Text Watermarking Technique For Hindi Language Documents Using Structural Approach

Nitin Namdeo Patil, Jayantrao Bhaurao Patil

Abstract: The ease of availability of Internet in recent decades for sharing ideas and knowledge has motivated authors to publish their contents digitally instead of conventional paperwork. This leads to widespread availability of the author's information to a large volume of readers. But with this global connectivity, many security issues have raised threats to authorship and copyright of genuine authors. The digital contents are vulnerable to illegal copy, distribution, reproduction and authentication. For digital information in numerous natural languages, significant attempts have been made to secure the text from unauthentic access. Each of these concentrate on distinguished language construct(s) to attain uniqueness. In this paper we propose a new robust watermarking technique for Hindi language. Hindi belongs to Devanagari script and no watermarking technique for it is developed till now. We use pratyaya as important Hindi language construct in our structural approach to protect Hindi language digital text across the globe.

Index Terms: Authorship, Copyright, Devanagari, Pratyaya, Robust, Structural Approach

1. INTRODUCTION

NUMEROUS kind of digital information like image, audio, video and text is globally available across the Internet. This electronic form of all kind of information is greatly popular compared to the bulky traditional paperwork. Due to widespread availability, this information is likely to be vulnerable to illegitimate copying, redistribution, authentication, fingerprinting and tampering. Among various existing solutions, digital watermarking has emerged as a prominent solution against several security issues. A general watermarking technique is characterized by properties like robustness, security, capacity and imperceptibility [1, 2]. The digital text watermarking embeds a watermark in the text to uniquely identify the genuine owner of the text. Many researchers have developed text watermarking techniques to enhance security of digital text and to protect intellectual property rights of genuine authors of various natural languages like English, Turkish, Chinese, Arabic and Marathi [3, 4, 5, 6, 7]. For each of these techniques, the authors have appropriately used unique and specific approaches focusing on the features of respective language to make the contents secure from unauthentic access and use. All these techniques can be categorized in one of the following approaches: Image Based, Syntactic, Semantic or Structural. Each of these approaches has uniqueness and shortfalls compared to generic requirement of watermarking technique. But those also have some uniqueness depending upon their application accordingly. Here we focus on Hindi language which is national language of India. Currently no such technique still exists to protect the intellectual property rights of Hindi language text documents. In this paper we propose a new robust watermarking technique for Hindi language. We identify 'pratyaya' (suffix) as a distinguished unique language construct of Hindi text to be used for generation of required watermark. Considering all the facts about Hindi language, we follow the structural approach to develop our proposed text watermarking technique which focuses on text form specifically and avoids the image (copy or scan) format. The structural approach uses logical embedding of text as watermark instead of text alteration [8]. This paper is organized in four main sections: Section 1 gives Introduction, the section 2 focus on Hindi language and its features. The Section 3 describes all the previous work carried out in text watermarking; specifically, in linguistic text watermarking. The proposed algorithms for watermark embedding and extraction are demonstrated in Section 4. Section 5 gives experimental results and discussion. The final section gives summary with conclusions.

- *Author Prof. Dr. Nitin Namdeo Patil has completed his PhD degree in Computer Engineering from Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India & is currently working as Associate professor in Computer Engineering Department, R. C. Patel Institute of Technology, Shirpur (MS) India.*
  E-mail: er.nitinpatil@gmail.com
- *Co-Author Prof. Dr. Jayantrao Bhaurao Patil, is currently working as Principal, R. C. Patel Institute of Technology, Shirpur (MS) India.*

2 ABOUT HINDI LANGUAGE

As per the Constitution of India (Article 343), the use of Hindi in the Devanagari script is the official language of the Union [9]. Hindi speakers across the world are estimated from 150 to 350 million. Hindi is also spoken in some countries outside India, such as in Mauritius, Fiji, Suriname, Guyana, Trinidad Tobago and Nepal. Just like European languages, Hindi has most common linguistic features like other Indo-Aryan languages. Hindi is written from left to right. Hindi is easy to read. Each Hindi character sounds different and one can write Hindi words as per their pronunciation. Comparative to English, Hindi has different word structure. Devanagari script is used to write modern Hindi. It consists of 33 consonants and 11 vowels. All vowels are phonemic and they can be nasalized. Hindi is free order language where no specific word order like English exists. It is also difficult to handle spelling variants in Hindi where a word with same meaning can appear having different spelling variations. The morphological richness of Hindi is additional challenge while dealing with Hindi language words. Mainly the lack of sufficient amount of standard tools and corpus for experiments makes it difficult to experiment with Hindi language text [10]. Several authors have contributed to Hindi language digital content in the form of various resource materials, literature, news articles, government documents and many more on the Internet. The widespread use of Internet has greatly prop up the authors to share their contents across the globe

