IST-Africa

IST-Africa 2010 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2010
ISBN: 978-1-905824-15-1

Digital Watermarking: A State-of-the-Art Review

Ademola O. ADESINA, Henry O. NYONGESA, Kehinde K. AGBELE
University of the Western Cape, Dept. of Computer Science,
Natural Language Research Group, Cape Town, South Africa
Email: inadesina@gmail.com, henrynyongesa@gmail.com, agbelek@yahoo.com

ABSTRACT: Digital watermarking is the art of embedding data, called a watermark, into a multimedia object such that the watermark can be detected or extracted later without impairing the object. Concealment of secret messages inside a natural language, known as steganography, has been in existence as early as the 16th century. However, the increase in electronic/digital information transmission and distribution has resulted in the spread of watermarking from ordinary text to multimedia transmission. In this paper, we review various approaches and methods that have been used to conceal and preserve messages. Examples of real-world applications are also discussed.

Keywords: Text Watermarking, Linguistic Steganography, Natural Language Processing

1. Introduction

Steganography is the ancient art and science of hiding information by embedding messages within other, seemingly innocent-looking messages. The name steganography name is taken from a work by Trithemus (1462-1516) entitled "Steganographia" comes from a Greek word meaning "cover writing". A good modern example of the use of steganography is in "Prisoners Problem", in which two inmates Alice and Bob wish to communicate across different cells about their breaking jail with the knowledge that the warden, Wendy (who is their adversary) was watching and monitoring their conversations [21]. This example explained further the difference between cryptography (obscured message meaning) and steganography (hiding the facts of the message being communicated). The communication medium is referred to as cover object and the 'stego' object is the embedded message, which together forms the stegosystem. A stego key keeps the operation secured and stego objects cannot be extracted from cover objects within the stegosystem [1, 21].

Natural language processing can be used to implement steganography through, for example, word substitution without changing the meaning of the message. The objective is to obscure knowledge of existence of covert communication, as opposed to decoding a hidden message. Automated steganographic analysis of natural languages remains a very difficult problem in Computer Science thus making automated attacks very hard [5, 6, 9]. The cover objects may be in form of image, audio, graph signals, video, or text. The use of natural language processing for information hiding is categorized as linguistic steganography. It has become more favourable in recent years because of the large volume of communication traffic and the more complex processing used by non-linguistic steganography. This is a logical outcome, especially with the emergence of an e-society in which day to day activities are conducted through the Internet: e-government, e-health and

e-shopping. The security of the communication involved in these transactions remains a concern, necessitating mechanisms to provide continuous data protection. In this paper, we examine the potential of digital watermarking technologies to protect such communication. Digital watermarking of data provides a means to protect information in cases where access control to the information may be compromised.

The remainder of this paper is organized as follows. Section 2 discusses various techniques in linguistic techniques with appropriate examples. Section 3 examines non linguistic steganography techniques and made comparison on the type of watermarking like text, image, video, audio. Section 4 reviews the state-of-the-art under different natural language processing techniques. Section 5 discusses the real-world application of digital watermarking especially to e-HIS (the purpose of our research.) Finally, Section 6 provides conclusion to the paper.

2. Linguistic Steganography

Natural language based information hiding technology relies on modifying information in text documents by manipulating their lexical, syntactic, and semantic properties while preserving the meaning as much as possible. These methods are more robust than techniques that just modify the appearance of text elements like fonts, line space, interword distance etc in achieving texts covert. There are several examples of linguistic steganography, which can be categorised as follows:

a). Null Cipher: Null ciphers applies a series of characters and words intended to confuse a cryptanalyst. The communicated appears as obvious nonsense, but can be decoded to a meaningful message. This is an ancient form of encrypted communication in which a message is surrounded by a large amount of redundant characters (known as null ciphers). This form of communication is, in fact, known to have been used by the German army during WW II. The following is an example of a null cipher form of steganography:

> "Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit.
> Blockade issue affects pretext for embargo on by-products, ejecting suets and vegetable oils".

Decoding this message by extracting the second letter in each word reveals:

> "Pershing sails from NY June 1".

