PRINTED DOCUMENT FORGERY DETECTION USING TEXT REORDERING AND MIXING OF MATRICES IN ZERO WATERMARKING

AFFANDI HUSAIN

A project report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Computer Science (Information Security)

Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2015

This project report is dedicated to my beloved parents, wife and my family for their endless prayer, support and encouragement.

ACKNOWLEDGEMENT

First and foremost, all praise to Allah SWT, the Almighty and the All Great, Most Beneficent and the Most Merciful for His guidance and mercy. Peace and blessings of Allah SWT to our beloved prophet Muhammad S.A.W (PBUH). I would like to express my appreciation to my supervisor, Associate Professor Dr. Subariah Ibrahim for her time and effort in teaching and guiding me throughout this project. Without her guidance, this project would not be completed. I also would like to give my heartfelt gratitude to my beloved parents, Hj. Husain Perumal and Puan Noraini Chong, my beloved wife Puan Roslina Ridzwan and also my family for their endless prayer and duā that gives me strength and motivation during the hard time and tolerating with my busy schedule. Finally, I would like to extend my appreciation and thanks to my fellow friends and lecturers of Faculty of Computing, Universiti Teknologi Malaysia for their support and contribution towards the completion of this thesis.

ABSTRACT

Printed documents are still needed in our daily life even though we are living in digital era. Information either in printed document or in digital form must be protected from threats and attacks such as forgery or unauthorized modification. Such threats makes the document lose its integrity and authenticity as well as the value of the information. There are several methods that have been used and created to maintain authenticity, integrity and detecting forgery of printed documents. However some of the methods are not suitable for public use due to its complexity, hard to obtain special materials to secure the document and expensive. This project studies on several text watermarking methods that have been used for document security. There are four main approaches in text watermarking with their own advantages and disadvantages. Zero watermarking is another simpler yet effective method to verify the integrity and to detect forgery of text document which can be used as an alternative to text watermarking. Based on the studies that have been conducted, a zero watermarking algorithm was proposed to improve weakness found in one of the content-based zero watermarking algorithm. This document forgery detection solution uses text content to generate watermark by implementing text reordering and mixing of matrices. The generated watermark is registered into the trusted third party instead of embedding the watermark into the text document. The generated original watermark stored in the trusted third party is compared with extracted watermark from the received text document later on for verification and forgery detection. A performance analysis of the proposed zero watermarking algorithm showed that the forgery detection accuracy of text document is improved from the average of 70% detection to 100% detection based on the content-based zero watermarking algorithm. The execution time of the proposed algorithm is calculated to be less than one second which can be concluded that the execution time performance is almost similar to other watermarking algorithm.

ABSTRAK

Dokumen bercetak masih diperlukan dalam kehidupan harian kita walaupun kita hidup dalam era digital. Maklumat bercetak atau digital perlu dilindungi daripada sebarang ancaman seperti pemalsuan atau pengubahsuaian yang tidak dibenarkan. Ancaman-ancaman tersebut menyebabkan integriti, kesahihan dan nilai maklumat dokumen teks tergugat. Terdapat beberapa kaedah yang telah digunakan untuk melindungi kesahihan, integriti serta mengesan pemalsuan dokumen bercetak. Namun begitu, sebahagian besar daripada kaedah tersebut tidak sesuai untuk digunakan secara umum disebabkan oleh kerumitan penggunaan, kesukaran mendapatkan bahan dan kos yang tinggi. Kajian ini menganalisa penggunaan beberapa kaedah tera-air teks dalam menjamin keselamatan dokumen. Terdapat empat pendekatan dengan kelebihan dan kelemahan masing-masing. Kaedah tera-air sifar merupakan kaedah yang lebih mudah dan berkesan bagi menjamin integriti dan mengesan pemalsuan dokumen teks yang boleh digunakan sebagai alternatif kepada kaedah tera-air teks. Berdasarkan kepada kajian tersebut, algoritma tera-air sifar telah dicadangkan untuk menangani kelemahan di dalam salah satu algoritma tera-air sifar berasaskan kandungan. Algoritma ini menggunakan kandungan teks untuk menjana tera-air dengan penyusunan semula teks dan penggabungan matriks. Tera-air yang dijana didaftarkan dengan pihak ketiga yang dipercayai tanpa memasukkan tera-air tersebut ke dalam dokumen teks asal. Tera-air asal dibandingkan dengan tera-air yang diekstrak daripada dokumen teks yang diterima untuk pengesahan dan pengesanan pemalsuan. Analisis prestasi algoritma yang dicadangkan menunjukkan bahawa algoritma tersebut dapat memperbaiki ketepatan pengesanan pemalsuan dokumen dari purata ketepatan 70% kepada 100% berbanding dengan algoritma tera-air sifar berasaskan kandungan. Analisis masa perlaksanaan algoritma tersebut mendapati algoritma tersebut memerlukan masa perlaksanaan kurang daripada satu saat di mana prestasi masa perlaksanaan algoritma tersebut adalah lebih kurang sama dengan algoritma tera-air yang lain.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2MOL	-	Second maximum occurrence letter
AES	-	Advance Encryption Standard
ASCII	-	American Standard Code for Information Interchange
BCH	-	Bose-Chaudhuri-Hocquenghem
BP	-	Back propagation
CA	-	Certification Authority
CUEPACS	-	Congress of Unions of Employees in the Public and Civil
DCT	-	Discrete cosine transform
DFT	-	Discrete Fourier transform
DLL	-	Dynamic Link Library
Dpi	-	Dots per inch
DRBG	-	Deterministic random bit generator
DWT	-	Discrete wave transform
GHz	-	Gigahertz
ICR	-	Intelligent character recognition
ICT	-	Information and Communication Technology
HCSN	-	Health Care Solutions Networks Inc.
HMM	-	Hidden Markov model
HVS	-	Human visual system
ID	-	Identification
LSB	-	Least significant bit
LST	-	Large Size Text
MAC	-	Message Authentication Code
MD5	-	Message Digest 5
MOFL	-	Maximum occurring first letter
MST	-	Medium Size Text