3 PREVIOUS WORK

The domain of digital watermarking mainly includes image, audio, video and text watermarking. Each of these have contributed with significant research work except text watermarking. The text information is major part of digital content. Maxemchuck et al. initiated research in text watermarking by restricting black and white image redistribution. The individual character or word position/ location is changed to reform the sentence with insertion of watermark [11]. Further modifications were made by Brassil et al. [12] and Low et al. [13]to evaluate this approach. Later on the researchers proposed various methods like line-shifting, word-shifting and character modification by embedding data in text image [13, 14]. Mercan Topkara et al. proposed two fold effort using available tools of Natural Language Processing by analyzing their performance. The authors used sentence structure like characters, words or lines to implement new approach of Natural Language watermarking [15]. C. Culane et al. experimented on multi-set formatted text documents to improve watermarking capacity and robustness in printing and scanning [16]. S. Ranganathan et al. developed three stage technique which performs identification of document type, preprocessing according to the type of document and then appropriately embedding of watermark [17]. Further linguistic text watermarking was elaborated by some researchers to propose language specific text watermarking for some natural languages like Arabic, Chinese, English, German and Turkish. In each technique, unique feature(s) of each language is used in generation of watermark. Yingli Zhang et al. focused on object's special properties of MS-Word file for information hiding. Water- mark created by legal user's information is encrypted and then embedded circularly in the MS-Word's object [3]. Hasan M. Meral concentrated on Turkish language for watermarking. Turkish is agglutinative language which has flexible word order and morphosyntactic structure. Authors used morphosyntactic tools to generate syntax based natural language watermark [4]. Adnan Abdul Aziz Gulab et al. used an extension Arabic character 'kashida' to develop a scheme for Arabic e-text documents [6]. Z. Jalil et al. calculated occurrence frequency of non-vowel ASCII characters and words of English text to propose a technique called as zero text watermarking algorithm [18]. Thus the linguistic watermarking is supported with the language specific constructs to preserve the uniqueness of each implementation. The above discussed approaches of linguistic watermarking attempt to feature all required properties of text watermarking. Although those may vary according to application domain, but the ultimate goal is to protect the concern language text from unauthorized access and illegal dis- semination of the information.

Table 1 Summary of Linguistic Text Watermarking Techniques

Name of Author	Watermark to be Embedded	Time-stamp	Author wise Unique Key	File with Watermark
Author 1	CgmXZb1w7820FNAJDk6Ch6cyMogn4ofgmoepmclaMdleeewOqtEWt1uz1kjqxhu9I3Sqr0FpnpDMzW9jrlWmM2K6mGX6Ulefjafi24iigsogj5ppgpmp1d5rtnKK9seQ	26/8/2015	2upn9sj3	H1.pdf
Author 2	CeNOotTRGhszAPveSGgoqbyYR59scbltp59scbltppwlD3moo9afclxp5gdonpU59mxlIpadlQVgX6Ule+7xOFfzlIS6RVc2z6OwZoP+xV51bUpG+asl45klVnmaIeRl	31/8/2015	4e0b6c79	H2.pdf
Author 3	9qM1jxjrLA6olvdmgQh7oWIEqNzhmM2K6mdR8Y6HS6J+W69QVc2PtDG3kJXymdx5drgbklpqyujano8cvslpp9Ljgzbnm2jalpeUmfhS7lapriovnkaL7edkat6wscvn	16/9/2015	d7eaec3e	H3.pdf
Author 4	o06akTmNB6W6DPb3mwSZJNieflMNAr6bF12DkIFzhoCEisE5DOORxmXcwQpHtsVoIVYPagi0MCu0vNFkT4uqw=ivjsui2oipelasipti5lafjksirpdu+jdl34flfotp	27/8/2015	47f93aac	H4.pdf
Author 5	zVDrD/UDAvgDo1SRKOi3Xh0LrwZHGlDmcpajpDfxqPnh72JbxilYNmyB5F67FwKTmvkOQuI7kpromwdc4rfgbn9asdfg1pkjbv7asdfbnmyhnk5zdfghjlp0tgvcxzAd	29/8/2015	10f8ji56	H5.pdf

