STEGANALYSIS: DETECTING EXISTENCE OF FILE TYPES EMBEDDED IN GREY SCALE IMAGES

This thesis is presented to the
College of Arts and Sciences
in fulfillment of the requirements for
Master of Science (Information Technology)
Universiti Utara Malaysia

By
AMIRULIKHSAN ZOLKAFLI

STEGANALYSIS: DETECTING EXISTENCE OF FILE TYPES EMBEDDED IN GREY SCALE IMAGES

This thesis is presented to the
College of Arts and Sciences
in fulfillment of the requirements for
Master of Science (Information Technology)
Universiti Utara Malaysia

By
AMIRULIKHSAN ZOLKAFLI

KOLEJ SASTERA DAN SAINS
(COLLEGE OF ARTS AND SCIENCES)
UNIVERSITI UTARA MALAYSIA

PERAKUAN KERJA/TESIS
(Certification of Thesis Work)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

AMIRULIKHSAN ZOLKAFLI

calon untuk Ijazah
(candidate for the degree of) **SARJANA SAINS (TEKNOLOGI MAKLUMAT)**

telah mengemukakan tesis/disertasinya yang bertajuk
(has presented his/her thesis work of the following title)

STEGANALYSIS: DETECTING EXISTENCE OF FILE TYPES EMBEDDED IN GREY SCALE IMAGES

seperti yang tercatat di muka surat tajuk dan kulit tesis/disertasi
(as it appears on the title page and front cover of thesis work)

bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan, dan liputan bidang ilmu yang memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : **20 Mei 2008**

(that the thesis/dissertation is acceptable in form and content, and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate through an oral examination held on

Pengerusi Viva *(Chairman for Viva)*	:	Dr. Nor Laily binti Hashim	Tandatangan: *(Signature)*	
Pemeriksa Luar *(External Examiner)*	:	Dr. Norafida binti Ithnin	Tandatangan: *(Signature)*	
Pemeriksa Dalaman *(Internal Examiner)*	:	En. Ahmad Hisham bin Zainal Abidin	Tandatangan: *(Signature)*	
Penyelia Utama *(Principal Supervisor)*	:	Dr. Shafiz Affendi bin Mohd Yusof	Tandatangan: *(Signature)*	
Setiausaha Panel *(Panel Secretariat)*	:	En. Syamsul Bahrin bin Zaibon	Tandatangan: *(Signature)*	

Tarikh
(Date) : **20 MEI 2008**

PERMISSION TO USE

In presenting this thesis in full fulfillment of the requirement for the postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may take it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor or, in absence, by the Applied Science Chair. It is understood that any copying or publication or use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests foe permission to copy or make other use of materials in this thesis, in whole or in part, should be addressed to:

Applied Science Chair
College of Arts and Sciences
Information Technology Building
Universiti Utara Malaysia
06010 Sintok
Kedah Darul Aman

