Information Hiding in Multimedia Files: A Literature Review

Abdullah Mohammed Abdo Ali Ja'far (PhD student), Prof. Dr. Safaai Deris
Artificial Intelligence and Bioinformatics Laboratory
Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
Email: Abdullah_19762003@yahoo.com ; safaai@fsksm.utm.my

Abstract: There are two basic idea to keep something as a secret, the object could be changed by rules to a form could not be recognized the original object just for people who know these rules and every thing about it such that thy can obtain the original object, his is called "Cryptography" which is a Greek term and mean "Secret Writing". Or an object can be hidden at a secret place; one can hope that nobody will find this object except people who are familiar with the secret sender the methodology concerned with this kind of information security is called "Steganography". In this paper we speak about literature review of steganography.

Keywords: Information hiding, Steganography, AI, Security.

Introduction

There is no perfect protection to any type of information specially in these days because the modern programs and tools that solve any mathematical or statistical problems has been developing every day to make the difficult be easy. So that the protection in general has a relation with the time of protection, that means we protect the data for the time which is equal or little longer than the time that is spent in transmission of the secret data. One of the most important things that serve the data security is the development of communication methods and protocols, that is clear in the international network (Internet) which makes our home a small village and makes any user sends large amount of data in standard time not more than minutes. Therefore, the attacker has hard and little time to perform his attack [Abdullah, 2003].

Information hiding is one of the modern ways to keep something a secret and it depends on finding the best place to hide an object in away that is not to be recognized for others and in same time does not effect the place that keeps the object. That is called "Steganography". The word Steganography comes from Greek, which literally means "covered" or "hidden writing" and includes a vast array of methods of secret communications that conceal the very existence of the message [Abdullah, 2003; Neil et al., 2001].

Digital Steganography

Steganography 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 secret message present [Abdullah, 2003].

Steganography is the art and science of hiding the fact that communication is happening. While classical steganographic depend on keeping the encoding system secret, modern steganography is detectable only if secret information's known, e.g. a secret key. Because of their invasive nature steganographic systems leave detectable traces within an image's characteristics, e.g., its Fourier signature. This allows an eavesdropper to detect images that have been modified, revealing that secret communication is happening. Although, the secrecy of the information is not degraded, its hidden nature is revealed which defeats the whole purpose of steganography [Polpitiya et al., 2001; Provos, 2001].

Steganography is a technique to make confidential message imperceptible to human eyes by using some other data like an image. The data which hides the secret message is called "Vessel", "Carrier", "Container", or "Dummy data" of the secret message, it looks innocent, attackers can not see anything to attack. Therefore, steganography is more an "information imperceptualizing technique" than an "information hiding technique." Encryption of the embedded data would further improve security. This scenario is analogous to putting something in a very secure safe and then hiding the safe in hard to find place [Polpitiya et al., 2001; Kawaguchi, 2001].

In the following another definitions of Steganography:

- Steganography is the art and science of hiding data into innocent-Looking cover data so that no one can detect the very existence of the hidden data [Shin, 1999].
- Steganography is the art and science of communicating in a way which hides the existence of the communication [Mukherjee, 2002].
- Steganography in ideal word users would all be able to openly send encrypted mail or files to each other with no fear of reprisal [Johnson et al., 2001; Saleh, 2001].
- Steganography encompasses methods of transmitting secret messages through innocuous cover carriers in such a manner that the very existence of the embedded messages is undetectable [Johnson and Jajodia, 1998].
- In data security, the concealment of the existence of messages, literally covered writing. In data security this can take the form of filling in intermessage gaps with padding characters [L. et al., 1988].
- Steganography is the art of passing information in a manner that the very existence of the message is unknown [Johnson and Jajodia, 1998].
- Steganography is the hiding of information within a more obvious kind of communication. Although not widely used, digital steganography involves the hiding of data inside a sound or image file [Radcliff, 2002].
- Steganography today is a computer technique (or an art) to make confidential information imperceptible to human eyes by embedding it in some innocent looking vessel data (aka. a "carrier" or "dummy" data) such as a digital image or a speech sound [Kawaguchi et al., 2001].

The Main Goals of Steganography are:

1. To avoid drawing suspicion to the transmission of a hidden (secret) message. If suspicion is raised, then this goal is defeated [Johnson et al., 1998].
2. To hide messages inside other "harmless" messages in a way that does not allow any enemy to even detect that there is a second secret message present [Sellars, 2000].
3. The goal of steganography is undetectability, not secrecy only [Shin, 1999].