The drawback of this form of steganography is that the message sender is forced to make a text cover according to a preset procedure, hence defeats the purpose of steganography. Also, applying a 'brute force' approach to decoding will reveal the message.

b). Statistical-based [5]: In this approach, a language model, Probabilistic Context Free Grammar (PCFG), is used to generate sequences of words starting from the root node (and recursively applying selected rules from a context free grammar (CFG)). Conversely, a word sequence belonging to the language produced by a PCFG can be parsed to reveal a possible sequence of likely rules that can produce it. Mimicry functions generated cover text using PCFG that has statistical properties close to the normal text. This is achievable by passing on the Huffman code to each grammar rule based on the probabilistic rule. The string is therefore embedded by choosing the grammar rule whose code corresponds to the portion of the message being embedded. The output from mimic functions is garbage, therefore making it suspicious but the combination of mimic functions and CFG improves the readability of the text. The text cover produced is meaningless and semantically incoherent [22]. These shortcomings bring suspicion in covert communications that could be noticed. The following example illustrates generation of cover text for the secret binary message "1011001".

Table 1: A Simple Probabilistic Context-Free Grammar.

The Huffman code corresponding to each rule is also listed.

Rule #	Rule	H. Code	Prob.
1	S → AB	0	0.5
2	S → CB	1	0.5
3	A → She	00	0.25
4	A → He	01	0.25
5	A → Sherran	10	0.25
6	A → Davies	11	0.25
7	B → likes	0	0.5
8	B → detests	10	0.25
9	B → wants	110	0.125
10	B → hates	111	0.125
11	C → Everybody	0	0.5
12	C → The cleaning lady	10	0.25
13	C → A nice kid	11	0.25
14	D → milk.	00	0.25
15	D → apples.	01	0.25
16	D → pumpkin.	10	0.25
17	D → cookies.	11	0.25

Table 2: Generation of Cover Text Using Rules Determined by the Payload

Position	Prefix	Rule	Output string
•1011001	1	2	CB
1•011001	0	11	Everybody B
10•11001	110	9	Everybody wants D
10110•01	01	15	Everybody wants apples.

c) Synonym-based: The original message is hidden through the use of a cover text which is shared between sender and receiver. There are quite a number of the approaches:

1) *Confusion Approach*[13] where typographic errors, abbreviations and acronyms are used for to cover messages. This is achieved through the use of , so called, Computationally Asymmetric Transformation (CAT). The transformation can be carried out easilly, but reversing it may not be semantically easy for humans to carry out but very difficult to automate. The following are examples:

i. “my nopia is kenbro” instead of “my piano is broken”
ii. “lol” as an abbreviation of “lots of love”, “laugh out loud”, or “limits of liability”.
iii. “good to see you” can be “GTCY”, “GTSY”, “G2CY” or “ G2SY”.

One of the ways to make such correction by the adversary is to use spelling checkers to suggest correction and additional information such as previously observed regularity of typos, pronunciation similarity to prioritize the correction the correction list.

2) *Homographic Approach (Synonym substitution)* Chapman et.al. [10] use an algorithm, NICETEXT that is based on a context free grammar (CFG) to build syntactically correct sentences. By running the cover text through part-of-speech tagging, NICETEXT obtains a set of “sentence frames” like nouns, verbs, prepositions, and determiners. NICETEXT fails on the count of semantics because the output text becomes ungrammatical and semantically incorrect. For example, using a word that is spelt the same like another word and might be pronounced the same or differently but which has a different meaning e.g.

i. “Bank” as in “financial institution”, or “the edge of a stream slope” and “the turn of a road”
ii. “Cry” as in “to produce tears as the result of a strong emotion” and “to call out or speak loudly”.

3) *Homophonic Approach* having a word pronounced the same way as another word, but, which has different meaning or a different spelling, for example,

i) “so” and “sew” ii) “Ice cream” and “I scream” iii) “I see” and “IC”.