Tayan et al. experimented with logical embedding of watermark data rather than physical embedding. In this approach, digital Image is embedded into the document by converting into bits. In embedding phase, the watermark bits are distributed into word sets with variable set size. The key generated using water- mark is registered to CA (Certifying Authority). In the decoding phase, similar key generator is used to generate the key using input characteristics of the document. The generated key and registered key at CA is compared to check the originality and ownership of the document [19]. In the zero-watermarking approach by Jalil et al., watermark gets embedded in host text document without any alteration. The characteristics of text are utilized to generate required watermark. But here, the watermark is of fragile nature which is later used for authentication. This algorithm effectively uses the text characteristics [20]. The enhanced kashida based watermarking approach by Gutub et al. uses pointed letters with extension (Kashida) to hold secret bit 'one' and the un-pointed letters with Kashida to hold secret bit 'zero'. In this method, Kashida can be added be- fore or after the letters. For consistency purpose the location of the extensions should be the same throughout the complete document with watermarking. This algorithm has limitation of not using kashida to all letters as per standard writ- ing structure of Arabic language. The Kashida can only be added in locations between connected letters of Arabic text [21]. Another experimentation using kashida characters for Arabic language text watermarking was proposed by Alginahi et al. In this, the embedding process first calculates the number of words in the documents. If the character contains the special feature then it appends the kashida character else moves to next character. The process is continued till the end of the document. If the end of key reached before the end of the document then key is repeated for the rest of the document [22].

Thus the linguistic watermarking is supported with the language specific con- structs to preserve the uniqueness of each implementation. The above discussed approaches of linguistic watermarking attempt to feature all required properties of text watermarking. Although those may vary according to application domain, but the ultimate goal is to protect the concern language text from unauthorized access and illegal dissemination of the information. Here in Table 1, we summaries the various linguistic text watermarking techniques and their specific approaches with respect to the concern language features.

4 PROPOSED WORK

Understanding the requirement of preserving the authorship of genuine authors, we propose the first digital text watermarking technique for Hindi language text. We used 'pratyaya', a unique feature of Hindi language, to implement this technique. Generally an alphabet or group of letters which are included at the end of any word to form a different word or to change the meaning of the word is called suffix. Pratyaya (suffix) is an important Hindi language construct which in com- bination with 'kriyapad' (verb) can be used to derive 'naam' (noun) to lead the creation of majority of Hindi language words Following are some examples of how pratyaya is formed by combination in Table 2.

Table 2 Example of Formation of Words Using Pratyaya

Pratyaya	Word Formed using Pratyaya
ऊ (oo)	टिकावू (Tikaavoo)
ता (ta)	चलाता (Chalaataa)
ते (te)	घबराते (Ghabaraate)
ती (tee)	समझाती (Samajhaatee)
ना (naa)	लूटना (Lootanaa)
नी (nee)	सुमरनी (Sumaranee)
ने (ne)	ईश्वरने (Ishwarane)
गे (ge)	भुलाओगे (Bhulaavoge)

Our proposed technique consists of two algorithms: Embedding and Extraction. Initially we define an array of pratyaya by randomly selecting fifteen pratyaya from Hindi language. In the embedding algorithm, for the input Hindi text, we first search for predefined fifteen pratyaya for their occurrences. We use them as keywords for watermark generation. These pratyaya are searched in input text file of the author for their occurrence count. The count of each pratyaya is repeatedly added to calculate their digital root. The value of digital root is used to sequentially select the actual pratyaya from our predefined array to concatenate with 128-bit user preferred author ID. This combined string is encrypted using Advanced Encryption Standard (AES) algorithm to generate the encrypted key comprising watermark. By embedding this encrypted key randomly in the input text, the watermarked file containing watermark is created. Thus the text document now contains a digital watermark which can uniquely identify the genuine author. Any text document may undergo attacks like duplication of content, substitution of synonym and reordering of content. These attacks can be categorized as: unauthorized insertion, unauthorized detection, unauthorized deletion, re- ordering attack and combination of the all. In unauthorized insertion, some un- wanted additional information is inserted in the original text. The unauthorized detection is violation of secrecy of detection ability. Whereas in unauthorized deletion, the attacker may delete the content randomly including the watermark from the original author's text. In case of reordering attack, the attacker reorders the author's content to make them unlike the original content. The attacker select the attack volume depending on the targeted intention. The intention may be to use small or large part of the original content. The nature of intended attack may be localized or dispersed. The localized attack means insertion and deletion from particular location like start or end of file. In dispersed attack, the insertion or deletion can take place at multiple locations [18]. In this paper we will be using insertion and deletion attacks for checking the accuracy of the proposed technique. This watermarked document needs to be authenticated by third party to confirm the authorship of genuine author and to resolve any conflict arising about the authorship. The 'Certifying Authority' (CA) is the neutral third party official administration where the author name and the watermark can be registered officially [23]. Our proposed technique facilitates a CA to ensure this authentication with some additional features. The watermarked file generated in embedding algorithm is now registered with our CA. The CA performs time-stamping and issues a unique authorization key to each user. Thus, in authentication process CA maintains an exclusive record of each user at a time including author-name, watermark, time stamp, unique authorization key and the watermarked file itself. This information is sufficient enough to clear any authorship conflict to justify the ownership of original author and to invalidate the unauthentic owner. Table 3 represents some sample entries of original authors with all related parameters. The extraction algorithm facilitates the document authentication by CA. For this the author ID and the text file (either watermarked or attacked) are pro- vided initially. The encrypted key is decrypted to separate pratyaya and author ID. The occurrences of all pratyaya in the provided file is counted and repeatedly added till we get single digit value i.e. the digital root. If this pratyaya and author ID matches with those in originally registered file kept with CA, the author is recognized as genuine owner of the text file; failing to which the authorship claim is denied. The time-stamp and unique authorization key are the supporting features to justify the original authorship. The count of each pratyaya in file is read and digital root of all the occurrences is calculated. The value of digital root indicates the actual pratyaya to be selected to generate concatenated string in combination with author ID. The AES is applied to this string to get encrypted key. If this key and the key registered with CA matches then the authorship is validated. The additional features like the time-stamp, unique authorization key are used to support the original authorship and to deny the claim of unauthentic authors.

