

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 terganggu. 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 memasukkankan tera-air tersebut ke dalam dokumen teks asal. Tera-air asal dibandingkan dengan tera-air yang diekstrak daripada dokumen teks yang diterima untuk pengesanan 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 pelaksanaan algoritma tersebut mendapati algoritma tersebut memerlukan masa pelaksanaan kurang daripada satu saat di mana prestasi masa pelaksanaan algoritma tersebut adalah lebih kurang sama dengan algoritma tera-air yang lain.