4) *Replacement Approach (Acronym substitution)* [14] from the synonym set words can be replaced with another word e.g.

i) "he went without water and food for 3 days" can still stand as "he survive without water and food for 3 days" but replacing "went" with "die" may mean semantically different thing entirely

5) *Syntactic Transformation* [9] this creates little effect on the meaning of the sentence e.g.

i) "what!" Ben cried ⟹ "What!" cried Ben

ii) I like God! ⟹ God, i like

d). Noise-based: Topkara et al. [11] described noise-based approach that employs typographic errors and ungrammatical abbreviations in a text, e.g., emails, blogs, forums, etc., for hiding data. For example, unintentional typographical errors (character typing errors such as "teh" instead of "the"), well known abbreviation and acronyms (e.g. using "ur" instead of "you are" or "omg" for "oh my god"), colloquial words or phrases (e.g. "gonna" or "ain't nothin"). Another steganographic scheme is translation-based which keep a message in the errors (noise) that are always encountered in machine translation (MT). A substitution procedure on the translated text by the use of variations of multiple MT systems is used to embed the message. The MT errors that are inserted and uses make the synonyms substitutions to increase the noise. Translation-based techniques cannot be applied to all languages because the fundamental structures are different. This brings severely incoherent and unreadable text, however, another method known as list-based steganography (Listega) was shown to be applicable to several different languages[1]. Shirali-Shahreza et al. [1] have introduced an abbreviation-based scheme to conceal data using the short message service (SMS) of mobile phones. Because of the size constraints of SMS and the use of the phone keypad instead of the keyboard, language called SMS-texting make this practical. But it is also sensitive to amount of noise (errors) that occurs in the users writing. The shortcoming narrow the effectiveness of the scheme and rate of hiding data, unlike Listega that neither employs errors nor uses noisy text to conceal data.

3. Non-Linguistic Steganography

Non-linguistic steganography approaches are characterized based on file type, such as, text, image, audio, and video. Textual steganography hides data by a textual format manipulation (TFM) process. TFM amends an original text by employing different attributes of text like: spaces, misspellings, fonts, font size, font style, colors, and non-color (as invisible ink) to embed an encoded message. However, comparing the original text against the modified text prompts suspicion and enables an adversary to detect where a message is hidden. In addition, TFM can be distorted and may be discerned by human eyes or detected by a computer. Similarly, in image steganography the idea is to manipulate digital images to conceal a message. Generally, image steganography suffers from several issues like distortion, increased image size or limitation of the messages that can be embedded, and the increased vulnerability to detection through digital image processing techniques.

Echo hiding is a form of audio steganography that has applications in providing proof of the ownership, annotation, and assurance of content integrity. Therefore, the data should not be sensitive to removal by common transforms to the stego audio (encoded audio signal), such as filtering, re-sampling, block editing, or lossy data compression. Hiding data in audio signals presents a variety of challenges, due in part to the wider dynamic and differential. These techniques, in general, are very complex, and like their image-based counterparts, are still subject to distortion and are vulnerable to detection. The hidden message may become to a great extent a foreign body in the cover and thus makes those schemes vulnerable to detection. In addition, contemporary steganography schemes rely on

private or restricted access to the original unaltered cover in order to avoid the potential of comparison attacks. In other words, the presence of a hidden messages can be detected by comparing an example of an original message and a modified message [1,5].

4. Review of State-of-the-Art

Linguistic steganography is a relatively new research field consequently there is sparse literature on the subject. *Dung Huang et.al* [13] and Mohan S. K. et. al [14] describe the digital watermarking methods for text documents as limited because of the binary nature of the text documents having space pattern as a distinct feature in text documents, a new approach in text watermarking where interword spaces of different text lines were modified slightly. This gave average spaces of various lines the characteristic of a sine wave which constitutes a mark after the modification. The marks are used as the basis for the shifting codes. Based on this there are three types of text watermarking 1)Line-shift coding, which vertically shifts text lines to encode a document; 2)Word-shift coding, which horizontally shifts the location of words within the text lines in order to encode a document and 3) Feature coding, which selects particular text features and alters their attributes. Hoehn[6], studied algorithms for the sophisticated embedding watermarks in natural texts, using both semantic and syntactic transformations of the original cover text, instead of modifying its appearance. The author suggested techniques for Natural Language watermarking: 1) synonyms substitution- simply means that words are substituted for in the sentence e.g. It is good and change the "good" to "fine" 2) semantic transformation using synonym substitutions (i.e. selected words are substituted by an appropriate synonym), i.e. sentences of the text are first transformed into a corresponding tree structure by applying syntactic transformations where branches and leafs of the tree represent part and objects of the sentence with the root standing as the verb for the sentence, the left branch the subject and the other branches as objects or auxiliary verbs and 3) Semantic transformation using text meaning representation (TMR) - using semantic transformation branches and leafs of the tree represent, the TMRs which describe the semantical meaning of parts of the sentences instead of representing the word. *Adjunct movement,* is taken for example for syntactic transformation. This is moved or added in a sentence e.g. the sentence "The dog often chased the cat" would become "The dog chased the cat often". For semantic transformation e.g. *Pruning* is applied for cutting or copying information from one sentence to another e.g. name of country, the meaning never changed. It is hardly possible to reconstruct the original wording of the represented sentence having only its semantics available.