NHCAA	-	National Health Care Anti-Fraud Association Services
NLP	-	Natural language processing
OCR	-	Optical character recognition
OLE	-	Object Linking and Embedding
PMP	-	Pattern matching percentage
PPC	-	Print, Paste and Copy
PRNG	-	Pseudo random number generator
REI	-	Reverse Engineering Imitation
RGB	-	Red, Green, Blue
RSA	-	Rivest Shamir Adleman
SEP	-	Scan, Edit and Print
SHA	-	Secure Hash Algorithm
SSM	-	Spread Spectrum Modulation
SST	-	Small Size Text
TDR	-	Tampering distortion rate
TIFF	-	Tagged Image File Format
TMR	-	Text meaning representation
VA	-	Verification Authority
WAR	-	Watermark accuracy rate
WDP	-	Watermark distortion percentage
WDR	-	Watermark distortion rate
WMP	-	Watermark matching percentage
WPM	-	Watermark pattern matching

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Data is a collection of facts and statistical that have been used for reference, analysis and as knowledge. Data has been preserved in many kind of medium and forms, in digital and analog world. Humans preserved data or information since the beginning of this world in many forms and materials such stones, bones, woods, bamboo, silk and so on. Then people started to use material called paper that has been invented by the ancient Chinese in the second century A.D. to imprint memory or information to preserve it (Biermann, 1996). The manufacturing and use of paper spread throughout the world in 10^{th} century that begins in Europe and there have been a lot of techniques used to write information into the paper such as handwriting or printing. Even until now people are still preserving data in the analog world using paper documents such as newspaper, books, printed records and much more. But in these few decades back, the world were introduced with Internet. From that moment, data has been 'digitized' into the digital world due to the effectiveness and easiness off conveying data or information over the Internet. The world become more and more paperless and the growth on ICT makes the data preservation in digital form become popular. However, in this digital era we are still using printed documents to represent information such as birth certificates, educational transcript and certificates, land titles, official letters, wills, contracts and many more.

Data or information need to be protected as it is a valuable element in life. People can manipulate data to do something which is damaging to the real world. Information must be secured from attacks or unauthorized access. Security main goals is to maintain confidentiality, integrity and availability of assets, in this case data or information. In this modern age, a lot of investments goes to protecting digital data rather than the printed documents as the threats on digital world is much more severe and growing fast since the introduction of Internet. There are a lot of developed methods that have been used in protecting the digital data and one of available techniques that is widely used is cryptography. Cryptography is a practice of methods that converts data from readable form into something that cannot be read by unintended third parties. In contrast with cryptography, there is also a method which conceal information or messages instead of converting it known as steganography. There are also watermarking methods that have been used for document authentication and forgery detection using several elements such as image, audio and text. But all of this methods are used to secure data in digital but rarely used for printed data. Data in analog world are the same with the digital world, they are susceptible to attacks. Main threats on printed documents such as counterfeiting and forgeries are common threats to printed document security over these years. Printed documents are often being forged, altered or faked to deceive intended parties who thought that the documents are real in order to gain benefits from it such as money, contracts and even freedom (Netto and Carter,2013).