ABSTRAK

Steganografi ialah satu seni menyembunyikan maklumat rahsia ke dalam medium digital di mana hanya penerima yang sah tahu akan kewujudan maklumat yang tersembunyi. Sejak peristiwa 11 September 2001, steganografi telah dikaitkan dengan kegiatan keganasan. Ada juga yang mengaitkan kumpulan Al-Qaeda dengan penggunaan steganografi sebagai salah satu cara untuk berkomunikasi secara rahsia dalam menyebarkan fahaman mereka dan pelan-pelan keganasan dengan menyembunyikannya ke dalam imej digital di dalam internet. Untuk menyekat sebarang kemungkinan ancaman kini dan masa hadapan, penyiasat forensik komputer perlu cuba sedaya upaya untuk mengenalpasti dan menyekat sebarang komunikasi yang berbentuk rahsia. Dalam pada itu, bukan sahaja keupayaan untuk mengenalpasti kewujudan maklumat rahsia yang diperlukan malahan usaha-usaha untuk mengenalpasti sifat atau jenis maklumat yang disembunyikan dan mengeluarkannya adalah penting supaya sebarang pelan-pelan jahat di masa hadapan dapat dijangka dan ditewaskan pada peringkat yang lebih awal. Kami membentangkan kaedah yang dikenali sebagai "active steganalysis" sebagai satu kaedah untuk membongkar sebarang kemungkinan mengenalpasti kewujudan maklumat tersembunyi. Membuat ramalan mengenai parameter algoritma penyembunyian adalah pendekatan yang digunakan dalam kajian ini memandangkan maklumat rahsia adalah salah satu daripada parameternya. Pendekatan eksperimen digunakan untuk mengesan pelbagai format fail maklumat tersembunyi di dalam imej digital hitam putih. Format fail yang berbeza akan bertindak sebagai pelbagai kemungkinan jenis maklumat tersembunyi yang digunakan oleh pihak pengganas. Kami juga ingin membuktikan bahawa imej digital yang telah dimasukkan dengan jenis format maklumat yang berbeza akan meninggalkan kesan atau bukti yang unik di mana ianya dapat digunakan untuk mengesan dan mengagak sifat atau jenis maklumat tersembunyi dengan bantuan pemprosesan imej (perbezaan imej) dan juga analisis statistik (Paired T dan ANOVA).

ABSTRACT

Steganography is an art and science of hiding secret messages into other digital mediums in such a way that no one apart of the intended recipient knows the existence of the message hidden. Recently, steganography have been linked with terrorism activities prior to the September 11th 2001 tragedy. There have been claims that a terrorist group known as Al-Qaeda has been using steganography as a way of secretive communication in spreading their ideologies and attack plans by hiding malicious materials in digital images via the internet. To intercept possible current and future threats, computer forensic examiners must try to identify and intercept any possible secret communication. In doing so, not only the capability to detect the existence of hidden messages is vital but efforts to determine the nature of the hidden message and extract it are also needed so that any malicious plans in the future could be predicted and intercepted earlier. Active steganalysis is presented as ways to expose the possibility of detecting existence of hidden messages, hence defeating steganography. Estimating some parameters of embedding algorithm is the approach that the study used since secret message is one of its essential parameters. Experimental research method was used to detect hidden messages with various file formats in grayscale digital images by conducting two experiments. The various file formats will act as the different possible types of hidden messages that could have been used by terrorists. Here we show that digital images embedded with different file format leaves unique statistical evidence that could be used for detection and estimating the nature of the hidden message with the aid of image processing (image subtraction) and statistical analysis (Paired T and ANOVA).

ACKNOWLEDGEMENT

In the name of Allah, The Most Gracious and The Most Merciful,

I would like to thank:

The Ministry of Science, Technology and Innovation (MOSTI) for the financial support,

Universiti Utara Malaysia for the facilities and resources provided,

My supervisor, Dr. Shafiz Affendi Mohd Yusof. I gratefully acknowledge his warm encouragement and patient guidance through my slow process of preparing this thesis,

Associate Professor Dr. Sharipah Soaad Syed Yahaya for her help in providing me with a very interesting guide in statistical analysis, which has helped me to complete the critical part of the study and chapter in this thesis,

My close friends have been a constant source of joy and gave me the strength I needed to go through the preparation of this thesis. Thank you.

Finally I would like to acknowledge my thanks to my family. My parents who supported me with their encouragement and comprehension, provided me constant support which helped me to overcome the many difficulties and discouragement on the way to completing this thesis.

◆

This thesis is dedicated to my family
especially my parents who lovingly encouraged
and supported me all the way since the
beginning of this research

◆

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ANOVA	**Analysis Of Variance**
BMP	**Bitmap**
CFE	**Computer Forensic Examiner**
DCT	**Discrete Cosine Transform**
FBI	**Federal Bureau of Investigation**
GLM	**Grey Level Modification**
HTML	**Hypertext Markup Language**
JPEG	**Joint Photographic Expert Group**
IT	**Image Tool**
LSB	**Least Significant Bit**
MRI	**Magnetic Resonance Imaging**
SNR	**Signal-to- Noise Ratio**
PSNR	**Peak Signal-to Noise Ratio**
POV	**Pair of Values**
UTHCSCA	**University of Texas Health Science Center at San Antonio**

