PROTOTYPE DEVELOPMENT FOR EMBEDDING LARGE AMOUNT OF INFORMATION USING SECURE LSB AND NEURAL BASED STEGANOGRAPHY

BASAM N. SALEH

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

Centre for Advanced Software Engineering (CASE)
Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia

APRIL 2009

ABSTRACT

The security of information became a very important issue. Steganography is an effective way to hide the desired secret information in seemingly innocent cover files which are mostly multimedia files. Using multimedia files as hosts to hide the information in will avoid the need to secure the communication when sending secret messages. The challenge to Steganography is the amount of information to be embedded in the host file without affecting the properties of that file and to avoid distortion of the image, the video, or the sound host file and as a result, to avoid detection of hidden information existence. The need for new methods, techniques and algorithms to make enhancements regarding increasing the amount the hidden information, preserving the host file quality, preserving the size of the file, and keep it robust against steganalysis. To achieve these goals, the embedding must be in suitable locations in the multimedia file, choosing the proper. A recent approach is using artificial intelligence that teaches the machine to give the best candidate bits to hide the information in. This approach is remarkably theoretically efficient, and this approach is the basis of this project to implement a prototype that uses this approach. In this project, for embedding, neural network with adaptive smoothing error back propagation that keeps trying to refine the Stego file until it reaches the best embedding results besides another adaptive Steganography method using concepts called main cases and sub cases. In this project, four layers of security will be used to secure the hidden information and to add more complexity for steganalysis and another point of focus in this project will be on embedding the maximum amount of information that can be embedded without affecting the other objectives.

ABSTRAK

Keselamatan maklumat merupakan isu terpenting terutamanya kepada pihak berkuasa dalam urusan pentadbiran harian. “Steganography” merupakan cara yang efektif untuk menyembunyikan maklumat sulit dalam file multimedia yang kelihatan biasa. Setelah “embedding” maklumat di “host file” selesai, sebarang kaedah yang digunakan untuk menghantar maklumat akan menjadi selamat dan kukuh kerana “host file” ini bukan lagi merupakan titik tumpuan utama. Cabaran untuk “Steganography” adalah jumlah maklumat yang dapat dimuatisikan ke “host file” tanpa mempengaruhi property file dan menghindari distorsi pada gambar, video, atau suara ”host file”, justeru dapat menyindari dari sebarang deteksi mengenai kehadiran maklumat tersembunyi. Pencarian terus untuk kaedah baru, teknik dan algorithms untuk membuat perangkat tambahan demi meningkatkan jumlah maklumat yang tersembunyi, melestarikan kualiti serta saiz “host file”, dan tetap kuat terhadap steganalysis adalah sangat penting. Untuk mencapai tujuan tersebut, kesesuaian lokasi untuk “embedding” di dalam file multimedia adalah penting dengan memilih bait yang paling sesuai di tempat bit bait. Ini juga merupakan cabaran yang sangat besar kepada steganographers. Pendekatan terkini adalah menggunakan kecerdasan buatan yang mengajar mesin untuk mengesan dan memberikan calon bit yang terbaik untuk menyembunyikan maklumat. Pendekatan ini bukan sahaja secara teoritis efisien, malahan merupakan dasar projek ini untuk menerapkan prototype yang menggunakan pendekatan ini. Dalam projek ini, untuk “embedding”, neural network dan adaptive smoothing error back propagation yang terus berusaha untuk memperbaiki stego file sehingga mencapai hasil yang terbaik untuk “embedding” selain daripada adaptif “Steganography” lain yang menggunakan kaedah yang dikenali sebagai main cases dan sub cases. Dalam projek ini, empat lapisan sekuriti akan digunakan untuk meneguhkan maklumat tersembunyi tersebut dan untuk menambahkan kerumitan untuk steganalysis lain, titik focus dalam projek

ini akan “embedding” jumlah maklumat yang maksimum tanpa mempengaruhi tujuan yang lain.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	viii
	LIST OF TABLES	xii
	LIST OF FIGURES	xiii
	LIST OF ABBREVIATIONS	xvi
1	**INTRODUCTION**	**1**
	1.1 Overview	1
	1.2 Background of the Problem	2
	1.3 Problem Statement	4
	1.4 Project Aim	5
	1.5 Project Objectives	5
	1.6 Project Scope	5
	1.7 Summary	6
2	**LITERATURE REVIEW**	**7**
	2.1 Introduction	7
	2.2 Image File Formats	8
	2.2.1 Image Parameters	9
	2.2.2 Most Common Used Image Format	12