1.2 Problem Background

In each part of this world, the usage of printed documents are still relevant. Both digital and printed documents need security protection. The information or data inside printed documents are susceptible to threats such as forgery and counterfeiting. These attacks on printed documents are successful even though there are few methods available to secure the documents. This is also due to the lack of practical authentication and forgery detection methods that can be use generally on all kinds of printed document. We are still hearing news on document forgery cases all over the

world. In United States of America, document forgery is one of the common problems faced by the authorities and private companies especially on health insurance where The National Health Care Anti-Fraud Association (NHCAA) estimated that United States of America lost 3% to 10% of total healthcare cost to fraud (Simborg, 2008). One of the healthcare insurance fraud case related to healthcare document forgery happened where former medical record director of Health Care Solutions Networks Inc. (HCSN), Wondera Eason found guilty of conspiracy on 25 April 2013 that resulted a total loss of USD63 million on Medicare and Medicaid Savings Programs in Florida. The company billed illegal claims to Medicare and Medicaid Florida between 2004 and 2011 based on thousands of healthcare document alteration, fabrication and forgery which have been oversaw by Eason (U.S. Department of Justice, 2013). Another case regarding document forgery reported on 18 October 2013 in Orlando, Florida where two inmates that were life sentenced without parole for murdering had mistakenly released by the guards at Franklin Correctional Institutes after being shown with forged documents. The falsified documents contains forged motions that appealed for sentences reduction of Charles Walker and Joseph Jenkins and also a court orders that granted the request and release of the inmates. A statement given by Jeffrey L. Ashton on behalf of Ninth Circuit State Attorney mentioned that the forged motions documents forged with fake signature of Orlando-state attorney or the assistant state attorney. While the court orders were filed with Orange County Clerk's office looks legitimate with the county's seal and forged signature of nationally famous Judge Belvin Perry. As the result, Jenkins released on 27 September and Walker freed on 8 October (Netto and Carter, 2013).

There are also document forgery cases reported in Malaysia, especially related to the government sectors. In one of the mainstream newspaper, Berita Harian (2010) reported that a statement given by the Congress of Unions of Employees in the Public and Civil Services (CUEPACS) stated that more than 45,000 or 3% of 1.5 million government's staff in Malaysia forged medical certificate as a reason to absent from work. The president of CUEPACS also said that some of the government's staff falsified the medical certificate to do part-time jobs. However, Director-General of Public Service Malaysia at that time said Public Service Department have yet to receive any report from the agencies regarding the percentage of staff that falsify the

medical certificate. There is another case of document falsification that gained national attention in 2012 where former Director of State Veterinary charged by the Seremban's Sessions Court on falsifying documents for the purpose of promotion since 2009. The former director faced four charges on falsifying documents for the purpose of promotion and adjustment of her salary and allowances between 16 March 2009 and 24 July 2011. The former director have been accused of giving three falsified reference letters of her promotion and post confirmation from a Grade G44 Veterinary Officer to a Grade G54 Veterinary Officer from the department's Putrajaya head office in three different occasion for administrative processes. By doing this the former director wrongfully obtained a total of RM50,315.58 in emoluments (Jaafar, 2012). Anoher recent fraud case related to document forging is when Johor Immigration Department successfully busted a syndicate counterfeiting the social visit passes on 05 November 2013. The syndicate is believed to have use a special ink obtained from India to print the fake social visit passes and sold it at RM8,400 each to the foreigners who were seeking for employment in Malaysia. This operation has been done by an Indian national aged in his thirties for almost two years before caught red-handed in a raid by the Immigration enforcement team in Taman Sri Putri, Skudai. The fake document produced looks like a genuine social visit pass, but Immigration's trained officers are able to detect the defects and deficiencies on the fake social pass (Kim, 2013). Other possible forgery and falsification cases that might have occurred are a forged government contracts with vendors where the vendor might have deleted certain legal quotes or changed the value of the contract that bring advantages to them. Another possible case is forgery of official letter to gain sensitive information or the right of supplying goods to ministries or departments in the government.

