A DETECTION METHOD FOR TEXT STEGANALYSIS USING EVOLUTION ALGORITHM (EA) APPROACH

PURIWAT LERTKRAI

UNIVERSITI UTARA MALAYSIA
2012

DETECTION METHOD FOR TEXT STEGANALYSIS USING EVOLUTION ALGORITHM (EA) APPROACH

A project submitted to Dean of Research and Postgraduate Studies Office in partial
Fulfillment of the requirement for the degree
Master of Science (Information Technology)
Universiti Utara Malaysia

By
Puriwat Lertkrai

Permission to Use

In presenting this project in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence by the Dean of Postgraduate and Research. It is understood that any copying or publication or use of this project or parts 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 to Universiti Utara Malaysia for any scholarly use which may be made of any material from my project.

Requests for permission to copy or to make other use of materials in this project, in whole or in part, should be addressed to

Dean of Research and Postgraduate Studies
College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman
Malaysia

Abstract

The ability of sending a secret message through a network nowadays has become a more challenging and complex process. One of the reasons why we need tools to detect the hidden message is because of security and safety. Secret messages can be both for good or bad, and subversive groups have been known to send secret messages to coordinate their terrorist activities. Thus, a technique such as steganalysis is one method example to detect a secret message. Much of the technical steganalysis work had been carried out on image, video, and audio steganalysis, but in any agency or organisation, all business documents generally uses the natural language in text form or document form. Therefore, this research employed a detection factor based on the evolution algorithm method for text steganalysis. The aim of this project was to detect a hidden message in an observed message using text steganalysis.

Acknowledgement

Firstly, I would like to thank to my supervisors, Mr. Roshidi Din and Dr. Shahrul Azmi always have faith in me and give ideas and suggestions to complete this project. I also would like to use this opportunity to thank my evaluators, Dr. Azizi Ab Aziz and Mdm. Nur Azzah Abu Bakar, who had given me the comments for improving my project report. I used what I had learnt in this project.

I am most grateful to my family. I am deeply and forever indebted to the people in my life that touched my heart and gave me the strength to move forward for achieving something better, especially to my beloved father, mother, and my young sister, Mr. Siri, Mrs. Suvapa, Ms. Doungdusade for their love and support.

I would also like to thank to my two best friends in UUM, Sittichai and Jutaporn. Thank for your never ending encouragement to complete this project.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

SVM	Support Vector Machine
CI	computational intelligence
EA	Evolution algorithm
JEAP	Java Evolution Algorithms Package (JEAP)
JGAP	Java Genetic Algorithm Package
PoS	Part of Speech
PCFG	Probabilistic Context-Free grammar
TID	Topic Identification
IDF	Inverse Document Frequency
DMR	Degree of Machine Reversibility
DMP	Degree of Machine Preference
MT	Machine Translation
TBS	Translation-Based Steganography
GA	Genetic Algorithms
SI	Swarm Intelligence
EC	Evolutionary Computation
TSP	Travelling Salesman Problem
EDSS	Evolution Detector Steganalysis System
QIM	Quantization Index Modulation
MRF	Markov Random Field

CHAPTER ONE
INTRODUCTION

1.1 Introduction

This chapter provides an overview of the entire study. The first section describes the background of the study that leads to the implementation of the whole research. This is followed by problem statement, research questions, objectives, scope of the study, significance of the study, and research organisation. The last section provides the way this research report is summarised.

1.2 Background of the Study

In recent decades, computer development and the expansion to use in different areas of life and work has come to the fore, and one of the many issues that is being raised is security of information which has gained special significance. This is because the information is now being treated as a commodity or a resource comparable to labour and capital (Hillman, 1982). One of the concerns in the area of information security is the concept of information hiding (Shirali-Shahreza and Shirali-Shahreza, 2006). The information hiding concept has received attention from the research community and from industry since before the 1990s and it is changing fast since the first academic conference on the subject organised in 1996. There is a difference between information hiding and encrypting where the goal of information hiding is to avoid intrusion or discovery of hidden data. Meanwhile, the goal of encrypting is to restrict data access.