2.2.3 Red Green Blue (RGB) Images 14
2.3 Steganography 16
2.3.1 MSE and PSNR Formulas 17
2.3.2 Different Forms of Steganography 18
2.3.3 Steganographic Methods 19
2.3.4 Least Significant Bit (LSB) Insertion Steganography 22
2.3.4.1 LSB in BMP 23
2.3.4.2 LSB in PNG Image 26
2.3.4.3 Steganography in GIF image 27
2.4 Steganalysis 27
2.4.1 Steganalysis Methods 29
2.4.2 Steganalysis Against LSB 30
2.5 Choosing the Best Location in the Cover Image to Hide Information 32
2.5.1 Coding Framework 36
2.6 Intelligent Data Embedding Method for LSB Steganography 37
2.6.1 Neural Networks Learning System 39
2.7 Summary 40

3 RESEARCH METHODOLOGY 41

3.1 Introduction 41
3.2 Requirements Specifications 42
3.3 Prototype Architectural Design 42
3.4 Prototype Development 44
3.5 Testing the Results 44

4 PROTOTYPE DESIGN 45

4.1 Introduction 45
4.2 Design Challenges 46
4.3 The Prototype Architecture 48

4.3.1 First Security Layer (AES Encryption) 49
4.3.2 Second Security Layer (Adaptive Segmentation) 50
4.3.3 Third Security Layer (Main Cases and Sub Cases) 51
4.3.4 Fourth Security Layer (Neural network) 57
4.3.5 Extraction and Decryption Layer 60
4.4 Operational Phases 63
4.5 Summary 63

5 PROTOTYPE IMPLEMENTATION 64

5.1 Introduction 64
5.2 Implementation Phases 64
5.3 Prototype Code Structure and UML Diagrams 68
5.4 Summary 83

6 TESTING THE RESULTS AND CONCLUSION 84

6.1 Introduction 84
6.2 The Benchmark 84
6.3 Prototype Usage Limitations 87
6.4 Testing Approaches And Methods 87
6.4.1 Program Performance 88
6.4.2 Results Listing and Analyzing 89
6.4.2.1 Experiment to Hide Very Small Amount of Information (459bytes) 91
6.4.2.2 Experiment to Hide (9 K.B) of Information 93
6.4.2.3 Embedding Experiment Using This Prototype and Maximum Embedding Ability of S-Tools 95
6.4.2.4 Embedding Experiment Using Maximum Embedding Ability of Both This Prototype and S-Tools 98
6.5 Meeting The Objectives 101

6.6 Summary And Conclusion 102

REFERENCES **103**

CHAPTER 1

INTRODUCTION

1.1 Overview

Steganography is the art of passing information in a manner that the very existence of the message is unknown. The goal of Steganography is to avoid drawing suspicion to the transmission of a hidden message. If suspicion is raised, then this goal is defeated. [21]

Steganography is defined also as it is the art and science of communicating in a way which hides the existence of the communication. In contrast to Cryptography, where the enemy is allowed to detect, intercept and modify messages without being able to violate certain security premises guaranteed by a cryptosystem, the goal of Steganography is to hide messages inside other harmless messages in a way that does not allow any enemy to even detect that there is a second message present".[21]

Steganography applications conceal information in other, seemingly innocent media. Steganographic results may masquerade as other file for data types, be concealed within various media, or even hidden in network traffic or disk space. We are only limited by our imagination in the many ways information and data can be exploited to conceal additional information. [46]

Redundant or noisy data can be removed from the original image and replaced with a hidden message. Steganographic technologies are a very important part of the future of Internet security and privacy on open systems such as the Internet. Steganographic research is primarily driven by the lack of strength in the

cryptographic systems on their own, and the desire to have complete secrecy in an open-systems environment. [47]

There are a number of uses for Steganography besides the mere novelty. One of the most widely used applications is for so-called digital watermarking. A watermark, historically, is the replication of an image, logo, or text on paper stock so that the source of the document can be at least partially authenticated. A digital watermark can accomplish the same function; a graphic artist, for example, might post sample images on her Web site complete with an embedded signature so that she can later prove her ownership in case others attempt to portray her work as their own. [17]

Steganography can also be used to allow communication within an underground community. There are several reports, for example, of persecuted religious minorities using Steganography to embed messages for the group within images that are posted to known Web sites. [17]

Hiding the information in an image is known as the Embedding process, It can be done using various techniques of Steganography, taking in consideration lossless information or image quality, and also it is very important to keep the original file size so that the detection of hidden information will be harder, and the image will not be suspicious.