There are few initiatives that have been taken to strengthen the document security, and yet document forgery cases are still occurring. In digital data security, the integrity protection provided by cryptographic hash algorithm such as Secure Hash Algorithm (SHA), Message Digest 5 (MD5) and Message Authentication Code (MAC). While the authentication that verifies the originality of the data can be achieved by applying digital signatures and challenge-response authentication. There are also watermarking methods using image, audio and text to authenticate the integrity of the digital information and for copyright protection. But for the

authentication printed documents are something that is still consider a challenge in information security. Security printing is the field in printing industry that prevents forgery, tampering or counterfeiting of printed documents or items such as banknotes, certificates, passports and others. Security printing uses few technical techniques, for example the usage of special paper like cotton fiber and polymer, special watermarks that looks lighter or darker when viewed with lights from behind, intaglio printing that makes images in the printed document raised, microprinting, optically variable color-changing inks, holograms, security threads, halo that created images that can be hidden in the background or picture on the document and prismatic coloration which blends two or more color together to create prismatic effect.

There are also several other methods that have been introduced which are cheaper and effective. Integrity verification using two-dimensional barcode have become one of the alternatives as the barcode is able to store information that being used in the text document verification. But one of the disadvantage in integrity verification using barcode is that it can only carry a small amount of information and spatial resource issue of the barcodes. Watermarking method is the widely used method for document integrity verification and forgery detection. Watermark is a perceptible pattern or image in document that carries key information regarding the copyright or description of authenticity of the document. In printed watermarking, there are two available categories which are image and text watermarking. Image watermarking uses image as the medium to carry watermark and have two different processing methods, spatial-domain and transform-domain techniques. Each of the methods have different set of techniques that can be used in hiding the watermark such as print-and-scan image watermarking using discrete cosine transform (DCT) by Agani *et al.* (2013) and color printed document watermarking using color modulation and two dimensional fast Fourier transform (Mayer and Simske, 2012). Another category is in printed watermarking is text watermarking. Text watermarking uses text as the medium to carry the watermark. There are four approaches in text watermarking that have been discussed in the next chapter which have their weaknesses that led this study to explore another text watermarking method. Text-based zero watermarking method is another less-complex and newer method in text watermarking. This method uses text content in the printed document to generate watermark pattern for forgery

detection (Jalil *et al.*, 2010a; Kaur and Babbar, 2013). However, there is a weakness in the zero watermarking algorithm proposed by the authors that enables the attacker to avoid the forgery detection. Each printed watermark methods have their own advantages and disadvantages that will be elaborated more in Chapter 2.

1.3 Problem Statement

There are several methods that can be used to detect printed document forgery and falsification. However, these methods required special printing machines, materials and sufficient knowledge to implement these methods in everyday printed documents which is unsuitable to normal users. One of the most widely used method is watermarking method. This study is focused on forgery detection accuracy of text-based zero watermarking for printed text document. Based on Kerckhoff's principle that assumes the watermark generation algorithm is accessible and known to the attacker, the content-based zero watermarking algorithm proposed by Jalil *et al.* (2010a) has a weakness where attacker is able to avoid the forgery detection by understanding the watermark pattern generation algorithm. Thus, the main question to be answered in this study is :

"How to increase the zero watermarking forgery detection accuracy for printed text document?"

The supporting questions are :

(a) How to improve unpredictability of the text document watermark pattern?
(b) How to make use of all ASCII printable characters in the printed text document for watermark generation to eliminate the possibility of detection bypass?
(c) How to proof that the proposed zero watermarking algorithm is able to improve the forgery detection rate?

1.4 Objectives

There are several objectives that are set to measure the success of this study. The objectives are listed as below :

(a) To study and analyze text watermarking methods that are available for text document forgery detection.

(b) To improve the text-based zero watermarking forgery detection accuracy of the content-based zero watermarking algorithm using text reordering and mixing of matrices.

(c) To analyze and compare the forgery detection result and accuracy of the proposed zero watermarking algorithm with the content-based zero watermarking algorithm.

1.5 Project Scopes

This study is carried out with the limitation based on several scopes as mentioned below :

(a) This study is not focused on Verification Authority (VA), a trusted third party verifier for document verification and forgery detection. Instead, a database is used to represent Verification Authority.

(b) The conversion of text in the document image is done after the selection of text area have been verified in OCR. The converted text is manually checked and corrected by user based on the printed text document.

(c) This study only covered ASCII printable characters excluding whitespaces for forgery detection.