CHAPTER ONE

INTRODUCTION

Steganography have been widely part of the secret communication world, alongside cryptography. The resurgent of this ancient art of hiding information into the Internet world has spur a lot of attention, good and bad. It has brought a new meaning into the secret communication society. But after the September 11th 2001 tragedy, the malicious side of it has been in the spotlight. The investigation of the terrorist attacks on the United States has drawn new attention to this innocuous method of sending messages via the Internet. The method has the ability to hide messages in digital images or in music files but leave no outward trace that the files were altered. To protect secrecy, terrorists continuously discover new and better cover mediums as well as design and develop robust algorithms. The hidden message may be plaintext, ciphertext, or anything that can be represented as a bit stream. According to Potdar, Khan, Chang, Ulieru and Worthington:

> *"We focus on the use of steganography to achieve the goal of disguising the existence of secret communication and steganalysis to recover secret communication. If the existence of communication is discovered, as such the e-forensics system proves successful. We think steganalysis can be useful companion to e-forensics."*
>
> (2004, para. 4)

The contents of the thesis is for internal user only

hidden information. The study could be improved by using colour digital images. Here we could see the effect caused by the embedding scheme towards the three colour channel; red (R), green (G) and blue (B).

There are several aspects of the study that could be enhanced for future studies. Since the result of the study only to proof that we can estimate some of the parameters of embedding algorithm, it could only provide a solid platform for other research and studies in active steganalysis especially in the attempt to estimate the secret message but more parameters of embedding algorithm could be taken into consideration for future studies thus aiding the effort to achieve the main goal of steganalysis, which is extracting the embedded secret message regardless of its types or version.

REFERENCES

Anderson, R.J & Petitcolas, F.A.P. (1998*). On The Limits Of Steganography*, IEEE Journal of Selected Areas in Communications, Special Issue on Copyright and Privacy Protection, 16(4), 474-481.

Aura, T. (1996). Practical *Invisibility In Digital Communication*, Lecture Notes in Computer Science, vol.1174, Springer-Verlag, 265-278.

Bromiley, P.A., Thacker, N.A., & Courtney, P. (2002). *Non-Parametric Image Subtraction Using Grey Level Scattergrams*, Image and Vision Computing, 20, 609-617.

Chalapathi V.N. (2003). *Steganalysis for BMP Image*, Indian Institute of Technology, Kanpur, India.

Chandramouli, R., & Memon, N. (2001). *Analysis of LSB based Image Steganography Techniques*, *Proceedings of ICIP 2001*, Thessaloniki, Greece, October 7-10, 2001.

Chandramouli, R. (2002). *Mathematical Approach to Steganalysis,* Proceeding of the SPIE Security and Watermarking of Multimedia Contents IV, vol. 4675, 21-24

Chadramouli, R. (2003). A *Mathematical Framework for Active Steganalysis*, ACM Multimedia Systems, Vol. 9-3, 303-311.

Chandramouli, R. & Memon, N. (2003) *Steganography Capacity: A Steganalysis Perspective*, Proceeding of the SPIE Security and Watermarking of Multimedia Contents V, vol. 5020.

Dang, X. & Kota, K. C. S. (2006) *Case Study: An Implementation of a Secure Steganoigraphic System*, In Proceeding of International Conference of Security and Management (SAM 2006), Las Vegas, 26-29.

Dittmann, J. & Kraetzer, C. (2006). *Audio Benchmarking Tools and Steganalysis*, ECRYPT, Eurapoen Network of Excellence in Cryptology. Retrieved July 24, 2006 from www.cs.uni-magdeburg.de/-kraetzer/publications/DWVL10.pdf.

Farid, H. (2001). *Detecting Steganographic Message in Digital Image*, Report TR2001-412, Dartmouth College, Hanover, NH.