Meral et al [2] used a morphosyntactic approach in developing a syntax-based natural language watermarking technique. In this approach, unmarked text is initially transformed into syntactic tree structure and their hierarchy and functional dependencies coded. Subsequently, the watermarking software operates on these sentences in a syntax free format and executes binary changes under control of a tool called Wordnet to avoid semantic drops. The Turkish language was used in the research. Turkish is an agglutinative language in which morphemes (phonemes) are systematically added to base words. For example, *Avustralyalılaştıramadıklarımızdan* is pronounce as one word but in translated into English it means "one of those whom we could not make resemble the Australian people". Turkish was, thus, appropriate for a syntax-based natural language watermarking because of its relatively free word order possibilities and rich collection of the morpho-syntactic structure. In Mercan et al [8] identified the low embedding bandwidth, application to all sentences, relatively small alternative forms of sentence are all factors that bring limitations and control to the grammar and vocabulary of natural languages, because of the limitations and the requirement to preserve style and fluency. The paper suggested lower-level marking as a tool to improve the flexibility and embedding properties of higher levels (i.e. from word-based to sentence-based). This is achievable by using word-based methods

separately from the sentence-based methods. The sentence level watermarking technique relies on multiple features of each sentence and exploits the notion of orthogonality between these features.

e.g. Two transformations can be performed on the same sentence as follows:

Original : he said nigeria and south africa recently rejected the idea

After passivization : he said the idea was recently rejected by nigeria and south africa

After adjunct movement : he said the idea was rejected recently by nigeria and south africa

Kankanhalli *et al* [14] addressed the problem of illegal transmission or copying of documents by proposing electronic text documents fingerprinted with a few semantics-preserving modifications to the documents text. The text modifications are done from multiple copies of the same master document so as to have the same meaning. The unauthorized copy can be identified from the authorized source and the recipient by proper examination of the text modified. The technique was shown to be accurate, robust against attacks, scalable and secure. Meaning preserving modifications of cover text are considered favourable for any natural language text information hiding technique. Tokpara *et a*l [12] observed that previous natural language text information hiding schemes did not use numeric quantification of the distortions brought about by transformation, but mainly used heuristic measures of quality based on conformity to a language model that have no reference to the original cover text. Ambiguous synonym alternatives are favoured when there are many words to be transformed through synonym substitution. Maximum ambiguity makes us move closer to distortion limits, which this discourages further transformation without exceeding the damage threshold.

Tokpara et al [9] discussed the use of natural language text structure of sentence constituents to insert a watermark. This is different from text watermarking which embeds information by modification of the appearance of text. In this research, the authors showed how techniques used in image watermarking domain may be applicable to the natural language watermarking domain. Embedding information in the syntactic structure of sentences is the most promising technique of all natural language watermarking techniques. Evaluation of Natural Language watermarking is more tasking than other types of watermarking because of the issues of meaning, grammar and text style. No known algorithm or studies on the robustness of Natural Language watermarking schemes for objective assessment on human perception is available.