(d) Matrix with the size of 4×4 is used in the proposed zero watermark algorithm.

(e) Forgery activity in this study only refers to text unauthorized modification in the text document.

1.6 Significant of the Project

In the end of this study, a document forgery detection method using zero watermarking algorithm is proposed for detecting forgery in the text content of printed document that have been always exposed to document falsification and forgery attacks. The proposed algorithm can be used in any printed text document such as certificates, transcript, legal documents and other confidential documents. The proposed algorithm is one of text watermarking alternatives that does not require special machines and materials for generating watermark and document verification process where author or the document sender can use existing devices to perform it. The algorithm is easy to understand and can be deploy without any hassle. The proposed algorithm is also expected to detect any unauthorized modification on the text document even if one ASCII printable character changed in the text document.

1.7 Chapter Organization

This study is divided into five chapters. Chapter 1 briefly describes about the summary of the whole study and background of the problem that motivates the need to study for current methods and alternatives to text watermarking. There are also objectives and scopes that need to be achieved. In Chapter 2, the discussion on the goals of information security and also literature reviews on various existing printed watermarking methods which are divided into image and text watermarking are done. From the literature review, research gap of this study is stated to improve weaknesses found in the current method or algorithm. The methodology that describes on how this study is executed are explained in Chapter 3. The methodology is based on the objectives stated in Chapter 1 that consists of methods and phases which are organized in a systematic way to ensure the objectives are successfully achieved.

The design and implementation of the proposed forgery detection method are defined in Chapter 4. In this chapter, the proposed forgery detection method using zero watermarking algorithm is explained in detail. The components involved in the

algorithm are discussed based on the research framework developed in Chapter 3. The design of the system gives an overview idea on how the system should work, what are the processes that involved in this method and how to implement the prototype of the forgery detection method. Chapter 5 discusses on several tests and results that have been captured on the performance of the proposed zero watermarking algorithm. These results are compared with the results of the performance of content-based zero watermarking algorithm to view the improvement that have been achieved on the text document forgery detection accuracy. The execution time performance of the proposed zero watermarking algorithm is also discussed in this chapter. Finally, Chapter 6 reviews and concludes the findings of forgery detection method using proposed zero watermarking algorithm with discussion of future works that can be done to further improve the proposed algorithm.

REFERENCES

Agani, N., Wahyudi, M. I. and Riyanto, 2013. Document Authentication Using Print-Scan Image Watermarking Based on DCT (Discrete Cosine Transform) Algorithm. Information Systems International Conference (ISICO). 2-4 December. Bali, Indonesia, 465-470.

Al-Wesabi, F. N., Alshakaf, A. Z. and Vasantrao, K. U., 2012. A Zero Text Watermarking Algorithm Based On The Probabilistic Weights For Content Authentication Of Text Documents. IJCA Proceedings on National Conference on Recent Trends in Computing (NCRTC). 7, 26-31. International Journal of Computer Applications.

Alattar, A. M. and Alattar O. M., 2004. Watermarking Electronic Text Documents Containing Justified Paragraphs and Irregular Line Spacing. SPIE Proceedings of Security, Steganography, and Watermarking of Multimedia Contents VI. 5306, 685-695. SPIE.

Atallah, M. J., McDonough, C. J. and Raskin, V., 2000. Natural Language Processing for Information Assurance and Security : An Overview and Implementations. 9th ACM/SIGSAC New Security Paradigms Workshop. 18-21 September, 2000. Cork, Ireland : ACM Press, 51-56.

Atallah, M. J., Raskin, V. Hempelmann, C. F. and Karahan, M., 2002. Natural Language Watermarking and Tamperproofing. In Petitcolas, F. A. P. (Ed.). Information Hiding (pp. 196-212). Berlin-Heidelberg : Springer.

Berita Harian, 2010, May 25. 45,000 Kakitangan Awam Tipu Sijil Cuti Sakit. Retrieved on November 20, 2013 from http://www2.bharian.com.my/articles/45_000kakitanganawamtipusijilcutisakit/Article/print_html

Bertrand, R., Gomez-Krämer, P., Terrades, O. R., Franco, P. and Ogier, J., 2013. A System Based On Intrinsic Features for Fraudulent Document Detection.12th International Conference on Document Analysis and Recognition (ICDAR). 25-28 August. Washington DC, United States : IEEE, 106-110.

Beusekom, J. v. and Shafait, F., 2012. Text-Line Examination for Document Forgery Detection. International Journal on Document Analysis and Recognition (IJDAR). 16(2), 189-207. Springer-Verlag.