Foss, K. (2001, October 27). *Does The Medium Conceal The Message?* Retrieved November 25, 2006 from http://www.workopolis.com/servlet/News/fasttrack/20011027/FOCSCIE

Fridrich, J., Du, R. & Meng, L. (2000). *Steganalysis of LSB Encoding in Color Images*, Proceedings IEEE International Conference on Multimedia and Expo, New York City, NY.

Fridrich, J., Goljan M., & Du, R. (2001a). *Steganalysis Based On JPEG Compatibility*, Proceedings Of SPIE: Multimedia Systems And Applications IV, Denver, 275–280.

Fridrich, J., Goljan, M. & Du, R. (2001b). *Reliable Detection of LSB Steganography in Grayscale and Color Images*, Proc. ACM, Special Session on Multimedia Security and Watermarking, Ottawa, Canada, 27–30.

Fridrich, J., Goljan, M., & Du, R. (2001c). *Detecting LSB Steganography in Color and Gray-Scale Images*, Magazine of IEEE Multimedia, Special Issue on Security, 22–28.

Fridrich, J, Goljan, M., & Soukal, D. (2003a). *Higher-Order Statistical Steganalysis Of Palette Images,* Proceedings Of SPIE: Electronic Imaging 2003, Security And Watermarking Of Multimedia Contents, Santa Clara, CA.

Fridrich, J., Goljan, M., & Hogea, D. (2003b). *Steganalysis of JPEG Images: Breaking The F5 Algorithm.* Lecture notes in computer science, vol 2578. Springer, Berlin Heidelberg NewYork, 310– 323.

Grgic, S., Grcig, M. & Zovko-Cihlar, B. (2001). *Performance Analysis of Image Using Wavelets*, IEEE Transaction on Industrial Electronics, vol.48-3, 682-695.

Jackson, J.T., Gunsch, G.H., Claypoole, R.L. & Lamont, G.B. (2003) *Blind Steganography Detection Using a Computational Immune System Approach: A Proposal.* International Journal of Digital Evidence. Retieved on 25th November, 2006 from http://www.ijde.org/docs/02_winter_art4.pdf.

Johnson, N.F., & Jajodia, S. (1998a). *Exploring Steganography: Seeing the Unseen*, IEEE Computer, pp.26-34.

Johnson, N.F., & Jajodia, S. (1998b). *Steganalysis of Images Created Using Current Steganography Software*, Lecture Notes in Computer Science, vol.1525, Springer-Verlag, Berlin, pp. 273-289.

Johnson, N.F., & Jajodia, S. (1998c) *Steganalysis: The Investigation of Hidden Information*, In Proceeding of the IEEE Information Technology Conference.

Katzenbeisser, S., & Petitcolas, F.A.P. (2000). *Information Hiding Techniques For Steganography And Digital Watermarking,* Artech House, Boston

Katzenbeisser, S., & Petitcolas, F.A.P. (2002). *On Defining Security in Steganographic Systems*, Proc. Electronic Imaging, Photonics West, SPIE 2002, San Jose, California,

Kelley, J. (2001, 19 June). *Terror Groups Hide Behind Web Encryption*, USA Today. Retrieved November 25, 2006 from http://www.usatoday.com/life/cyber/tech/2001-02-05-binladen.htm

Ker A. D. (2005) *Steganalysis of LSB Matching in Grayscale Images*, IEEE Signal Processing Letters, Vol. 12(6), pp 441-444.

Kessler, G.C. (2004a). *An Overview of Steganography for The Computer Forensics Examiner*, Forensic Science Communication, Vol 6-3.

Kessler, G.C. (2004b). *Steganography: Implications for the Prosecutor and Computer Forensics Examiner*, American Prosecutors Research Institute, Child Sexual Exploitation Program Update, Vol. 1-1. Retrieved November 6, 2006 from http://www.ndaa.org/publications/newsletters/child_sexual_exploitation_update_volume_1_number_1_2004.html.

Key, J.P. (1997). *Research design in Occupational Education*, Oklahoma State University, Retrieved 25 March 2005 from www.okstate.edu/ag/agedcm4h/academic/aged5980a/5980/newpage.html.