Table 3 Sample Entries of Registered Authors

Name of Author	Watermark to be Embedded	Time-stamp	Author wise Unique Key	File with Watermark
Author 1	CgmXZb1w7820FNAJDk6Ch6cy Mogn4ofgmoepmclaMdleeewOqt EWt1uz1kjqxhu9I3Sqr0FpnpDM zW9jrlWmM2K6mGX6Ulefjafi2 4iigsogj5ppgpmp1d5rtnKK9seQ	26/8/2015	2upn9sj3	H1.pdf
Author 2	CeNOotTRGhszAPveSGgoqbyY R59scbltp59scbltppwlD3moo9afc lxp5gdonpU59mxlIpadlQVgX6Ul e+7xOFfzlIS6RVc2z6OwZoP+x	31/8/2015	4e0b6c79	H2.pdf

4.1 Watermark Embedding Algorithm

Variable Declaration:
Digital Root -> Digi_root,
Author ID -> A_{ID},
Pratyaya Array -> P_A,
Array Count -> A_C, a variable to hold generated string value -> KEY,
Compare function - > CMP,
Advanced Encryption Standard -> AES

Precondition:
Hindi Text Document -> D_H,
Pratyaya array's maximum size -> P_{MAX},
Sum and variable i as counter initialized to zero -> Sum=0; and i=0;
Post condition: Embedded Watermark -> E_{WM}

1) Enter D_H, A_{ID}
2) For each value of P_{MAX}
   a) A_C = CMP (P_A, D_H)
   b) P_A = Sum + P_A where $P_A > 9$
   c) Digi_root = $\sum_{i=1}^{9} \mathrm{Sum} + \mathrm{i}$
3) KEY= P_A+ Digi_root + A_{ID}
4) D_H = AES (KEY)
5) Stop.

In the embedding algorithm, a genuine user is given 16-character unique author id. An array pratyaya is predefined for 11 selected pratyaya. The genuine author's input text is searched for the occurrences of respective pratyaya as per predefined array. Further the digital root of total pratyaya count is calculated. The raw key is formed using unique author id and digital root value. This key is encrypted using AES and randomly embedded into the input text. The unique- ness of embedding algorithm lies in selection of pratyaya as special feature of Hindi language text. Also the use of digital root value of actual pratyaya count in formation of embedding key is an additional distinct feature of proposed technique.

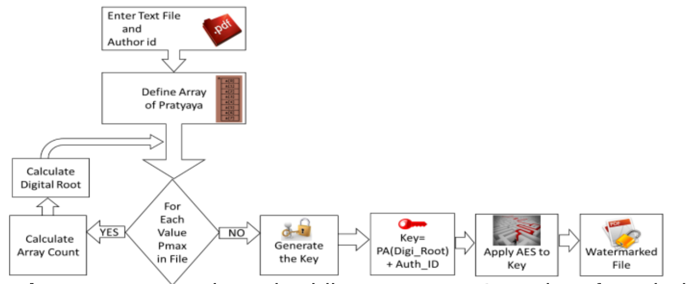

Fig. 1. *Watermark Embedding Process Overview for Hindi Text Watermarking Technique Using Pratyaya*