Bhambri, P. and Kaur, P., 2014. A Novel Approach of Zero Watermarking for Text Documents. International Journal of Ethics in Engineering & Management Education (IJEEE). 1(1), 34-38.

Bharati, P. D. and Nitin, P. N., 2012. Text Watermarking Algorithm Using Structural Approach. World Congress on Information and Communication Technologies (WICT). 30 October-2 November. Trivandrum, India : IEEE, 629-633.

Biermann, C. , 1996. Paper and Its Properties. Handbook of Pulping and Papermaking (pp. 158-189). London : Elsevier.

Brassil, J. T., Low, S. and Maxemchuk, N. F. (1995). Electronic Marking and Identification Techniques to Discourage Document Copying. IEEE Journal on Selected Areas in Communications. 13(8), 1495-1504. IEEE.

Chaoju, H. and Xuning, W., 2009. Zero Watermark Protocol Based on Time-stamp and Digital Signature. International Forum on Information Technology and Applications (IFITA). 15-17 May. Chengdu, China : IEEE, 193-196.

Chen, H., Chuang, H., Kung, T. and Huang, Y., 2010. An Enhanced Three-party Encrypted Key Exchange Protocol Using Digital Time-Stamp. Sixth International Conference on Networked Computing and Advanced Information Management (NCM). 16-18 August. Seoul, South Korea : IEEE, 665-670.

Ćosić, J. and Bača, M., 2010. Improving Chain of Custody and Digital Evidence Integrity with Time Stamp. Proceedings of the 33rd International Convention MIPRO. 24-28 May. Opatija, Croatia : IEEE, 1226-1230.

Cox, I. J., Kilian, J., Leighton, T. and Shamoon, T., 1996. Secure Spread Spectrum Watermarking For Images, Audio and Video. International Conference on Image Processing. 16-19 September. Lausanne, Switzerland : IEEE, 243-246.

Daemen, J. and Rijmen, V., 2002. Specification of Rijndael. In Daemen, J. and Rijmen, V. (Ed.). The Design of Rijndael : AES - The Advance Encryption Standard. (pp. 31-51). Berlin-Heidelberg : Springer-Verlag.

Daraee, F. and Mozaffari, S., 2014. Watermarking In Binary Document Images Using Fractal Codes. Journal of Pattern Recognition Letters. 35, 120-129. Elsevier.

Davarzani, R. and Yaghmaie, K., 2009. Farsi Text Watermarking Based on Character Coding. International Conference on Signal Processing Systems. 15-17 May. Singapore : IEEE, 152-156.

Fei, W. and Tang, X., 2011. A Chinese Text Watermark Algorithm Based on Polyphone. Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC). 26-30 July. Harbin, China : IEEE, 1215-1218.

Forouzan, B. A., 2008. Cryptography and Network Security. (International Edition). Singapore : McGraw-Hill Education (Asia).

Haber, S. and Massias, H. (2011). Time-Stamping. In Tilborg, H. C. A. v. and Jajodia, S. (Eds.) Encyclopedia of Cryptography and Security (pp. 1299-1303). United States of America : Springer US.

Huang, D. and Yan, H., 2001. Inter-word Distance Changes Represented by Sine Waves for Watermarking Text Images. IEEE Transactions on Circuits And Systems For Video Technology. 11(12), 1237-1245. IEEE.

Jaafar, M. Y., 2012, January 18. Bekas Pengarah Veterinar Didakwa. Utusan Malaysia. Retrieved on November 20, 2013 from http://www.utusan.com.my/utusan/info.asp?y=2012&dt=0118&pub=Utusan_Malaysia&sec=Mahkamah&pg=ma_01.htm

Jalil, Z. and Mirza, A. M., 2010. An Invisible Text Watermarking Algorithm using Image Watermark. In Sobh, T. and Elleithy, K. (Eds.). Innovations in Computing Sciences and Software Engineering (pp. 147 - 152). Netherlands : Springer.

Jalil, Z., Mirza, A. M. and Sabir, M., 2010a. Content based Zero-Watermarking Algorithm for Authentication of Text Documents. International Journal of Computer Science and Information Security (IJCSIS). 7(2). 212-217.

Jalil, Z., Mirza, A. M. and Jabeen, H., 2010b. Word Length Based Zero-Watermarking Algorithm for Tamper Detection in Text Documents. Second International Conference on Computer Engineering and Technology (ICCET). 16-18 April. Chengdu, China : IEEE, 378-382.

