UNIVERSITI PUTRA MALAYSIA

STEGANOGRAPHY FOR EMBEDDING DATA IN DIGITAL IMAGE

SALAH IBRAHEM SOWAN

FK 2003 25

STEGANOGRAPHY FOR EMBEDDING DATA IN DIGITAL IMAGE

By

SALAH IBRAHEM SOWAN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Partial Fulfilment of Requirements for the Degree of Master of Science

March 2003

In the name of God, Most Gracious, Most Merciful

ii

Dedication to

My father Mr. Ibrahim Sowan

My family

STEGANOGRAPHY FOR EMBEDDING DATA IN DIGITAL IMAGE

By

SALAH IBRAHEM SOWAN

March 2003

Chairman: Abdul Rahman Ramli, Ph.D.

Faculty: Engineering

The growth of the World Wide Web (WWW) has enabled the personal computer to be used as a general communications tool. As in the case of other forms of communication there is a wish for security and privacy. With literally millions of images moving on the Internet each year, it is safe to say that digital image Steganography is of real concern to many in the IT security field. Digital images could be used for a number of different types of security fear. In the business world, the sending of a harmless looking bitmap file could actually hide the latest company secrets.

Steganography (literally, covered writing) is concealing of a secret message within another seemingly innocuous message, or carrier. Digital carriers include e-mail, audio, and images. Steganography, like cryptography, is a means of providing

secrecy. Steganography does so by hiding the very existence of the communication, while cryptography does so by scrambling a message so it cannot be understood. A cryptography message can be intercepted by an eavesdropper, but the eavesdropper may not even know the existence of a steganographic message.

This thesis discusses the issues regarding Steganography and its application to multimedia security and communication, addressing both theoretical and practical aspects, and tackling both design and attack problems. In the fundamental part, we identify a few key elements of Steganography through a layered structure. Data hiding is concerned to be as a communication problem where the embedded data is the signal to be transmitted. The tradeoff for two major categories of embedding data using spatial domain and frequency domain will be discussed.

In addition, we have found that unevenly distributed embedding capacity brings difficulty in data hiding. We propose a complete solution to this problem, addressing considerations for choosing constant or variable embedding rate and enhancing the performance for each case. In the design part, we present new data hiding algorithms for binary images, grayscale and color images, covering such applications as annotation, fingerprinting, and ownership protection.

Tesis abstrak yang dikemukan kepada Senat Universiti Putra Malaysia sebagai memenuhi sebahagian syarat-syarat Ijazah master Sains

PENGGUNAAN TEKNIK STEGANOGRAFI UNTUK PENEMPELAN DATA DI DALAM IMEJ DIGITAL

Oleh

SALAH IBRAHEM SOWAN

Mac 2003

Pengerusi:　　**Abdul Rahman Ramli, Ph.D.**

Faculti:　　　**Kejuruteraan**

Peningkatan penggunaan Internet dan anjakan paradigma dengan WWW telah membolehkan komputer peribadi digunakan sebagai alat komunikasi umum. Keupayaan ini telah menggalakkan penggunaan teknik-teknik tertentu untuk tujuan keselamatan dan kerahsiaan. Imej-imej yang ditukar sesama pengguna Internet melalui WWW telah menjangkau kepada jumlah yang sangat banyak telah memudahkan pihak yang ingin berkomunikasi secara rahsia menggunakan steganografi imej. Imej-imej digital boleh digunakan untuk pelbagai jenis ancaman keselamatan. Di dalam dunia korporat, penghantaran sesuatu file imej bitmap mungkin mengandungi rahsia teknologi terkini syarikat.