4.2 Watermark Extraction Algorithm

The extraction algorithm is used by CA to verify the genuine author of any Hindi text with embedded watermark as described in embedding phase. For any such claim of authorship, the AES is used to decrypt the encrypted key to get author id and pratyaya count in form of plain text. From this digital root value any tampering/attack on text document is verified. After comparison function leads to matching of extracted and embedded key. The successful comparison verifies the author as genuine else it denies its claimed authorship. In this extraction algorithm,

Precondition:
Hindi Text Document -> D_H,
Pratyaya arrays maximum size -> P_{MAX},
Sum and variable i as counter initialized to zero -> Sum=0; and i=0;
Post condition: Extracted Watermark -> E_{WM}

1) Enter D_H, A_{ID}
2) KEY = Extract (D_H)
3) EX_{KEY} = AES (KEY) // Apply Advanced Encryption Standard for Key Decryption
4) For i = 1 to P_{MAX}
   a) A_C = CMP (P_A, D_H)
   b) P_A = Sum + P_A where $P_A > 9$
   c) Digi_root = $\sum_{i=1}^{9} \mathrm{Sum} + 1$ while i ≥ 9
5) EM_{KEY} = Digi_root + P_A + A_C
6) If EX_{KEY} = = EM_{KEY}
   Genuine Author
   else
   Authorship denied!
7) Stop

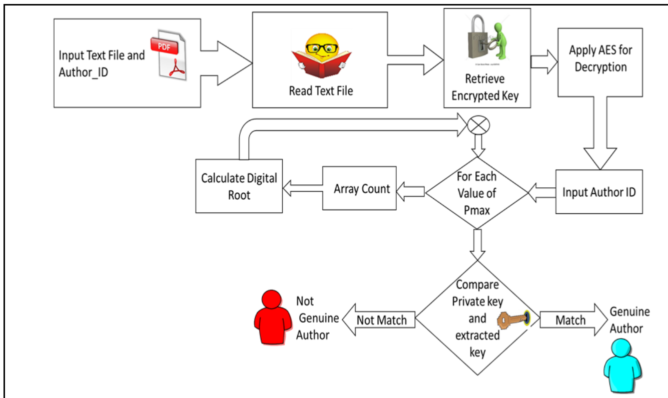

Fig. 2. *Overview of Extraction Algorithm for Hindi Text Watermarking Technique Using Pratyaya*

5 EXPERIMENTAL RESULTS AND DISCUSSION

To validate our proposed technique, we used corpus of Indian Institute of Technology, Bombay and Technology Development for Indian Languages (TDIL) [24]. These corpus need some preprocessing on them like correction in typo- graphical errors. Also we have selected krutidev as default font to input text in author's content. We categorized the contents from both these corpus into four text categories namely: Small Volume Hindi Text (SVHT) with less than 100 words, Medium Volume Hindi Text (MVHT) with words in between 100 and 500, Large Volume Hindi Text (LVHT) which comprise words up to 1000 but not less than 500 and Very Large Volume Hindi Text (VLVHT) containing number of words in between 1000 to 1500. We then tested these text categories for 10 samples of each category against typical types of text attacks like insertion and deletion by varying attack percentage capacity with 5%, 10% and 30%.Our observation shows higher accuracy of retrieved watermarks for small volume of text (SVHT) for insertion and deletion attacks. With increasing text volumes for rest of the categories (SVHT, MVHT, LVHT, VLVHT), the retrieved watermark accuracy is observed as decreasing relatively. Also we can notice from the Table 4 that the deletion attack can moderately destruct text volumes of all categories if compared with the insertion attack. This evaluation of the proposed technique also confirms the four basic and most important properties of watermarking techniques like robustness, security, imperceptibility and capacity. Even though in above discussed attack analysis the considered sample text is attacked with different capacity and type, the higher percentage of retrieved accuracy confirms the robustness of our implemented technique. The proposed technique meets all security parameters where the existence of the watermark and its payload remain secret from any unauthorized access. The imperceptibility feature is achieved by undetectable insertion of watermark. Also it is supported by identical appearance of original and the watermarked file observing same file size with and without watermark. In the implemented technique, the watermark is significantly acceptable with size of 128 bits as an upper limit of embedding which ensures the adequate capacity of our technique.

Table 4 Retrieved Watermark Accuracy for Different Text Categories against Varying Percentage of Attacked Text Volume.