Types of Steganography Protocols: There are basically three types of steganographic protocols: pure steganography, secret key steganography, and public key steganography. In the following subsections, all three types will be described [Katzenbeisser and Petitcolas, 2000].

Steganographic Techniques

Steganography has been widely used in historical times, especially before cryptographic systems were developed. Examples of historical usage include:

- Hidden messages in wax tablets: in ancient Greece, people wrote messages on the wood, then covered it with wax so that it looked like an ordinary, unused, tablet.

- Hidden messages on messenger's body: also in ancient Greece. Herodotus tells the story of a message tattooed on a slave's shaved head, hidden by the growth of his hair, and exposed by shaving his head again. The message, if the story is true, carried a warning to Greece about Persian invasion plans.

- Hidden messages on paper written in secret inks under other messages or on the blank parts of other messages.

- During and after World War II, espionage agents used microdots to send information back and forth. Since the dots were typically extremely small -- the size of a period produced by a typewriter (perhaps in a font with 10 or 12 characters per inch) or even smaller -- the stegotext was whatever the dot was hidden within. If a letter or an address, it was some alphabetic characters. If under a postage stamp, it was the presence of the stamp.

- More obscurely, during World War II, a Japanese spy in New York City, Velvalee Dickinson, sent information to accommodation addresses in neutral South America. She was a dealer in dolls, and her letters discussed how many of this or that doll to ship. The stegotext in this case was the doll orders; the 'plaintext' being concealed was itself a codetext giving information about ship movements, etc. Her case became somewhat famous and she became known as the Doll Woman.

- In the manga Lone Wolf and Cub, a main plot device is a Yagyū letter which has a message written in mulberry extract. They

see the message by placing it in a bowl of silkworms and seeing where they eat.

- The one-time pad is a theoretically unbreakable cipher that produces ciphertexts indistinguishable from random texts: only those who have the private key can distinguish these ciphertexts from any other perfectly random texts. Thus, any perfectly random data can be used as a covertext for a theoretically unbreakable steganography. A modern example of OTP: in most cryptosystems, private symmetric session keys are supposed to be perfectly random (i.e. generated by a good RNG). Even very weak ones (e.g. shorter than 128 bits). This means that users of weak crypto (in countries where strong crypto is forbidden) can safely hide OTP messages in their session keys.

Modern Steganographic Techniques

- Concealing ciphertext within ciphertext. The method builds upon the security of the underlying cipher. If the cipher is secure, this steganographic method probably resists any statistical analysis. In contrast, this cannot be said about methods that conceal ciphertext within image or sound files. Statistical analysis of such data can detect patterns that are unusual or unexpected in digitized images or sounds. The steganographic method of concealing ciphertext within ciphertext is used e.g. by TrueCrypt, which is open source-disk encryption software for Windows and Linux.
- Chaffing and winnowing
- Invisible ink
- Null ciphers
- Concealed messages in tampered executable files, exploiting redundancy in the i386 instruction set [2]

Literature Review

A large literature on steganography had been composed by the 16-17th centuries. The basis of this steganography relied on novel information encoding methods. Gaspar Schott (1608-1666), in his book Schola Steganographica, explains how using musical notes to correspond to a specific letter could be used to hide messages in music scores. He also expanded Johannes Trithemius' (1462-1516) Ave Maria code proposed in Steganographæ, one of the first known steganography-related books. Similarly, David Kahn, in his book: The Codebreakers, explained how acrostic methods were used for a monk to hide his lover's name in the first letters of successive chapters of a book he Wrote [Cochran, 2000].

Steganography literature was grown after 1992.Most of the published articles concerned with the description of some software tools designed and build to perform steganography on some text, image and audio cover media. The publications about the scheme of the stego systems is very primitive, and mostly does not offer a key solutions for some weak aspects may face the discussed, (proposed) system. Among the large number of published articles, we have chosen the followings:

A scheme of steganography system to embed sound data in wave file and voice file without effect to the subjective quality of the sound data. And make a comparison between the 1's bit and 2's bits for least significant bit hiding to ensure the useful and it's condition and limitation. We use some audio files (MIDI) files, (VOC) files, (MP3) files, and WAV files as secret data that embedded in wave file and voice file and extracting from the carrier media without being changed. (ZIP, RAR) files also able use as secret data. The proposed system use cipher methods to encryption data before hiding process it in (wave and voice) files to provide better security of the hidden data [Abdullah, 2003].Implements a technique for data hiding in audio images, known as Audi file steganography. They are also being incorporating encryption of the data to be hidden [Polpitiya and Khan, 2001]. S-Tools v4 is an excellent Windows 95/NT-based steganography tool that hides files in BMP and GIF graphic files, and WAV audio files. S-Tools provided many user options including encryption and compression. Even though S-Tools uses least significant bit substitution and pseudo-random dispersion of hiding bits, the quality of the output is extraordinarily good when paralleled with comparable tools [Polpitiya and Khan, 2001]. a scheme of steganography system for hiding different types of data in audio media file (. WAV), (Windows Audio-Visual) and (. VOC), (Creative Voce File) format were proposed, the hiding mechanism was based on using Low-bit Encoding or Least Significant bit substitution (LSB) techniques [Raed, 2001]. Steghide is a DOS command-line application that features hiding data in BMP graphic and WAV and AU audio files. It features blowfish encryption, 128 bit MD5 hashing

of passphrases to blowfish keys, and pseudo-random distribution of hidden bits in the container file. Steghide is available in precompiled binaries for Windows and Linux platforms [Cochran, 2000].

Research in steganalysis is motivated by the concern that communications associated with illicit activity could he hidden in seemingly innocent electronic transactions. By developing defensive tools before steganographic communication grows, computer security professionals may he better prepared for the threat. This research investigated an artificial immune system approach to novel steganography detection for digital images [Jacob et al., 2003].

the problem of prime factorization is by no means the only computational problem that we know is hard to solve. Of course the term "hard" is subject to debate, and there will clearly not be much debate going on in the case of prime factorization, since we have a very deep understanding of problems in numeric computation, and can rely upon complexity-theory to estimate the hardness of problems. If we believe a problem to be NP-complete, then we can rely on its complexity-theoretic hardness, and analogously we might think of a problem to be "Al-complete", so we can rely on its ontologic hardness. AI-complete */A-I k*m-pleet'/* [MIT, Stanford: by analogy with 'NP-complete' *(see NP-)*\ adj. Used to describe problems or subproblems in AI, to indicate that the solution presupposes a solution to the 'strong AI problem' (that is, the synthesis of a human-level intelligence). A problem that is AI-complete is, in other words, just too hard. Examples of AI-complete problems are 'The Vision Problem' (building a system that can see as well as a human) and 'The Natural Language Problem' (building a system that can understand and speak a natural language as well as a human). These may appear to be modular, but all attempts so far (1999) to solve them have foundered on the amount of context information and 'intelligence' they seem to require. See also *gedanken.* (Raymond 2000) It is interesting to note that the term already appeared in the jargon, before the use of problems in artificial intelligence as security primitives was first proposed, as the above quote from *the Jargon File* shows. (The first appearence of this term in the context of security applications, to my knowledge, was an article about CAPTCHAs in c't (Schwellinger 2003)). However the word-game on "Al-completeness" and NP-completeness gives a suprisingly accurate account of the analogous roles of these notions in the analysis of secure communication systems [Richard, 2004].We introduce captcha, an automated test that humans can pass, but current computer programs can't pass: any program that has high success over a captcha can be used to solve an unsolved Artificial Intelligence (AI) problem. We provide several novel constructions of captchas. Since captchas have many applications in practical security, our approach introduces a new class of hard problems that can be exploited for security purposes. Much like research in cryptography has had a positive impact on algorithms for factoring and discrete log, we hope that the use of hard AI problems for security purposes allows us to advance the field of Artificial Intelligence. We introduce two families of AI problems that can be used to construct captchas and they show that solutions to such problems can be used for steganographic communication. captchas based on these AI problem families, then, imply a win-win situation: either the problems remain unsolved and there is a way to differentiate humans from computers, or the problems are solved and there is a way to communicate covertly on some channels [Luis et al.,2003].

Conclusion

Steganography 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 secret message present. The security of data is of extreme importance in today's information-based society, including the fields of military, diplomacy, corporation, medicine, and even the individual, the information have to be safeguarded to avoid the unauthorized or illegal accesses and prevent the misuses and abuses. Any system, method or technique that deal with, processing information (data), and put this data in shapes or forms of media under the condition that it must be not visible in its new form for human observer. All such systems are called hiding systems for information.

References

Abdullah M. A. A. Ja'far, (2003) "Audio Hiding In Audio Files by Using Low-Bit Encoding ", Informatics Institute for Postgraduate Studies, M.Sc thesis, Baghdad, Iraq.