5. Real-World Applications of Digital Watermarking

As long as there is need for data movement there will be need for data security. The importance of digital-watermarking, therefore, cannot be over-emphasized. Several of the core thematic priorities considered in IST 2010 Conference bear witness to this: from health to security to digital libraries. Watermarks are often inserted into images that can be detected when the image is compared with the original. These watermarks used for copyright protection are designed to identify both the source of the image as well as its authorized users. Public key encryption, such as the RSA algorithm, does not completely prevent unauthorized copying because of the ease with which images could be reproduced from previously published documents. All encrypted documents and images must be decrypted before the inspection. After the encryption is taken off, the document can be readable and disseminated. In healthcare, for example, there is a need to secure the confidentiality of patient information and data. Chang-Tsun *et.al.* [4] presented a role-based access control framework using data hiding techniques for combating security threats in a Pictures Archival Communication System (PACS). Access to the databases and the information contained in the pictures, in this case mammograms, are controlled through the

issuance of a stego-key and a watermarking key. Catherine Quantin *et. al.* [3] discussed about the problem of health professionals accessing information that is distributed over different medical records and in different locations as a steganographic filing system was designed, with the intention to conceal the existence of any files. In this case, users were required to be aware of the existence of the file and to supply a file name and associated password to access the desired file. The system uses initialization of the file system with several randomly generated cover files. Newly created object is embedded as the exclusive or of a subset of cover files.

The idea of embedding an invisible watermark to identify ownership and recipients has attracted many interests in the printing and publishing industries. Copyright is a form of intellectual property that gives the author of an original work exclusive right for a certain time period in relation to that work, including its publication, distribution and adaptation, after which time the work is said to enter the public domain. Copyright applies to any expressible form of an idea or information that is substantive and discrete and fixed in a medium. Some jurisdictions also recognize "moral rights" of the creator of a work, such as the right to be credited for the work. Copyright is described under the umbrella term intellectual property along with patents and trademarks. Digital video can be copied repeatedly without loss of quality. Therefore, copyright protection for video data is more critical in digital video delivery networks. One copyright protection method is to add a watermark to the video stream that carries information about the sender and recipient. In this way, video watermarking can enable identification and tracking of different copies of video data. It can be applied to video distribution over the Internet, pay-per-view video broadcasting, and video disk labeling [1]. Through the mechanism of digital watermarking, copyright of digital materials were able to be ascertained.

Steganography is, in fact, used by some modern colour laser printers, whereby tiny yellow dots are added to each page. The dots are barely visible and contain encoded printer serial numbers, as well as date and time stamps [20]. Covert channels in Transmission Control Protocol/ Internet Protocol (TCP/IP) involve masking identification information in TCP/IP headers to hide the true identity of one or more systems. This can be very useful for any secure communications needs over open systems such as the internet when absolute secrecy is needed for the entire communication process and not just one document [1, 15]. There is always improvement in this sector, both at the proprietary level and open source basis and is a model that could be used to watermark text for electronic transfer of fund was proposed **[**15, 16**].**

6. Conclusion

The events of September 11, 2001 created an increased awareness of terrorism, and also raised interest in the ways the terrorists may have been communicating before the attack. Steganography is an obvious method that may have been used to send hidden information through Internet emails, blogs and web pages. [16, 19]. Digital watermarking similarly has played a significant role in the protection, validity, integrity, uniqueness, traceability of electronic documents, prohibition of the illegal duplication and illegal tampering with application documents ranging from security, e-government, financial institution, e-voting system, e-HIS, copyright, e-commerce.

The research reviewed in this paper shows the potential of the many techniques that have been studied recently, but, also highlights weaknesses in real-world applications. Many of the techniques are as yet not robust enough to prevent detection and removal of embedded data. Notably, the quality of the media should not noticeably be degradable upon addition of a mark, marks should be undetectable even in the presence of the payload message, multiple marks in a payload should not interfere with each other, marks should survive attacks and most importantly digital watermarks should not degrade the payload

message. Hence, it is suggested that implementation of digital watermarking should be complemented with data encryption mechanism to improve the assurance and integrity of data stored, retrieved or transmitted across electronics devices. Our research is in the realm of healthcare informatics and patient medical records. In this regard, it is vital both patient and health workers will have confidence in the confidentiality of the information and data, and integrity of the transmission channels.