1.2 Background of the Problem

There are several techniques for Steganography, some of which become very complicated to understand. One simple method is LSB (Least Significant Bit), or Least Significant Bit Steganography. The concept of LSB Embedding is simple. It exploits the fact that the level of precision in many image formats is far greater than that perceivable by average human vision. Therefore, an altered image with slight variations in its colors will be indistinguishable from the original by a human being, just by looking at it.

By using the least significant bits of the pixels' color data to store the hidden message, the image itself will seem unaltered. An image is nothing more than strings and strings of bytes, each byte representing a different color. The last few bits in a color byte, however, do not hold as much significance as the first few. This is to say that two bytes that only differ in the last few bits can represent two colors that are virtually indistinguishable to the human eye. For example, 00100110 and 00100111 can be two different shades of red, but since it is only the last bit that differs between the two, it is impossible to see the color difference. LSB Steganography, then, alters these last bits by hiding a message within them. [38]

As important as the Steganographic technique is, equally important is the choice of the cover image. In LSB Embedding, a poor choice of cover image can lead to a Stego-image that is easily differentiable from the original. Current image formats can be divided into two broad categories, lossy and lossless. Lossy images are those formats, which loses some of the image's data when stored. An example would be JPEG. The plus side of lossy images, in particular JPEG, is that it achieves extremely high compression, while maintaining fairly good quality. However, due to the very nature of lossy formats, it is not suitable for LSB Embedding. [34]

Since LSB Embedding spreads the hidden message throughout the image's data, the loss of the image's data by compression would lead to the lost of parts of the hidden message. On the other hand, lossless images are suitable for LSB Embedding, since the integrity of the image data is preserved. However, they do not have the high compression ratio that lossy formats do. Not all lossless images are good candidates as a cover image. 24-bit bitmaps, as well as grayscale images and other color images with small variations in its palette are good candidates as cover images. [34]

The main advantage of the LSB coding method is a very high watermark channel bit rate and a low computational complexity of the algorithm, while the main disadvantage is considerably low robustness against signal processing modifications. Increasing Robustness of LSB Audio Steganography by Reduced Distortion LSB Coding. Further More, LSB Embedding has the advantage that it is simple to

implement. This is especially true in the 24-bit bitmap case. It also allows for a relatively high payload, carrying one bit of the secret message per byte of pixel data. In addition, it is also seemingly undetectable by the average human if done right. However, the assumption has been that the Stego-image is indistinguishable from the original cover image by the human eye. There have been many statistical techniques developed to determine if an image has been subjected to LSB Embedding. [34]

It's noticed that almost all of the current LSB algorithms for RGB (Red Green Blue) color scheme are not considering an intelligent method or the use of artificial intelligence to perform the embedding process choosing the best candidates (pixels) to embed the data, even the existing automated implementations are not used a trained machine to embed the data, but the need for such kind of systems is growing due to the new techniques to detect the LSB hidden information, so the LSB Steganography needs a fast reliable method to embed the information in the host image.

1.3 Problem Statement

How can we reduce the lack of existence of an intelligent method to choose the best candidate pixels in the cover image file to embed the information in?

What method we can suggest for LSB Steganography to solve the problem of choosing the best location in the image to hide large amount of information in a fast accurate reliable way?

How can the suggested method be able to avoid detection of the existence of hidden information and avoid the failure of secrecy of the desired communication?

1.4 Project Aim

The aim of this proposed project is to develop a prototype of an intelligent method to choose the best candidate pixels locations in any RGB bitmap image file to hide large amount of information in those pixels.

1.5 Project Objectives

The objectives of this project are:

i. Developing and implementing a prototype that trains the machine to give the candidate pixels in an RGB image file for the best location to hide information using LSB Steganography.
ii. Comparing the resulted Stego images from this prototype to Stego images resulted by S-Tools (the benchmark).
iii. Preserving the size of the cover image in the Stego image produced by this prototype from that particular cover image.
iv. Preserving the quality of the cover image in the Stego image produced by this prototype from that particular cover image.
v. Producing Stego image from this prototype that is robust against specific visual and statistical measures by increasing the complexity of the statistical and visual steganalysis.