Attacked Text Volume	Attack Type		Localized Attack			Dispersed Attack		
	Insertion Attack (%)		5	10	30	5	10	30
	Deletion Attack (%)		5	10	30	5	10	30
Retrieved Watermark Accuracy	Insertion Attack Catagory	1	98.8	98.1	99.2	99.8	99	99.3
		2	95.5	95.8	97	95.9	97.5	98.1
		3	96.6	96.4	97.4	95.7	96.5	96.4
		4	97.4	96.2	96	95.5	96.8	96.2
	Deletion Attack Catagory	1	98.5	98.2	98.7	98.9	98.8	97.8
		2	97.1	97.6	96.7	96.8	97.1	96.8
		3	96.6	06.5	95.4	95.6	95.4	95.3
		4	95	95.3	94.4	94.7	95.1	95.1

The following graphs in Figure 3 (a-f) illustrate percentage accuracy of retrieved watermarks against different categories of text volumes. The graphical representations of our experimental results confirms that lower the text volume higher will be watermark retrieving accuracy. As the text volume increases the retrieving accuracy goes on decreasing for both the types of attacks. This is valid for 5%, 10% and 30% attacks on the selected text samples respectively. Also we obtain retrieving accuracy in between 90 to 100 percentage for samples of all categories.

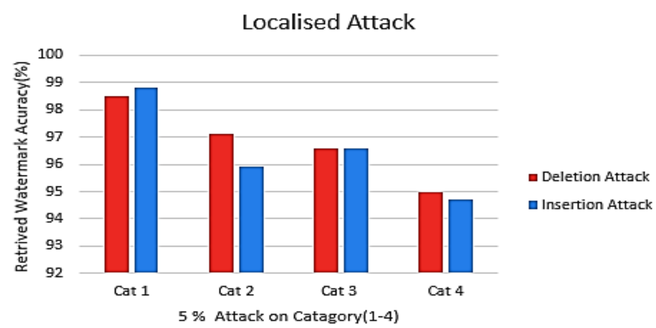

Fig. 3 (a). *Localized Attack (5 %) on Hindi Text*

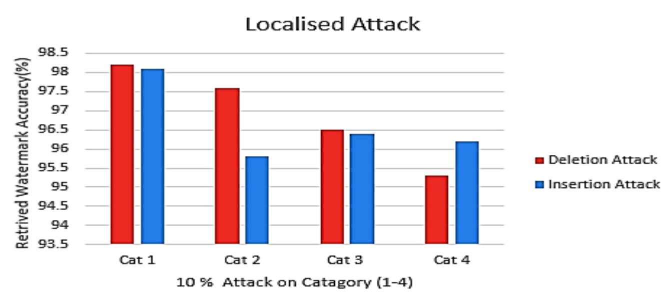

Fig. 3 (b) *Localized Attack (10 %) on Hindi Text*

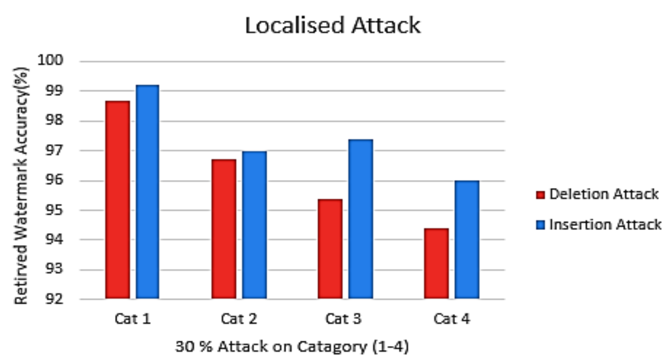

Fig. 3 (c) *Localized Attack (30 %) on Hindi Text*

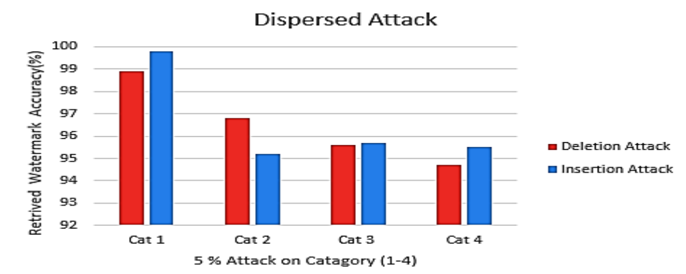

***Fig. 3 (d)** Dispersed Attack (5 %) on Hindi Text*

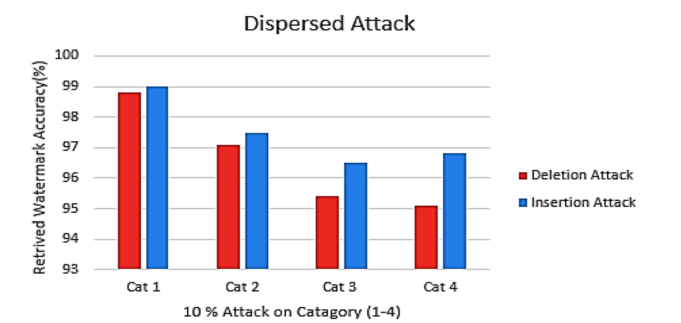

***Fig. 3 (e)** Dispersed Attack (10 %) on Hindi Text*

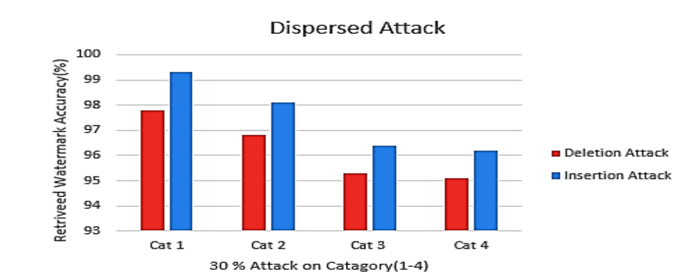

***Fig. 3 (f)** Dispersed Attack (30 %) on Hindi Text*

We compare the retrieved watermark accuracy in case of insertion and deletion types of attack of the proposed technique with that of watermarking technique for English text proposed by Z. Jalil et al. [18]. This comparison is irrespective of size of attacked text volumes. Although the results cannot be compared directly due to different implementation approaches, our technique is found to have better watermark retrieving accuracy. We can further add the details of our proposed technique to the Table 1 described in 'previous work' section of this paper. The novelty is summarized with the proposed way of implementation and crucial attack analysis in Table 5.

Table 5 Summary of Proposed Method

Proposed Method	Text Document Language	Security	Attack Analysis	Conclusion
Watermarking using special construct of hindi language	Hindi	AES Encryption	Analysis done for all possible attacks like insertion, deletion, combined at Localised and Dispersed locations of input text	Novel and Unique for Hindi Language

6 CONCLUSIONS AND ACKNOWLEDGEMENTS

6.1 Conclusions