Steganografi adalah suatu kaedah menyembunyikan mesej rahsia di dalam mesej biasa atau pembawa digital. Pembawa digital adalah seperti e-mail, audio dan imej. Steganografi adalah kaedah untuk yang menyediakan kerahsiaan seperti

kriptografi. Walaubagaimanapun, steganografi menyediakan kerahsiaan dengan menyembunyikan sebarang aktiviti komunikasi manakala kriptografi pula menjadikan sesuatu komunikasi rahsia agar ia tidak difahami. Seseorang yang ingin mencuri dengar sebarang komunikasi mungkin tidak langsung menyedari bahawa sebenarnya komunikasi sedang berlangsung dan tidak mengetahui akan kewujudan mesej-mesej rahsia yang dihantar.

Tesis ini membincangkan isu-isu mengenai steganografi dan aplikasinya kepada keselamatan multimedia dan komunikasi. Tesis ini membentangkan aspek-aspek teori dan praktikal serta menangani masalah-masalah rekabentuk dan serangan. Pada permulaan kajian, kami mengenalpasti beberapa komponen utama steganografi melalui struktur yang tersusun. Penyembunyian data adalah dimodelkan sebagai masalah komunikasi di mana data yang benam hendak dihantar ke destinasi. Kami mengkaji tolak-ansur yang perlu dilakukan untuk dua jenis kategori utama mekanisme penempelan. Selain itu, kami mendapati bahawa keupayaan penbenaman teragih yang tidak sama rata menyebabkan kesukaran di dalam penyembunyian data. Kami mencadangkan penyelesaian yang meneyeluruh kepada masalah ini dengan memilih kadar penbenbenaman pemalar atau pemboleh-ubah. Justeru itu kami dapat meningkatkan keupayaan untuk setiap suatu masalah. Di dalam bahagian rekabentuk, kami mempersembahkan algoritma baru penyembunyian data untuk imej binari, skala kelabu dan berwarna yang merangkumi aplikasi seperti anotasi, cap jari dan perlindungan hakmilik.

ACKNOWLEDGMENTS

First of all, I would like express my utmost thanks and gratitude to Almighty Allah S.W.T for giving me the ability to finish this thesis successfully.

The author gratefully expresses his profound appreciation and gratitude to his supervisor, Dr.Abdul Rahman Ramli, for his supervision, guidance, supporting, and constructive suggestions and comments throughout the duration of the project until it turn to real success

The author is also indebted to members of his supervisory committee, Dr V Prakash and Mr. Syed Abdul Rahman Al-Haddad, for their affectionate guidance, prompts decision and valuable assistance during this period.

The author is also like to thank lecturers, staffs, technicians of faculty of engineering for providing the facilities required for undertaking this project.

The author would like to thank his family for the encouragement and support without which is impossible for the success of this project, and my friends for offering helps all the time.

I certify that an Examination Committee met on 21st March 2003 to conduct the final examination of Salah Ibrahim Sowan on his Master of Science thesis entitled "Steganography for Embedding Data in Digital Image" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981 The Committee recommends that the candidate be awarded the relevant degree Members of the Examination Committee are as follows

Mohamad Khazani, Ph.D.
Associate Professor,
Department of Computer and Communication Engineering,
Faculty of Engineering,
Universiti Putra Malaysia
(Chairman)

Abdul Rahman Ramli, Ph.D.
Associate Professor,
Department of Computer and Communication Engineering,
Faculty of Engineering,
Universiti Putra Malaysia
(Member)

V.Prakash, Ph.D.
Department of Computer and Communication Engineering,
Faculty of Engineering,
Universiti Putra Malaysia
(Member)

Syed Abdul Rahman Al-Haddad,
Department of Computer and Communication Engineering,
Faculty of Engineering,
Universiti Putra Malaysia
(Member)

GULAM RUSUL RAHMAT ALI, Ph.D.
Professor / Deputy Dean,
School of Graduate Studies,
Universiti Putra Malaysia

Date - **5** MAY 2003

The thesis submitted to the Senate of Universiti Putra Malaysia has been accepted as fulfillment of the requirement for the degree of Master of Science. The members of the Supervisory Committee are as follows:

Mohamad Khazani, Ph.D.
Associate Professor,
Department of Computer and Communication Engineering,
Faculty of Engineering,
Universiti Putra Malaysia.
(Chairman)

Abdul Rahman Ramli, Ph.D.
Associate Professor,
Department of Computer and Communication Engineering,
Faculty of Engineering,
Universiti Putra Malaysia.
(Member)

V.Prakash, Ph.D.
Department of Computer and Communication Engineering,
Faculty of Engineering,
Universiti Putra Malaysia.
(Member)

Syed Abdul Rahman Al-Haddad,
Department of Computer and Communication Engineering,
Faculty of Engineering,
Universiti Putra Malaysia.
(Member)

AINI IDERIS, Ph.D.
Professor / Dean,
School of Graduate Studies,
Universiti Putra Malaysia.

Date: 1 0 JUL 2003

ix

DECLARATION

I hereby declare that the thesis is based on my original work except for equations and citations, which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

_____
(SALAH IBRAHIM SOWAN)

Date: 05 /05 /2003

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1-1 Background

The Internet has become the latest medium not just for communication through the advanced features available on the Internet such as file transfers, transaction applications, advertising and much more. The rapid expansion of the Internet in the past has rapidly increased the availability of digital data such as audio, images and videos to the public.

Just a few years were necessary before the academic and scientific network project known as the Internet became the medium through which people communicate, conduct business, exchange data, access information, and interact with systems and work. As the ever-increasing demand for the Internet bandwidth is satisfied with new technologies, the amount of information being offered for public access grows at an surprising rate. How can people deal with this amount of information? How can one maximize the information transferred to the user in the limited time that people have available to browse the Internet?

The answers to these questions came with the advent of multimedia applications. Systems that can mix sound, pictures, video and text are now standard

when one needs to interact with large amounts of information (e.g. business, distance education, human-machine interface and specialist systems applications).

With the development of Internet technologies, digital media can be transmitted conveniently over the network. Therefore, protecting secret messages during transmission has become an important issue. Using classic cryptography only, the encrypted message becomes clutter data that cannot pass the checkpoint on the network. Steganography provides another layer of protection on the secret message, which will be embedded in another media such that the transmitted data is meaningful and innocuous to everyone. Compared with cryptography techniques attempting to conceal the content of messages, Steganography conceals the existence of the secret message.

With literally millions of images moving on the Internet each year, it is safe to say that digital image Steganography is of real concern to many in the IT security field. In the corporate world, the sending of a harmless looking bitmap file could actually conceal the latest company secrets. JPEGs could be used by the government to hide the latest military secrets. It is believed that the September -11- 2001 crash in New York had aircraft configuration plans sent to them hidden inside a digital image. It is felt that the use of Steganography has allowed the communicate without the fear of being caught. (Fixmar, 2001)

Steganography is closely related to the fields of information hiding and watermarking. These three fields have a great deal of overlap and share many

technical approaches. However, there are three fundamental philosophical differences that affect the requirements, and thus the design, of a technical solution.

It is not surprising that Steganography has enjoyed resurgence in today's computerized world. As computers continue to permeate millions of people's daily Routines, their use as Steganography instruments makes perfect sense. Steganography takes advantage of covers that are commonplace – a role that computers fill in society today. Steganography's rise in popularity can be attributed, in part, to the United States government's cryptographic material export prohibition. This has driven some people to use Steganography as a means to reduce the casual interception of private information. Another reason for the increase in Steganography usage is due to the cover space abundance provided by digital media, particularly within the various computers file formats (e.g. BMP, GIF, JPG, PDF, WAV, HTML, TXT etc). With these almost-perfect digital media and the continuous technological advancements, there has been a rising concern for copyright abuses. This has driven much of the Steganography advancements with an immense focus on digital watermarking. This promising technology is proclaimed by the industry as an excellent anti-fraud and forgery mechanism.

