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

Table of Content

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.