With global availability of digital text on the Internet, no technique is offered till now to protect authorship and copyrights of genuine authors of Hindi language documents. We have proposed and successfully implemented a digital text watermarking technique for Hindi text which excels in all basic properties of text watermarking. We have tested our proposed technique using standard corpus against possible text watermarking attacks and proved the significant watermark retrieving accuracy. Also we are able to justify the robustness, security, imperceptibility and capacity of the proposed technique. In future we further aim to experiment this technique with our own developed corpus including all possible types of Hindi text.

6.2 Acknowledgements

We the authors of this paper express our sincere thanks to Government of Maharashtra (India) and North Maharashtra University, Jalgaon for providing funding and resources to accomplish this work under the scheme "Research Grant Assistance for Science & Technology Projects (RGSTC)." Also we are thankful to "Center For Indian Language Technology (CFILT) (http://www.cfilt.iitb.ac.in/Downloads.html) Indian Institute of Technology, Bombay" (IITB) and also to "Technology Development for Indian Languages" (TDIL) Programme initiated by the Department of Electronics & Information Technology (DeitY), Ministry of Communication & Information Technology (MC& IT), Govt. of India for providing valuable resources of Hindi Language corpus to experiment with our proposed technique and to evaluate the experimental results.

REFERENCES

[1] F. A. P. Petitcolas, R. Anderson, and M. G. Kuhn, "Information hiding – A survey," Proc. of the IEEE, pp. 1062- 1077, 1999. (IEEE Conference proceedings)

[2] F. Hartung, and M. Kutter, "Multimedia watermarking techniques," Proc. of the IEEE, pp. 1079-1107, 1999. (IEEE Conference proceedings)

[3] Nighat Mir, "Robust Techniques of Web Watermarking Using Verbs, Articles and Preposition", International Journal of Computer Science and Information Security, Vol. 9, No. 2, pp. 248-252, 2011. (Journal or magazine citation)

[4] Hasan M. Meral, Emre Sevinc, Ersin Unkar, Bulent Sankur, A. Sumru Ozsoy, and Tunga Gungor, "Syntactic Tool for Text Watermarking", Proc. of the in International society for Optical Engineering, pp. 904-909, 2007. (Conference proceedings)

[5] Yingli Zhang, and Huaiqing Qin, "A Novel Robust Text Watermarking For Word Document", 3rd International Congress on Image and Signal Processing, pp. 38-42, 2010. (Conference proceedings)

[6] Adnan Gutub, Fahd Al-Haidari, Khalid Al-Kahsah, and Jameel Hamodi, "e-Text Watermarking: Utilizing Kashida Extensions in Arabic Language Electronic Writing," Journal of Emerging Technologies in Web Intelligence (JETWI), pp. 48-55, 2010. (Journal or magazine citation)