References

[1.] Abdelraham Desoky. *Listega: List Based Steganography Methodology;* Int. J. Inf. Secur. (2009) 8: 247-261 DOI 10.1007/s 10207-009-0079-0

[2] Hasan M. Meral *et. al.*Natural Language Watermarking via Morphosyntactic alterations: Science Direct Computer Speech and Language 23 (2009) 107- 125

[3.] Catherine Quantin *et. al. New Advanced Technologies to Provide Decentralised and Secure Access to Medical Records: Case Studies in Oncology,* Cancer Informatics 2009: 7 217 -229 available from http://www.la-press.com

[4.] Chang-Tsun *et.al.* Protection of Digital Mammograms on PACSs Using Data Hiding Techniques: Department of Computer Science, University of Warwick, UK

[5.] Cuneyt M. Taskirana, Umut Topkarab, Mercan Topkarab, and Edward J. Delpc : *Attacks on Lexical Natural Language Steganography Systems* Motorola Labs, Multimedia Research Lab, Schaumburg, Illinois 60196 , Center for Education and Research in Information Assurance (CERIAS)

[6.] Helge Hoehn, *Natural Language Watermarking,* Seminar Series *Selected Topics of IT Security*, summer term 2007, Faculty of Security in Information Technology, Technische Universität Darmstadt hhoehn@konaktiva.tu-darmstadt.de

[7.] Aniket M. Nanbe et. al "Improved Synonym Approach to Linguistic steganography" Design and Proof of Concept Implement

[8.] M. Topkara, U. Topkara, M. J. Atallah, *Words are not enough: Sentence level natural language watermarking*, *MCPS'06*, Santa Barbara 2006

[9.] Mercan Topkara *et. al. Natural Language Watermarking* Center for Education and Research in Video and Image Processing Laboratory (*VIPER*) Information Assurance (CERIAS) School of Electrical and Computer Engineering Purdue University Purdue University West Lafayette, Indiana, 47907 West Lafayette, Indiana, 47907

[10.] Mikhail J Atallah *et. al. Natural Language Watermarking and Tamperproofing* CERIAS, Center for Education and Research in Information Assurance and Security, IN 47904, USA

[11.] Mercan Topkara et. al. Information Hiding Through Errors : A Confusing Approach Center for Education and Research in Video and Image Processing Laboratory (*VIPER*) Information Assurance (CERIAS) School of Electrical and Computer Engineering Purdue University Purdue University West Lafayette, Indiana, 47907 West Lafayette, Indiana, 47907

[12.] Umut Topkara *et.al. The Hiding Virtues of Ambiguity: Quantifiably Resilient Watermarking of Natural Language Text through Synonym Substitutions* Department of Computer Sciences Purdue University West Lafayette, IN, 47906, USA utopkara,mkarahan,mja@cs.purdue.edu

[13.] Ding Huang and Hong Yan *Inter-word Distance Changes Represented by Sine Waves for Watermarking Text Images.*

[14.] Mohan S. Kankanhalli *et. al.* (2002)*Watermarking of Electronic Text Documents,* Electronic Commerce Research, 2: 169–187 Kluwer Academic Publishers. Manufactured in the Netherlands.

[15] SANS Institute Infosec Reading Room: A Detailed look at Steganographic Techniques and their use in an Open Systems Environment

[16] SANS Institute Infosec Reading Room: Hiding plain View: Could Steganography be Terrorist Tool?

[17] Frank Y. Shih, Digital Watermarking and Steganography, Fundamental and Techniques; CRC Press Taylor & Francis Group (2008) ISBN-13: 978-1-4200-4757-8

[18] Liu xiaowei, The application of fragile watermark in e-governance; department of computer science, Technological academy of Jiangxi Normal University of Science and Technology, Nanchang, Jiangi China

[19] Maura Conway: Code Wars Steganogragraphy, Signals, and Tourism: Dept of Political Science 1, College Green Trinity College Dublin 2 Ireland

[20] www.http://osdir.comaccessed last Nov 2, 2009

[21] Christian Cachin 2005, *Digital Steganography,* IBM Research Zurich Research Laboratory CH-8803 Ruschlikon, cca@zurich.ibm.com

[22] Wayner, P.: Mimic functions. Cryptologia XVI(3), 193-214 (1992). doi: 10.1080/0161-119291866883