The music and movie industries have invested millions of dollars on techniques to conceal company logos and other proprietary markings in digital images, videos, and music recordings. The interest in creating a robust, tamperproof digital fingerprint has been the focus of much of the academic research in Steganography. Consequently, this anti-piracy technology has created a corresponding interest in basic Steganographic methods. Although this interest has

increased, there are relatively few companies that have tried to capitalize on any commercial Steganography products.

1.2 Thesis scope

The scope of this thesis is to organizational theory and design, and computer design and connectivity will be presented as necessary to clearly understand the practical applications of steganography. This research effort focuses on government Microsoft PCs processing sensitive, unclassified information in the national office automation environment. These machines are the most widely used. Embedded, or special purpose, computer systems were specifically excluded. In the case of networked systems, the research applies to the individual PCs that make up the network and leaves the peculiarities of interconnected computer networks and the various network architectures that support them to future research efforts. Ongoing research in operating system security and secure system development must be considered separate topics that cannot be addressed within the scope of this research.

1.3 Problem Definition

While the technology for information transmission has changed much with the advent of the digital age, the need for concealing information is as present today as it has ever been. In recent years, much attention has been paid to steganographic applications using digital images. Considering the frequency with which such images are transmitted on a daily basis, they serve as perfect containers for hidden messages. Although digital steganography is still a young field, a variety of techniques have been developed to implement the hiding of information within digital images. This

4

research attempts to determine strategies that can be used automatically to decode a steganography file. Emphasis is placed on techniques that can be readily employed in an automated environment and methods that are simply specific to any individual steganography application.

1.4 Objectives

- To Analyze Steganography methods.
- To implement data hiding in digital image by using spatial method
- To implement data hiding in digital image by using frequency domain.

1.5 Thesis Organization

This thesis includes five chapters. The first chapter summarizes the research plan by stating the problem, objectives. Chapter II discusses the origin of Steganography and how it has historically evolved to its present-day computer applications. Chapter III describes the methodology used to meet the research objectives followed by Chapter IV's research analysis and results. Finally, Chapter V addresses the research recommendations and conclusions.

CHAPTER II

LITERATURE REVIEW

Steganography is an ancient art that has been reborn in recent years. The word Steganography comes from Greek roots which literally means 'covered writing', and is usually interpreted to mean hiding information in between other information. Steganography researcher, Markus Kuhn, has submitted the more modern definition of Steganography as the "art and science of communicating in a way which hides the existence of the communication" (Jordon 2000).

The goal is to hide, in plain sight, information inside other innocent information to disallow an outsider or enemy the opportunity to detect that there is a second secret message present. In recent years, there has been an exponential increase in the research community and industries focus towards information hiding techniques as opposed to the traditional cryptography area. Figure 2.1 expressively depicts this rapid increase in topic publications.

In the field of Steganography, some terminology has been developed. The adjectives cover, embedded and stego were defined at the "Information Hiding Workshop" held in Cambridge, England. The term ``cover" is used to describe the original, innocent message, data, audio, still, video and so on. When referring to audio signal Steganography, the cover signal is sometimes called the "host" signal.