[7] Nitin N. Patil and J. B. Patil, "Implementation of a Novel Watermarking Technique for Devanagari Text", International Journal of Information and Electronics

Engineering, Vol. 5, No. 5, pp. 353-355, 2015. (Journal or magazine citation)
[8] Z. Jalil, and A. M. Mirza, ”An Invisible Text Watermarking Algorithm using Image Watermark”, International Conference on Systems, Computing Sciences, and Software Engineering (SCSS 2009), Innovations in Computing Sciences and Software Engineering, published by Springer, pp. 147-152,2009 (Conference proceedings)
[9] Constitution of India (as of 1 December 2007), The Constitution of India. Ministry of Law & Justice. (Constitution)
[10] Bharat R. Ambati, Samar Husain, Sambhav Jain, DiptiM.Sharma, Rajeev Sangal,(2010) “Two Methods to Incorporate Local Morph Syntactic Features in Hindi Dependency Parsing” In Proceedings of the NAACL HLT 1st Workshop on Statistical Parsing of Morphologically-Rich Languages, pages 22-30. (Conference proceedings)
[11] J. Brassil, S.Low, N. Maxemchuk, L. O’Gorman, ”Copyright Protection for the Electronic Distribution of Text Documents,” Proceedings of the IEEE (Invited Paper), Vol. 87 No. 7, pp. 1181-1196, 1999. (IEEE Invited Paper)
[12] J. T. Brassil, S. Low, N. F. Maxemchuk, L. O. Gorman, “Electronic Marking and Identification Techniques to Discourage Document Copying”, IEEE Journal on Selected Areas in Communications, pp. 1495-1504, 1995. (Journal & Conference proceedings)
[13] S. H. Low, N.F. Maxemchuk, A.M. Lapone, “Document Identification for Copyright Protection Using Centroid Detection”, IEEE Transactions on Communications, pp. 372-383, 1998. (IEEE Transactions)
[14] K. Bhattacharjya and H. Ancin, “Data embedding in text for a copier system”, Proc. of the IEEE International Conference on Image Processing, pp. 245-249, 1999. (IEEE Conference proceedings)
[15] Mercan Topkara, Cuneyt M. Taskiran and Edward J. Delp. “Natural Language Watermarking”, SPIE International Conference on Security, Steganography and Watermarking of Multimedia Contents, pp. 441-452, 2005. (Conference proceedings)
[16] Culnane, H. Treharne, and A. T. S. Ho, “Improving Mutliset Formatted Binary Text Watermarking Using Continuous Line Embedding”, 2nd International Conference on Innovative Computing, Information and Control, pp. 287-292, 2008. (Conference proceedings)
[17] Suganya Ranganathan, Ahamed Johnsha Ali, Kathirvel K., and Mohan Kumar M. “Combined Text Watermarking”, International Journal of Computer Science and Information Technologies, pp. 414-416, 2010. (Journal or magazine citation)
[18] Zunera JaliI,Hamza Aziz Saad Bin Shahid Muhammad Arif and Anwar M. Mirza, “A Zero Text Watermarking Algorithm based on Non-Vowel ASCII Characters”, International Conference on Educational and Information Technology (ICEIT 2010), pp. 503-507, 2010. (Conference proceedings)
[19] Tayan, Omar, Yasser M. Alginahi, and Muhammed N. Kabir. ”An Adaptive Zero-Watermarking Approach for Text Documents Protection.” (Journal or magazine citation)
[20] Jalil, Zunera, Anwar M. Mirza, and Maria Sabir. ”Content based zero-watermarking algorithm for authentication of text documents.” arXiv preprint arXiv: 1003.1796 (2010). (Preprint)
[21] Gutub, Adnan Abdul-Aziz, Lahouari Ghouti, Alaaeldin A. Amin, Talal M. Alkharobi, and Mohammad K. Ibrahim. ”Utilizing Extension Character ’Kashida’ with Pointed Letters for Arabic Text Digital Watermarking.” In SECRYPT, pp. 329-332. 2007. (Conference proceedings)
[22] Alginahi, Yasser M., Muhammad N. Kabir, and Omar Tayan. ”An enhanced Kashida-based watermarking approach for Arabic text-documents.” In Electronics, Computer and Computation (ICECCO), 2013 International Conference, pp. 301-304. IEEE, 2013. (Conference proceedings)
[23] Information Technology Act, Vol. 21 of 2000, 2000. (Act)
[24] Technology Development for Indian Languages Programme (Online)