REFERENCES

Abraham, A., Guo, H., and Liu, H. (2006). Swarm Intelligence: Foundations, Perspectives and Applications, *26*, 3-25. Retrieved from http://www.softcomputing.net/swarm-chapter.pdf

Cachin, C. (2004). An Information-Theoretic Model for Steganography. *Information and Computation Academic Press, 192*(1), 1-14. Retrieved from http://www.zurich.ibm.com/~cca/papers/stego.pdf

Chandramouli, R., and Memon, N. (2001). Analysis of LSB based image steganography techniques. *Proceedings of the International Conference on Image Processing, 3*, 1019-1022.

Chandramouli, R., and Subbalakshmi, S. K. (2007). In Steganalysis : A Critical Survey, Control, Automation, Robotics and Vision Conference (ICARCV), *2*, 964- 967.

Changder, S., Debnath, N. C., and Ghosh, D. (2011). A Greedy Apporach to Text steganography using Perperties of Sentences. *Eighth Internation Conference on Information Texhnology: New Generations*, 30-35.

Chapman, M., Davida G. I., and Rennhard M., (2001). A Practical and Effective Approach to Large-Scale Automated Linguistic Steganography. *Proceedings of the Information Security Conference (ISC '01),* 156-165.

Chen, Z., Huang, L., Yu, Z., Wei, Y., Li, L., Zhen, X., and Zhao, X. (2008). *Linguistix Steganography Detection Using Statistical Characteristics of Correlations between Words,* 224-235.

Chiang, Y. L., Chang, L. P., Hsieh, W. T., and Chen, W. C. (2004). Natural Language Watermarking Using Semantic Substitution for Chinese Text, 129-104. Retrieved from http://www.cs.uta.fi/~jt68641/tutkimuskurssi/ccr6.pdf

Chong, C. S., Low, M. Y. H., Sivakumar, A. I., and Gay, K. L. (2007). Using a Bee Colony Algorithm For Neighborhood Search in Job Shop Scheduling Problems. *21st European Conference on Modelling and Simulation.*

Davila, J. (1999). Genetic optimization of NN topologies for the task of natural language processing. *International Joint Conference on Neural Networks, 2*, 821-826.

Desoky, A., Younis, M., and El-Sayed, H. (2008). Auto-Summerization-Based Steganography. *International Conference on Innovations in information Technology*, 608-612.

Doerr, G. A., and Dugelay, J. L. (2004). Security Pitfalls of Frameby-Frame Approaches to Video Watermarking. *IEEE Transactions on Signal Processing, 51*(10), 2955-2964.

Fard, A. M., Akbarzadeh-T, M. R., and Varasteh-A, F. (2006). A New Genetic Algorithm Approach for Secure JPEG Steganography. *IEEE International conference on Engineering of Intelligent Systems*, 1-6.

Fleischer, M. (2003). Foundations of Swarm Intelligence:From Principles to Practice. *Swarming: Network Enabled C4ISR*, 1-13. Retrieved from http://arxiv.org/pdf/nlin/0502003.pdf

Fogel, D. B. (1997). The advantages of Evolutionary Computation. *Biocomputing and emergent computation: Proceedings of BCEC97*, 1-11.

Garnier, S., Gautrais, J., and Theraulaz, G. (2007). The biological principles of swarm intelligence. *Swarm Intell (2007)*, 3-37. Retrieved from http://cognition.ups-tlse.fr/_guyt/documents/articles/66.pdf

Geetha, S., Sivatha Sindhu, S.S., and Kannan, A. (2006). StegoBreaker: Audio Steganalysis using Ensemble Autonomous Multi-Agent and Genetic Algorithm. *India Conference, 2006 Annual IEEE*, 1-6.

Gopalan, K. (2003). Audio steganography using bit modification. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03), 2*, 421-424.

Hillman, D. J. (1982). Problems of the Information Age. *ACM '82 Proceedings of the ACM '82 conference*, 190-191.

Hinterding, R., Michalewicz, Z., and Eiben, A. E. (1997). Adaptation in Evolutionary Computation: A Survey. *IEEE International Conference on Evolutionary Computation*, 65-69.

Hlaing, Z. C. S. S., and Khine, M. A. (2011). An Ant Colony Optimization Algorithm for Solving Traveling Salesman Problem. *2011 International Conference on Information Communication and Management, 16*, 54-59.

Huang, H.-j., Sun, X.-m., Sun, G., and Huang, J.-w. (2007). Detection of Hidden Information in Tags of Webpage Based on Tag-Mismatch. *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 1,* 257-260.