The information hidden in the cover data is known as the ``embedded'' data. The ``stego'' data is the data containing both the cover signal and the ``embedded'' information. Logically, the process of putting the hidden or embedded data, into the cover data, is sometimes known as embedding. Occasionally, especially when referring to image Steganography, the cover image is known as the container.
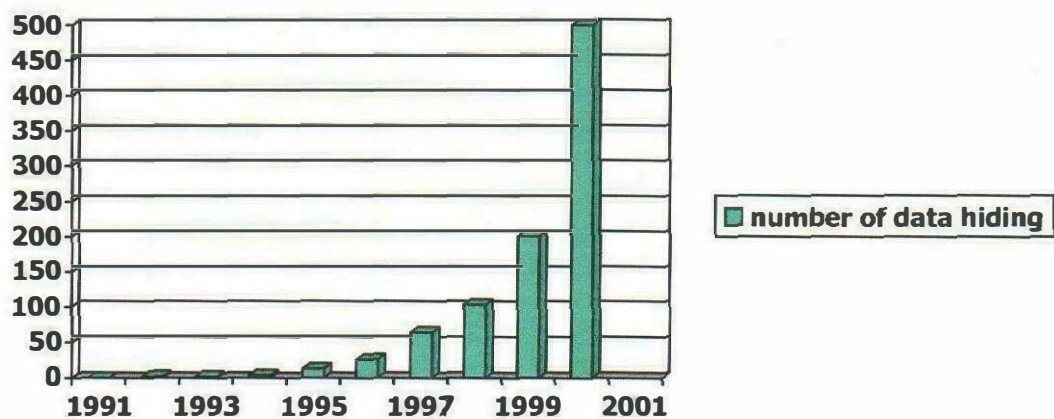
Figure 2.1 Number of Information Hiding Publications (Jordon 2000)

2.1 The Use of Steganography

One of the main drivers of the renewed interest in Steganography is for mitigating copyright abuses. As audio, video and other works become more readily available in digital forms, the ease with which perfect copies can be made may lead to large-scale unauthorized copying. This type of copying is naturally of great concern to the music, film, book, and software publishing industries. There has been significant recent research into digital watermarks or hidden copyright messages and digital fingerprints or hidden serial numbers. The idea is for file fingerprinting to be

7

used to help identify copyright offenders and then potentially prosecute them with the digital evidence.

2.3 History of Steganography

Histiaeus shaved the head of a slave and tattooed a message on his scalp. When the slave's hair had grown long enough he was dispatched to Miletus (Francesco 2002).

Another story from ancient Greece also comes to us via Herodotus. The writing medium of the time was text, written on wax-covered tablets. Demeratus, a Greek, needed to notify Sparta that Xerxes intended to invade Greece. To avoid capture, he scraped the wax off the tablets and wrote the message on the underlying wood. Then, he again covered the tablets with wax. The tablets appeared to be blank and unused so they passed inspection (Francesco 2002).

Invisible inks have always been a popular method of Steganography. Ancient Romans used to write between lines using invisible inks based on readily-available substances such as fruit juices, urine and milk. When heated, the invisible inks would darken, and become readable. Invisible inks were used as recently as World War II.

An early researcher in Steganography and Cryptography was Johannes Trithemius , a German monk. His first work on Steganography, Steganographia, described systems of magic and prophecy, but also contained a complex system of Cryptography. It was only published posthumously, as Trithemius had feared the reaction of the authorities if it was published (Johnson 1998).

Steganography continued to develop during the fifteenth and sixteenth centuries. Because they were often afraid of the wrath of powerful factions, authors of history books often concealed their names in their work. A treatise on this concept was written by Bishop John Wilkins, later the master of Trinity College, Cambridge. He devised a number of schemes ranging from coding messages in music and string knots to invisible inks, described the principles of cryptanalysis by letter frequencies, and argued against those who opposed publication in those fields.

As an interesting example of Steganography of this era, many scholars suspect the authorship of the Shakespearean plays can be attributed to Francis Bacon, the noted Elizabethan statesman and writer.

For example, a German spy sent the following message during the Second World War as in Figure 2.3 Apparently neutral's protest is thoroughly discounted and ignored. Blockade issue affects pretext for embargo on by decoding this message by taking the second letter in each word reveals the following secret message:

"Pershing sails from NY June 1".

Document layout was also used to reveal information. By modulating the position of lines and words, messages could be marked and identified.

Techniques such as writing messages on typewriter correction ribbons, and using pin punctures to mark selected letters were used. As new technologies that could pass more information and be even less conspicuous were developed, message detection improved. FBI Director J. Edgar Hoover as the enemy's masterpiece of espionage dubbed the German invention of the microdot.