1.6 Project Scope

The scope of this project will be working on the true 24-bits color RGB bitmap images, and the embedding algorithm will be the LSB algorithm. In testing phase of this project, only specific basic types of visual and statistical measures will be considered regarding the steganalysis complexity that might be performed on the Stego images produced by this prototype. Those measures are comparing the

brightness difference, the neighbor pixels difference, and the Euclidian norm. The desired file to be hidden in the cover image will be of the format of a Microsoft Windows notepad text file (.txt).

The project will consider the most important success factors for Steganography like preserving the file size and keep the size change very small that can be hardly noticed, lossless information, maximum extraction of the hidden information, and preservation of the image quality.

1.7 Summary

In this chapter we discussed the aim and objectives of this project and what is the background of the problem that was the reason to choose this topic of the project. The scope was identified for our work and the problem statement was declared.

REFERENCES

1. LK Tan, Image File formats, MBiomedEng, Department of Biomedical Imaging, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia,2006.
2. CompuServe Incorporated, A standard defining a mechanism for the storage and transmission of raster-based graphics information, GIF team, 1987
3. Patrik Lynch and Sarah Horton, Web Style Guide, 2004
4. Adrian Brown, Graphics file formats , Digital Archives Analystes Analyst, the national archive, 2003
5. Jpeg Image format, soap lab website http://www.soapplab.auckland.ac.nz/info/formats/jpeg.htm
6. Joe Burns, Image formats, 1997.
7. Ross Shannon, Image file formats, HTML source web site www.yourhtmlsource.com.
8. Richard H. Wiggins III, Christian Davidson, Ric Harnsberger, Jason R. Lauman, Patricia A. Goede, Image File Formats: Past, Present,and Future, BS, Radio Graphics, 2001.
9. TIFF, Revision 6.0,Final, Adobe Developers Association, Adobe Systems Incorporated,2003.
10. The MathWorks, Inc. www.mathworks.com, 2008
11. John Kielkopf, university of Louisville, 2004
12. Pixel Based Imaging – Unit 13 Channels and Masks technical report, faculty of the art staff and course pages, http://mercury.tvu.ac.uk
13. Octavian Henegariu. Nature Genet, Information supplementary to correspondence, (1999).

14. SVD Kotera, RGB to spectral image conversion using spectral pallete and compression, H. international conference on Image Processing, 2003

15. Jaroslav Fojtik and Vaclav Hlavac, Faculty of Electrical Engineering Center for Machine Perception, Czech Technical University.

16. Rajarathnam Chandramouli, Image Steganography and Steganalysis: Concepts and Practice, Department of Electrical and Computer Engineering Stevens Institute of Technology, Mehdi Kharrazi Department of Electrical and Computer Engineering, Polytechnic University, Brooklyn, and Nasir Memon Department of Computer and Information Science, Polytechnic University, Brooklyn,2003

17. Gary C. Kessler, Steganography: Hiding Data Within Data, September 2001.

18. Deshpande Neeta, Kamalapur Snehal, Implementation of LSB Steganography and Its Evaluation for Various Bits, Computer Science Dept ,K.K.Wagh Institute of Engineering, Education and Research, Nashik,India.

19. Hide and Seek: An Introduction to Steganography, NIELS PROVOS AND PETER HONEYMAN, University of Michigan, 2004

20. Sellars, D., An Introduction to Steganography, URL: http://www.cs.uct.ac.za/courses/CS400W/NIS/papers99/dsellars/stego.html

21. Johnson, Neil F., Steganography, 2000

22. KUNDUR, D. – HATZINAKOS, D., Mismatching Perceptual Models for EffectiveWatermarking in the Presence of Compression , Department of Electrical and Computer Engineering University of Toronto 10 King's College Road Toronto, Ontario Canada M5S 3G4.

23. von Ahn and Nicholas J. Hopper, Public-Key Steganography, Luis Computer Science Dept, Carnegie Mellon University, Pittsburgh PA 15213 USA

24. Munirajan, V.K.; Cole, E.; Ring, S., Transform domain Steganography detection using fuzzy inference systems, Multimedia Software Engineering, 2004.

25. BENDER, W. – GRUHL, D. – MORIMOTO, N., Techniques for Data Hiding, Massachusetts, Institute of Technology, Media Laboratory

Cambridge, Massachusetts 02139 USA, From the Proceedings of the SPIE, San Jose CA, February, 1995.