Johnson, N. F., and Jajodia, S. (1998). Steganalysis: The Investigation of Hidden Information. *Proceeding of IEEE Information Technology Conference*, 113-116.

Johnson, N. F. and Katzenbeisser S. (2000). A Survey of Steganographic Techniques. *Information Hiding: Techniques for Steganography and Digital Watermarking*, 43-78.

Karaboga, D., and Akay, B. (2009). A comparative study of Artificial Bee Colony algorithm. *Applied Mathematics and Computation 214, 108-132.*

Kevin, C., and Karen, B. (2006). An Evaluation of Image Based Steganography Methods. *Multimedia Tools and Applications, 30*(1), 55-88.

Kharullah, M. (2009). A Novel Text Steganography System Using Font Colors of the invisible Characters in Microsoft Word Documents. *2009 International Conference on Computer and Electrical Engineering*, 482-484.

Lenti, J. (2000). Steganographic Methods. *Periodica Polytechnica Ser. El. Eng., 44,* 249-258. Retrieved from http://www.pp.bme.hu/ee/2000_3/pdf/ee2000_3_04.pdf

Li, L., Huang, L., Zhao, X., Yang, W., and Chen, Z. (2008). A Statistical Attack on a Kind of Word-Shift Text-Steganography. *Internetional Conference on Intelligent Information Hiding and Mulitmefia Signal Processing*, 1503-1507.

Lin, I. C., and Hsu, P. K. (2010). A Data Hiding Secheme on Word Documents Using Multiple-base Notation System. *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 31-33.

Liu, T. T., and Tsai, W. H. (2007). A New Steganographic Method for Data Hiding in Microsoft Word Documents by a Change Tracking Technique. *IEEE Transactions on Information Forensics and Security*, 2, 24-30.

Mathew, T. V. (2006). Genetic Algorithm, 1-15. Retrieved from http://www.civil.iitb.ac.in/tvm/2701_dga/2701-ga-notes/gadoc.pdf

Meng, P., Huang, L., Chen, Z., Yang, W., and Li, D. (2008). Linguistic Steganography Detection Based on Perplexity. *2008 International Conference on MulitMedia and Information Technology*, 217-220.

Meng, P., Hang, L., Yang, W., and Chen, Z. (2009). Attack on Translation Based Steganaography. *IEEE Youth Conference on Information, Computing and Telecommunication*, 227-230.

Munassar, N. M. A., and Govardhan, A. (2010). A Comparison Between Five Models Of Software Engineering. *IJCSI International Journal of Computer Science Issues, 7*(8), 97- 101.

Petitcolas, F. A. P., Anderson, R. J. and Kuhn, M. G. (1999). Information Hiding : A Survey. *Proceedings of the IEEE, special issue on protection of multimedia content, 87*(7), 1062-1078.

Pham, D. T., and Karaboga, D. (2000). *Intelligent Optimisation Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks*.

Popov, A. (2005). Genetic Algorithms for Optimization – Application in Controller Design Problems, 1-21. Retrieved from http://p0p0v.com/science/downloads/Popov05a.pdf

Roshidi, D., ANI, Z. C., and SAMSUDIN, A. (2012). A Formulation of Conditional States on Steganalysis Approach. *Wseas Transactions on Mathematics, 11*(3), 173-182.

Roshidi, D., and Samsudin, A. (2009). Digital Steganalysis: Computational Intelligence Approach. *School of Computer Sciences, Universiti Sains Malaysia (USM)*.

Royce, W. W. (1970). Managing The Development of Large Software Systems, 1-9. Retrievedfrom http://leadinganswers.typepad.com/leading_answers /files/original_waterfall_paper_winston_royce.pdf

Salami, N. M. A. A. (2009). Ant Colony Optimization Algorithm. *UbiCC Journal, 4*, 823-829.

Saab, S. M., El-Omari, N. K. T., and Owaied, H. H. (2009). Developing Optimization Algorithm Using Artificial Bee Colony System. *Ubiquitous Computing and Communication Journal, 4*, 15-19.

