

University Id : 10532
Student Id : LX2003001
Security Level : Normal

Hunan University

MASTER'S THESIS

**New and Intelligent Embedding Algorithm and New
Techniques for Information Hiding in Web Pages**

By

Mohamed Lahcen Bensaad

College : Computer and Communication
Major : Computer Science and Technology
Research field : Information Security and Digital Watermarking
Supervisor : Prof. Sun XingMing
Submission Date : May, 10, 2006
Defense Date : May, 20, 2006
Defense Committee Chairman: Prof. Li Xiao Jian

**New and Intelligent Embedding Algorithm and New
Techniques for Information Hiding in Web Pages**

By
Mohamed Lahcen Bensaad

THIS THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER
IN
COMPUTER SCIENCE AND TECHNOLOGY
AT

Graduate School
Hunan University
ChangSha, P.R. China
May, 2006

© Copyright 2006, Mohamed Lahcen Bensaad, 2006

Table of Content

Table of Content	3
Table of Figures	6
Table of Tables	7
Abstract.....	8
Chapter 1 General Introduction	9
1.1 Preamble	9
1.2 Motivation.....	10
1.3 Purposes	10
1.4 Contribution	11
1.5 Manuscript Outline	11
Chapter 2 Cryptography and Steganography	12
2.1 Origins.....	12
2.2 Cryptography	12
2.3 Steganography.....	15
2.3.1 The Egyptians	15
2.3.2 The Chinese	16
2.3.3 The Greeks	16
2.3.4 The Arabs.....	17
2.3.5 World War II.....	18
2.3.6 Modern Techniques	20
Chapter 3 Information Hiding and Steganosystems	22
3.1 General Model of Steganographic System	22
3.2 Modern Applications of Steganography	23
3.2.1 Covert communication.....	23
3.2.2 Proof of ownership.....	23
3.2.3 Authentication and Information Integrity	23
3.2.4 Fingerprinting	24
3.2.5 Monitoring	24

3.3 Constraints, Goals and Difficulties in Algorithms.....	24
3.3.1 Fidelity	25
3.3.2 Attack Strength	25
3.3.3 Computational Complexity.....	25
3.3.4 Security	25
3.4 Formal Description of a Stegosystem.....	26
3.4.1 Pure Steganography	27
3.4.2 Secret Key Steganography.....	27
3.4.3 Public Key Steganography.....	28
3.5 Security in Stegosystems	28
3.6 Steganography Applied to Different Media.....	30
3.6.1 Still Images	30
3.6.2 Moving Images	30
3.6.3 Audio Files.....	31
3.6.4 Text Files	32
3.6.5 Steganographic File Systems	33
3.6.6 Hiding in Network Packets	35
3.7 Steganalysis and Attacks.....	43
3.8 Description of a New Intelligent Embedding Algorithm.....	44
Chapter 4 Digital Watermarking.....	47
4.1 Introduction.....	47
4.2 Visible and Invisible Watermarks.....	47
4.3 Basic Watermarking Principles.....	48
4.4 Classification of Watermarks.....	51
4.4.1 Fragile Watermark	51
4.4.2 Robust Watermark	52
4.5 Watermarking Using the Intelligent Embedding Algorithm	53
4.5.1 Types of Intelligent Watermarking Embeddings.....	53
4.5.2 Robustness	55
Chapter 5 Techniques for Information Hiding in Web Pages.....	56
5.1 Introduction.....	56

5.2 Web Pages and HTML.....	56
5.3 Where to Hide Information in HTML Code.....	57
5.3.1 White Space.....	57
5.3.2 Line Break.....	60
5.3.3 Order of Attributes.....	61
5.3.4 The Default Value of an Attribute:.....	62
5.3.5 Optional End Tags:.....	62
5.3.6 String delimiters:.....	62
5.3.7 Changing Color Values.....	62
5.4 Implementation and Experiments.....	63
5.4.1 Experiments.....	63
5.4.2 More Tests.....	64
5.4.3 Statistics.....	70
Conclusion and Future Work.....	74
References.....	75
Published papers.....	79

Abstract

Nowadays, the fascinating art of Information hiding is one of the hottest reach fields due to its significance and importance in solving many rising problems in the ear of information and internet. Digital watermarking, as a branch of information hiding, is a promising technique for the protection of intellectual property rights. This technology is mainly advanced for multimedia but the same has not been done for text. Because of the varieties of file formats, there has been no attempt to make one unified general embedding algorithm which can be applied to different types of data. The most general method of embedding a message M into a set of cover-objects C is to break down M into small units (bits) M_i then embed each unit M_i into one cover-object C_i , which is in most cases, if not always, selected sequentially from the cover data stream. This way of selecting the cover-objects C_i can be enhanced to make it more intelligent by studying the properties and features of the different kinds of C_i and use them accordingly, and that is what will be shown here. Web pages, as one of the many kinds of file formats, can be used as cover documents to server different purposes. In this work, some new techniques are presented to show how to hide information in web pages using some features of the markup language used to describe these pages. Most of these techniques use the white space to hide information or some varieties of the language in representing elements. Experiments on a very small page and analysis of five thousands web pages show that these techniques provide a wide bandwidth available for hiding a quite big amount of data. Tests show that the use of the presented techniques applying the new proposed embedding algorithm will enhance the robustness of the hidden information. This work might serve to form a solid base to develop robust watermarking algorithms not only for web pages but also for all other types of data.

Keywords: Steganography, Information Hiding, Covert Communication, Digital Watermarking, Robust, HTML and Web Pages.