Jalil, Z., Mirza, A. M. and Iqbal, T., 2010c. A Zero-Watermarking Algorithm for Text Documents Based on Structural Components. International Conference on Information and Emerging Technologies (ICIET). 14-16 June. Karachi, Pakistan : IEEE, 1-5.

Jaseena, K. U. and John, A., 2011. Text Watermarking using Combined Image and Text for Authentication and Protection. International Journal of Computer Applications. 20(4), 8-13. Foundation of Computer Science.

Kaur, S. and Babbar, G., 2013. A Zero-Watermarking Algorithm on Multiple Occurrences of Letters for Text Tampering Detection. International Journal on Computer Science and Engineering (IJCSE). 5(5), 294-301. Engg Journals Publications.

Khullar, S. and Singh, B., 2013. A Novel Content Based Zero Watermarking Algorithm for Tamper-Proofing Plaintext Document. International Journal of Computer Science and Engineering (IJCSE). 2(5), 1-10.

Kim, C. B., 2013, November 5. Johor Immigration Busts Syndicate Producing Fake Social Visit Passes. New Straits Times. Retrieved on November 22, 2013 from http://www.nst.com.my/latest/johor-immigration-busts-syndicate-producing-fake-social-visit-passes-1.392796

Kim, M., 2009. Natural Language Watermarking by Morpheme Segmentation. First Asian Conference on Intelligent Information and Database Systems (ACIIDS). 1-3 April. Dong Hoi, Vietnam : IEEE, 144-149.

Kim, M., Zaiane, O. R. and Goebel, R., 2010. Natural Language Watermarking Based on Syntactic Displacement and Morphological Division. 34th Annual IEEE Computer Software and Applications Conference Workshops. 19-23 July. Seoul, South Korea : IEEE, 164-169.

Kim, Y., Moon, K. and Oh, I., 2003. A Text Watermarking Algorithm based on Word Classification and Inter-word Space Statistics. Proceedings of the Seventh International Conference on Document Analysis and Recognition. 2, 775-779. IEEE.

Kulkarni, G., Patel, B. and Laxkar, P., 2013. Time stamp based Cross layer MANET security protocol. Third International Conference on Computational Intelligence and Information Technology (CIIT). 18-19 October. Mumbai, India : IEEE, 191-199.

Laine, M. and Nevalainen, O. S., 2006. A Standalone OCR System for Mobile Cameraphones. The 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. 11-14 September. Helsinki, Finland : IEEE, 1-5.

Liu, Y., Sun, X. and Wu, Y., 2005. A Natural Language Watermarking Based on Chinese Syntax. In Wang, L., Chen, K. and Ong, Y. S. (Eds.). Advances in Natural Computation (pp. 958-961). Berlin-Heidelberg : Springer.

Lu, H., Yi, F. D., XiaoLin, G., XiaoJiang, C., XinBai, X. and An, L. J., 2009. A New Chinese Text Digital Watermarking for Copyright Protecting Word Document. WRI International Conference on Communications and Mobile Computing (CMC). 6-8 January. Yunnan, China : IEEE, 435-439.

Lu, P., Lu, Z., Zhou, Z. and Gu, J., 2008. An Optimized Natural Language Watermarking Algorithm Based on TMR. The 9th International Conference for Young Computer Scientists (ICYCS). 18-21 November. Hunan, China : IEEE, 1459-1463.

Mali, M. L., Patil, N. N. and Patil, J. B., 2013. Implementation of Text Watermarking Technique Using Natural Language Watermarks. International Conference on Communication Systems and Network Technologies (CSNT). 6-8 April. Gwalior, India : IEEE, 482- 486.

Mayer, J. and Simske, S. J., 2012. Modulation in the HVS Domain for Hardcopy Watermarking of Color Documents. Eighth International Conference on Signal Image Technology and Internet Based Systems. 25-29 November. Naples, Italy : IEEE, 188-194.

Meral, H. M., Sankur, B. and Özsoy, A. S., 2008. Natural Language Watermarking Via Morphosyntactic Alterations. Journal of Computer Speech & Language. 23(1), 107-125. Elsevier.

Netto, J. and Carter, C. J., 2013, October 18. Official: Forged Documents Used In Prison Break from Fla. Prison. Channel News Network (CNN). Retrieved on November 20, 2013 from http://edition.cnn.com/2013/10/16/us/florida-inmates-mistakenly-freed/v

Oliveira, A. L., 2001. Techniques for the Creation Of Digital Watermarks In Sequential Circuit Designs. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 20(9), 1101-1117.