Johnson N. F., Z. Duric, and S. Jajodia, (2001) "Information Hiding: steganography and Watermarking Attacks and Countermeasures", Kluwer a Cadmic Publishers, Boston, USA.

Katzenbeisser S., and F. A. P. Petitcolas, (2000) "Information Hiding Techniques for Steganography and digital watermarking", Artech House, INC., Norwood, London.

Luis von Ahn[1], Manuel Blum[1], Nicholas J. Hopper[1], and John Langford[2], (2003),"Using Hard AI Problems For Security", Computer Science Dept., Carnegie Mellon University, Pittsburgh PA 15213, USA [2] IBM T.J. Watson Research Center, Yorktown Heights NY 10598, USA.

Jacob T. Jackson, Gregg H. Gunsch, Roger L. Claypoole, Jr., and Gary B. Lamont, (2003)" Novel Steganography Detection Using an Artificial Immune System Approach ", Air Force Institute of Technology Department of Electrical and computer Engineering 2950 Hobson Way, Bldg 640 Wright Patterson Air Force Base, OH 45433. USA.

Richard Bergmair, (September 10, 2004) "Natural Language Steganography and an "Al-complete" Security Primitive ".

Cochran J. T., (March 2000) "Steganographic Computer Warfare", M.Sc. thesis of Science in Computer Systems, Air Force Institute Of Technology, Ohio, USA.
http://www.iwar.org.uk/iwar/resources/usaf/maxwell/students/2000/afit-gcs-eng-00m-03.htm
http://www.iwar.org.uk/iwar/resources/usaf/maxwell/students/2000/afit-gcs-eng-00m-03.pdf

Read M. Saleh, (2001), "Information Hiding In Wave Media File By Using Low Bit Encoding", M.Sc thesis, University of Technology, Baghdad, Iraq.

Provos N., (January 31, 2001) "Probabilistic Methods for Improving Information Hiding", Center for Information Technology Integration, University of Michigan, USA.
Email: provos@citi.umich.edu
http://citeseer.nj.nec.com/cache/papers/cs/19866/http:zSzzSzwww.citi.umich.eduzSztechreportszSzreportszSzciti-tr-01-1.pdf/provos01probabilistic.pdf

Kawaguchi E., (Oct. 14, 2001.). "Steganography in General", Knowledge Engineering Lab, Kyushu Institute of Technology, Internet Survey, Japan.
Email: kawaguch@kawa.comp.kyutech.ac.jp.
http://www.know.comp.kyutech.ac.jp/BPCSe/BPCSe-general.html

Johnson N. F. and S. Jajodia, " Steganalysis of Images Created Using Current Steganography Software", Information Hiding Second International Workshop, vol. 1525 of Lecture Notes in Computer Science, Portland, Oregon, U.S.A., Springer-Verlag, Berlin, Germany, and ISBN 3-540-65386-4.

Polpitiya D. and W. J. Khan, (2001) "Information Hiding in Audio Files with Encryption", Washington University, USA, Version 1.0.http://www.cbcis.wustl.edu/~adpol/courses/cs502/project/
http://netra.wustl.edu/~adpol/courses/cs502/project/report.pdf

L. D., S. M. (1988)"Data & Computer Security", Dictionary of standers and terms, Stockton press, New York, USA.

Radcliff D., (JUNE 10 2002.),"Steganography: Hidden Data", Quickstudy: steganography, Source: Computerworld.
http://www.computerworld.com/securitytopics/security/story/0,10801,71726,00.html

Kawaguchi E. , Ph.D. Professor, (Oct. 18, 2001)"A Large Capacity Steganography", Information Hiding Technique-Invitation to BPCS-Steganography, Knowledge Engineering Lab, Kyushu Institute of Technology, Japan.
E-mail: kawaguch@know.comp.kyutech.ac.jp
http://www.know.comp.kyutech.ac.jp/BPCSe/
http://www.know.comp.kyutech.ac.jp/BPCSe/BPCSe-general.html

Sellars D., (2000.)"An Introduction to Steganography", An Internet Survey.
http://www.cs.uct.ac.za/courses/CS400W/NIS/papers99/dsellars/stego.html

Shin N., (1999) ***"One-Time Hash Steganography"***, Information Hiding Third International Workshop, vol. 1768 of Lecture Notes in Computer Science, Dresden, Germany, Springer-Verlag, Berlin, Germany and ISBN 3-540-67182-X.