Lenti, J. (2000). *Steganographic Methods*, Periodica Polytechnica SER. EL. ENG., Vol. 44, pp. 249-258

Liu Q. & Sung A. H. (2007) *Feature Mining and Neuro-Fuzzy Inference System for Steganalysis of LSB Matching Steganography in Grayscale Images*, International Joint Conference on Artificial Intelligence 2007, pp 2808-2813.

Marvel, L.M., Boncelet, C.G., & Retter, C.T. (1998*). Reliable Blind Information Hiding for Images*, Lecture Notes on Computer Science, vol. 1525, Springer-Verlag, New York, 1998, pp. 48–61.

Ojala, T., Valkealahti K, Oja, E. & Pietikäinen, M. (2001). *Texture Discrimination with Multidimensional Distributions of Signed Gray Level Differences*. Pattern Recognition 34:727-739

Ozer, H., Avcibas, I., Sankur, B., & Memon, N. (2003). *Steganalysis of Audio based on Audio Quality Metrics*, Security and watermarking of multimedia contents V, SPIE Santa Clara. Retrieved July 20, 2005 from www.busm.ee.boun.edu.tr/~sankur/SankurFolder/Audio_Steganalysis_16.doc

Petitcolas, F.A.P., Anderson, R.J., & Kuhn, M.G. (1999). *Information Hiding-A Survey*, In Proceeding of the IEEE, special Issue on Protection of Multimedia Content, 87(7), pp. 1062-1078.

Potdar, V., & Chang, E. (2004) *Grey Level Modification Steganography for Secret Communication*, Proceedings of the 2nd IEEE International Conference on Industrial Informatics, June 27-29, 2004, Berlin, Germany, pp. 223-228.

Potdar, V., Khan, M.A., Ulieru, M and Worthington, P.R (2004). *e-Forensics Steganography System for Terrorist Information Retrieval*,

Provos, N. (2001). *Defending Against Statistical Steganalysis, 10th USENIX Security Symposium*, Washington, DC.

Provos N. & Honeyman, P. (2001). *Detecting Steganographic Content on the Internet, CITI Technical Report 01-11,* August 2001, submitted for publication.

Provos, N. & Honeyman, P (2003). *Hide and Seek: An Introduction to Steganography*, IEEE Security and Privacy, vol. 1-3, pp. 32-44.

Ross, S.M. & Morrison, G.R. (2004). *Experimental Research Methods*, In D.J. Jonasson (Ed.). Handbook of Research on Educational Communications and Technology, (2nd Edition, pp 1021-1043). Mahwah N.J.: Lawrence Exlbaum Associates Publishers.

Sahoo G., & Tiwari R. K. (2008) *Designing an Embedded Algorithm for Data Hiding using Steganographic Technique by File Hybridization*, International Journal of Computer Science and Network Security, Vol. 8, pp 228-233.

Shi, Y.Q., Xuan G., Zou D., Gao J., Yang C., Zhang Z., Chai P., Chen W., & Chen C. (2005). *Steganalysis Based on Moments of Characteristic Functions Using Wavelet Decomposition, Prediction-Error Image, and Neural Network*, IEEE International Conference on Multimedia and Expo 2005.

Trivedi, S. & Chandramouli, R. (2003). *Active Steganalysis of Sequential Steganography*, SPIE Conference California 2003, Vol. 44-13, pp. 123-130.

Wang, H. & Wang, S. (2004). *Cyber Warfare: Steganography vs. Steganalysis,* Communication of the ACM, vol. 47-10, pp. 76-82.

Westfeld, A. & Pfitzmann, A. (2000). *Attacks on Steganographic Systems*, Lecture Notes in Computer Science, vol.1768, Springer-Verlag, Berlin, 2000, pp. 61-75.

Young, I.T., Gerbrands, J.J., & Van Vliet, L.J. (1995). *Fundamentals of Image Processing*, Vol. ISBN 90-75691-01-7, Delft: PH Publications.