proceedings of the IEEE International Conference on Image Processing, Washington, DC, Oct. 26 – 29, 1997, pp. 13 – 16.
2. Adesina, A.O., Nyongesa, H.O., Agbele, K.K., "Digital watermarking: A state-of-the-art review", IST-Africa, 2010.
3. Jalil Z., Mirza A.M., A review of digital watermarking techniques for text documents, IEEE International Conference on Information and Multimedia Technology, pp. 230-234, 2009.
4. Brassil J. T, Low S., and Maxemchuk N. F., Copyright Protection for the Electronic Distribution of Text Documents, Proceedings of the IEEE, Vol. 87, No. 7, July 1999, pp. 1181 – 1196.
5. Huang D. and Yan H., Interword distance changes represented by sine waves for watermarking text images, IEEE Journal Transactions on Circuits and Systems for Video Technology, Volume 11 Issue 12, December 2001, Page 1237-1245.
6. Davarzani R., and Yaghmaie K., Farsi Text Watermarking Based on Character Coding, International Conference on Signal Processing Systems, 2009, pp. 152 – 155.
7. Jensen, C. D., Fingerprinting Text in Logical Markup Languages G.I. Davida and Y. Frankel (Eds.): LSC 2001, LNCS 2200, pp. 433-445, 2001. Springer Verlag, Berlin, Heidelberg.
8. Lu P., An optimized natural language watermarking algorithm based on TMR, Proceedings of 9th International Conference for Young Computer Scientists, 2009.
9. Kaur S. and Babbar G., A Zero-Watermarking algorithm on multiple occurrences of letters for text tampering detection, International Journal on Computer Science and Engineering (IJCSE), Vol. 5, No. 5, 2013, pp. 294 – 301.
10. Jalil Z., Farooq M., Zafar H., Sabir M., and Ashraf E., Improved Zero Text Watermarking Algorithm against Meaning Preserving Attacks, World Academy of Science, Engineering and Technology 46 (2010), pp. 592 – 596.
11. Jalil Z., Mirza A.M., and Jabeen H., Word Length Based Zero-Watermarking Algorithm for Tamper Detection in Text Documents, 2nd International Conference on Computer Engineering and Technology, 2010, Vol. 6, pp. 378 – 382.
12. Jalil Z., Mirza A.M., and Iqbal T., A Zero-Watermarking Algorithm for Text Documents based on Structural Components, International Conference on Information and Emerging Technologies (ICIET), Karachi, 2010, pp. 1 – 5.
13. Jalil Z., Mirza A.M., and Sabir M., Content based Zero-Watermarking Algorithm for Authentication of Text Documents, International Journal of Computer Science and Information and Security 2010, Vol. 7, No. 2, pp. 212 – 217.
14. He L., Zhang L., Ma G., Fang D., and Gui X., A Part-of-speeach Tag Sequence Text Zero-watermarking, Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT '09), *Huangshan, China, 2009, pp. 187-190.*
15. Yingjie M., Tao G., Zhihua G., and Liming G., Chinese Text Zero-Watermarking Based on Sentence's Entropy, International Conference on Multimedia Technology (ICMT), 2010, pp. 1 – 4.
16. Yingjie M., Liming G., Xianlong W., Guo Tao, Chinese Text Zero-Watermarking Based on Space Model, 3rd International Workshop on Intelligent Systems and Applications (ISA), 2011, pp. 1 – 5.
17. Tayan O., Alginahi Y. M., and Kabir M. N., An Adaptive Zero-Watermarking Approach for Authentication and Protection of Sensitive Text Documents, The second international conference on advances in computer and Information technology (ACIT 2013), Kuala Lumpur, Malaysia, May 4 – 5, 2013.
18. Al-Haidari F., Gutub, A., Al-Kahsah, K., Hamodi, J., Improving security and capacity for Arabic text steganography using 'Kashida' extensions, IEEE/ACS International Conference on Computer Systems and Applications, 2009.
19. Aabed M.A., Awaideh S.M., Elshafei A.-R.M., Gutub A.A., Arabic Diacritics based Steganography, IEEE International Conference on Signal Processing and Communications, 2007.
20. Shirali-Shahreza M., Shirali-Shahreza S., Persian/Arabic Unicode Text Steganography, Fourth International Conference on Information Assurance and Security, 2008.

9/12/2013