Schemmel, J., and Meier, K. (2011 May 19). Introduction to Evolutionary Algorithms. Retrieved 25, 2012, from Heidelberg University website: http://www.kip.uniheidelberg.de/cms/vision/projects/recent_projects/hardware_perceptron_systems/training_with_evolutionary_algorithms/introduction_to_evolutionary_algorithms/

Shirali-Shahreza M. H., and Shirali-Shahreza M., (2006). A New Approach to Persian/Arabic Text Steganography. *Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science and 1st IEEE/ACIS,* 310-315.

Shirali-Shahreza, M., and Shirali-Shahreza, M. H. (2007). Text steganography in SMS. *International Conference n Convergence information Technology,* 2260-2265. doi:10.1109/ICCIT.2007.369

Shirali-Shahreza, M. (2008). Text Steganography by Chaning Words Spelling. *Internation Conference on Advanced Communication Technology, 3,* 1912-1913.

Shirali-Shaherza, M., and Shirali-Shaherza, S. (2008). Seganography in TeX Documents. *3rd Internetional Conference on Intelligent System and Knowledge Engineering, 1*, 1363-1366.

Shyu, S. J., Yin, P. Y., and Lin, B. M. T. (2004). An Ant Colony Optimization Algorithm for the Minimum Weight Vertex Cover Problem. *Annals of Operations Research, 131,* 283-304.

Wang, G., and Zhang, W. (2006). A Steganalysis-Based Approach to Comprehensive Identification and Characterization of Functional Regulatory Elements, *7*(6). Retrieved from http://www.biomedcentral.com/content/pdf/gb-2006-7-6-r49.pdf

Wang, Y., He, Q., Wang, H., Yin, B., and Ding, S. (2010). Steganographic Method Based on Keyword Shit. *2nd IEEE International Conference on Information Management and Engineering,* 454-456.

Weise, T. (2009). Global Optimization Algorithms Theory and Application, 1-820. Retrieved from http://www.it-weise.de/projects/book.pdf

Westfeld, A., and Wolf, G., (1998). Steganography in a Video Conferencing System. *Proceedings of Information Hiding - Second International Workshop,* 32-47.

Whitley, D. (1994). Agenetical Gorithm Tutorial. *Stat&tics and Computing 4*, 65-85. Retrieved from http://sci2s.ugr.es/docencia/cursoMieres/Whitley94.pdf

Whitley, D. (2001). An Overview of Evolutionary Algorithms: Practical Issues and Common Pitfalls. *Information and Software Technology 43,* 817-831.

Wu, G. (2000). A Neural Network used for PD Pattern Recognition in Large, Turbine Generators with Genetic Algorithm. *Conference Record of the 2000 IEEE International Symposium on Electrical Insulation*, 1-4.

Wyant, G. (2008). Drowning in the Waterfall: The Benefits of Agile versus the Predominance of Waterfall, 1-8. Retrieved from http://uenics.evansville.edu/~hwang/f08-courses/cs390/GuyWyantPaper.pdf

Yazdani, N., Herrmann, H., Daolio, F., and Tomassini, M. (2011). EA opitmization of Networks Against Malicious Attacks, 1-10. Retrieved from http://www.comphys.ethz.ch/hans/p/558.pdf

Xiang, L., Sun, X., Luo, G., and Gan, C. (2007). Research on Steganalysis for Text Steganography Based on Font Format. *Third International Symposium on Information Assurance and Security,* 490-495.

Yu, Z., Huang, L., Chen, Z., Li, L., Zhao, X., and Zhu, Y. (2009). Steganalysis of Synonym-Substitution Based Natural Language Watermarking. *International Journal Of Multimedia and Ubiquitous Engineering, 4*, 21-34.

Zhao, X., Huang, L., Li, L., Yang, W., Chen, Z., and Yu, Z. (2009). Steganalysis on Character Substitution Using Support Vector Machine. *Second International Workshop on Konwledge Discoverty and Data Mining*, 84-88.

Zhi-li, C., Lius-heng, H., Yu, Z., Zhao, X., and Xue-ling, Z. (2008). Effiective Linguistic Steganography Detection. *IEEE 8th International Conference on Computer and Information Technology Workshops*, 224-229.

Zuxu, D., Fan, H., Muxiang, Y., and Guohua, C. (2007). Text Information Hiding Based on Part of Speech Grammar. *Internation Conference on Computational Intelligent and Security Workshops*, 632-635.