26. Warfare Center, 2555 Amphibious Drive NAB Little Creek Norfolk, VA 23521 3225, Cynthia E. Irvine Computer Science Department Code CS/Ic Naval Postgraduate School Monterey, CA 93943 5118 ,Proceedings of the 19th, National Information System Security Conference, Baltimore, Md, October 1996.

27. CURRIE, D. L., Surmounting the Effects of Lossy Compression on Steganography, III Fleet Information

28. ZHAO, J. – KOCH, E., Fraunhofer,Embedding Robust Labels Into Images For Copyright Protection, Institute for Computer Graphics Wilhelminenstr. 7, 64283 Darmstadt, Germany, Proc. of the Int. Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies, Vienna, August 1995.

29. Fridrich Jiri, A new steganographic method for palette-based images, Center for Intelligent Systems, SUNY Binghamton, Binghamton, New York. ISandT_s PICS Conference, 1999

30. Daisy Jacobs, Stego two bits, School of Information Technology University of Pretoria, Pretoria 002 ,South Africa.

31. Fridrich Jiri, Rui Du, Secure steganographic methods for palette images, Center for Intelligent Systems, Dept. of SSandIE, SUNY Binghamton, Binghamton, New York. Information Hiding, Proceedings of the Third International Workshop, IH_99 Dresden Germany, Computer Science, 2000

32. Eiji Kawaguchi, Steganography and Steganalysis, (KIT STEGROUP),2008

33. Dr.Talal Alkharobi, Steganography, King Fahd University of Petroleum and Minerals Dhahran, Saudi Arabia ,2007

34. N. F. Johnson and S. Jojodia, Exploring Steganography: Seeing the Unseen ,1998

35. Wikipedia, the free encyclopedia www.wikipedia.com

36. Brittnee Morgan, Steganography Detection, university of rhode island

37. Michael T. Raggo, Steganography, Steganalysis and Cryptanalysis, CISSP Principal Security Consultant, VeriSign Co.

38. Ilana Marcus, University of Rhode island , Steganography Detection URL:http://www.uri.edu/personal2/imarcus/stegdetect.htm

39. Sonali Gupta, Steganalysis, SANS GCIH, 2005

40. Andrew D. Ker, Quantitative Evaluation of Pairs and RS Steganalysis, Oxford University Computing Laboratory, Parks Road, Oxford OX1 3QD, England

41. J. Fridrich, M. Goljan, and D. Soukal, V, E. J. Delp III and P. W. Wong, "Higher-order statistical steganalysis of palette images," in Security and Watermarking of Multimedia Contents, eds., Proc. SPIE 5020, 2003.

42. J. Fridrich, M. Goljan, and R. Du, Reliable detection of LSB Steganography in color and grayscale images, Proc. ACM Workshop on Multimedia and Security, 2001.

43. K. Solanki, K. Sullivan, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, provably secure Steganography: achieving zero K-L divergence using statistical restoration, Dept. of Electrical and Computer Engineering ,University of California at Santa Barbara, Santa Barbara, CA 93106.

44. Kaushal Solanki, Anindya Sarkar, Y.A.S.S. (Yet Another Steganographic Scheme That Resists Blind Steganalysis), and B.S. Manjunath, Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, Mayachitra Inc., 5266 Hollister Avenue, Santa Barbara, CA 93111

45. Nameer N. EL-Emam, Embedding a Large Amount of Information Using High Secure Neural Based Steganography Algorithm, International Journal of Signal Processing, 2006

46. Dr. Neil F. Johnson, Information Hiding: Steganography and Digital Watermarking, 2006 URL:http://www.jjtc.com/Steganography/

47. Bret Dunbar, SANS Institute, A detailed look at Steganographic Techniques and their use in an Open-Systems Environment, 01/18/2002.

48. Walter Maner, bowling green state university, Ohio, 1997.

49. Basil Tesler, www.webspacestation.com, 2005.

50. Webopedia the online computer encyclopedia, www.webopedia.com.

51. James Rumbaugh, The Unified Modeling Language Reference Manual, (2nd Edition) (Addison-Wesley Object Technology Series).

52. Steganography Obliterator: An Attack on the Least Significant Bits, Guillermo A. Francia, III, Ph.D, Tyler S. Gomez and Jacksonville State University Jacksonville, Alabama.

53. Gray C. Kessler, Champlain College in Burlington, Vermont, September 2001.

54. David Salomon , Data privacy and security, 2003

55. The U.S Military Digital Challenge, http://www.dc3.mil/2006_challenge, 2006