Patel, C., Patel, A. and Patel, D., 2012. Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study.International Journal of Computer Applications. 55(10), 50-56.

Renesse, R. L. V., 2002. Hidden And Scrambled Images : A Review. Proceedings of the 2002 SPIE Optical Security and Counterfeit Deterrence Techniques IV. 19 January. San Jose, CA : SPIE, 333-348.

Royster, P., 2011. The Art of Scanning. Library and Information Science Commons, 67, 1-28. Digital Commons / Institutional Repository Information@University of Nebraska-Lincoln.

Shirali-Shahreza, M. H. and Shirali-Shahreza, M., 2006. A New Approach to Persian/Arabic Text Steganography. Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science and 1st IEEE/ACIS International Workshop on Component-Based Software Engineering, Software Architecture and Reuse (ICIS-COMSAR). 10-12 July. Honolulu, Hawaii : IEEE, 310-315.

Simborg, D. W., 2008. Healthcare Fraud: Whose Problem is it Anyway?. Journal of The American Medical Informatics Association. 15, 278-280.

Topkara, M., Taskiran, C. M. and Delp, E. J., 2005. Natural Language Watermarking. SPIE Proceedings of Security, Steganography, and Watermarking of Multimedia Contents VI. 5681, 441-452. SPIE.

U.S. Department of Justice, 2013, July 8. Supervisor of $63 Million Health Care Fraud Scheme Sentenced in Florida to 10 Years in Prison. Official Press Release. Retrieved on November 20, 2013 from http://www.justice.gov/opa/ pr/2013/ July/13-crm-763.html

Vybornova, O. and Macq, B., 2007. Natural Language Watermarking and Robust Hashing Based on Presuppositional Analysis. IEEE International Conference on Information Reuse and Integration (IRI). 13-15 August. Las Vegas, United States : IEEE, 177-182.

Wang, H., Sun, X. Liu, Y. and Liu, Y., 2008. Natural Language Watermarking Using Chinese Syntactic Transformation. Journal of Information Technology. 7(6), 904-910.

Whitman, M. E. and Mattord, H. J., 2010. Management of Information Security (Third Edition). United States of America : Course Technology.

Yang, H. and Kot, A. C., 2004. Text Document Authentication By Integrating Inter Character And Word Spaces Watermarking. IEEE International Conference on Multimedia and Expo (ICME). 27-30 June. Taipei, China : IEEE, 955-958.

Yang, J., Wang, J., Wang, C. and Li, D., 2007. A Novel Scheme for Watermarking Natural Language Text. Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP). 26-28 November. Kaohsiung, China : IEEE, 481-484.

Yawai,W. and Hiransakolwong,N., 2013. Grid-Line Watermarking: A Novel Method for Creating A High-Performance Text-Image Watermark. Journal of Science Asia. 39(4). 423-435.

Yingjie, M., Xianlong, W., Wenjun, L. and Wei, C., 2013. Text Zero-Watermark Based on Chinese Edit Distance. Fifth International Conference on Computational and Information Sciences (ICCIS). 21-23 June. Shiyang, China : IEEE, 686-689.

Yu, Z. and Liu, X., 2009. A New Digital Watermarking Scheme Based on Text. International Conference on Multimedia Information Networking and Security (MINES). 18-20 November. Hubei, China : IEEE, 138-140.

Zhou, X., Wang, Z., Zhao, W., Wang, S. and Yu, J., 2009a. Performance Analysis and Evaluation of Text Watermarking. International Symposium on Computer Network and Multimedia Technology (CNMT). 18-20 January. Wuhan, China : IEEE, 1-4.

Zhou, X., Zhao, W., Wang, Z. and Wei, G., 2009b. Zero-Watermarking Algorithm for Content Authentication of Chinese Text Documents. Journal of Computers. 20(1). 11-17.

Zhou, Y., 2010. Are Your Digital Documents Web Friendly?: Making Scanned Documents Web Accessible. Information Technology and Libraries, 29(3), 151-160. The Library and Information Technology Association.

Zhu, L. and Zhu, L., 2012. Electronic Signature Based on Digital Signature and Digital Watermarking. 5th International Congress on Image and Signal Processing (CISP). 16-18 October. Chongqing, Sichuan, China : IEEE, 1644-1647.