Volume 5, Issue 1, January 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Study of Digital Watermarking and Stegnography

Dr. Govind N Sarage
Department of Computer Science,
National Defence Academy, Khadakwasla, Pune,
Maharashtra, India

Abstract*— Steganography and Digital Watermarking are important tools that allows transmission of information over and over communications channel. The purpose of steganographic communication is to hide the mere existence of a secret message. Steganography refers to the technology of hiding data into digital media without drawing any suspicion, while Digital Watermarking is the asteganalysis is the art of embedding the information about the data. This paper provides a brief study on steganography and digital watermarking , mainly covering the fundamental concepts, applications and various techniques. Some commonly used techniques for improving steganographic security and enhancing digital watermarking capability are summarized and possible research trends are discussed.

***Keywords*— Stegnography, digital watermarking, security, hiding, classification.**

I. INTRODUCTION

Digital watermarking[1] is a relatively new research area that has attracted the interest of numerous researchers both in academia and industry and has become one of the hottest research topics in the multimedia signal processing community. Watermarking is the practice of imperceptibly altering a piece of data in order to embed information about the data. The above definition reveals two important characteristics of watermarking. First, information embedding should not cause perceptible changes to the host medium (sometimes called cover medium or cover data). Second, the message should be related to the host medium. In this sense, the watermarking techniques[2] form a subset of information hiding techniques, which also include cases where the hidden information is not related to the host medium (e.g., in covert communications). A watermarking system should consist of two distinct modules: a module that embeds the information in the host data and a module that detects if a given piece of data hosts a watermark and subsequently retrieves the conveyed information.

Information hiding is the science of concealing the *existence* of data even when it is being sought.

Steganography[11] is a sub-discipline of the broader science of information hiding and employs numerous technologies to achieve its goals: digital signal processing, cryptography, information theory, data compression, math, and human audio/visual perception, just to name a few. Steganography has two primary goals: 1) Security – is the hidden data perceptible by either a person or a computer, and 2) Capacity – how much data can be hidden in a given cover file. These two goals are often in competition. Steganography is the art of hiding information[12]. It includes techniques to hide an image, a text file, and even an executable program inside a “cover" image without distorting the cover image. The word steganography comes from the Greek and literally means “hidden writing." It includes a vast array of secret communications methods that conceal the message’s very existence. These methods include invisible inks, microdots, character arrangement, digital signatures, covert channels, and spread spectrum communications. Steganography and cryptography are cousins in the spy craft family. People have used steganography through the centuries to hide messages. The messages are hidden in plain sight, because they are visible to people who know where to look. Cryptography scrambles a message so it cannot be understood. Steganography hides the message so it cannot be seen. Consider the sentence “Where real interesting technical exchanges can overcome dull entertainment." The first letter of each word spells the message “write code." This is not hidden well. Better hiding methods use the second or third letter of each word or the first letter of the first word, second letter of the second word, etc. Steganography and cryptography are closely related. Cryptography scrambles a message to produce something that looks scrambled. The “write code" example could be scrambled to be “xsjuf dpef" (replace each letter with the letter that follows it in the alphabet). The scramble sometimes encourages prying eyes who see it as a challenge to unscramble. Steganography instead hides a message in a cover message. The result looks like something innocent, so prying eyes often dismiss it. Lawyers and libertarians debate if steganography is close enough to cryptography to regulate its use. To date, steganography remains unregulated.

II. AN APPLICATIONS OF WATERMARKING TECHNIQUES

Watermarking can be the enabling technology for a number of important applications. Obviously, each application imposes different requirements on the watermarking system.

A. Owner identification and proof of ownership.

This class of applications was the first to be considered in the watermarking literature. In this case, the embedded data can carry information about the legal owner or distributor or any rights holder of a digital item and be used for notifying/warning a user that the item is copyrighted, for tracking illegal copies of the item, or for possibly proving the ownership of the item in the case of a legal dispute.

B. Broadcast monitoring.

In this case, the embedded information is utilized for various functions that are related to digital media (audio, video) broadcasting. The embedded data can be used to verify whether the actual broadcasting of commercials took place as scheduled, i.e., whether proper airtime allocation occurred, for devising an automated royalty collection scheme for copyrighted material (songs, movies) that is aired by broadcasting operators[3], or to collect information about the number of people who watched/listened to a certain broadcast (audience metering). Broadcast monitoring is usually performed by automated monitoring stations and is one of the watermarking applications that has found its way toward successful commercialization.

C. Transaction tracking.

In this application, each copy of a digital item that is distributed as part of a transaction bears a different watermark. The aim of this watermark is not only to carry information about the legal owner/distributor of the digital item but also to mark the specific transaction copy. As a consequence, the embedded information can be used for the identification of entities that illegally distributed the digital item or did not adopt efficient security measures for preventing the item from being copied or distributed and for deterring such actions. Identification of movie theaters where illegal recording of a movie with a handheld camera took place is a scenario that belongs to this category of applications. The watermarks used in such cases are often termed fingerprints and the corresponding application fingerprinting. However, the same term is sometimes used for the class of techniques that try to extract a unique descriptor (fingerprint) for each digital item, which is invariant to content manipulation[4]. Obviously these techniques (which are sometimes called perceptual or robust hashing or replica detection techniques) are totally different fromwatermark-based fingerprinting, since they do not embed any data on the digital item, i.e., they are passive techniques.

D. Usage control.

In contrast to the applications mentioned above, where watermarking is used to deter intellectual rights infringement or to help in identifying such infringements, in usage control applications, the watermarking plays an active protection role by controlling the terms of use of the digital content. The embedded information can be used in conjunction with appropriate compliant devices to prohibit unauthorized recording of a digital item (copy control) or playback of unauthorized copies (playback control). The DVD copy and playback control using watermarking complemented by content scrambling is an example of this application [2, 4].

E. Authentication and tamper-proofing.

In this case, the role of the watermark is to verify the authenticity and integrity of a digital item for the benefit of either the owner/distributor or the user. Example applications include the authentication of surveillance videos in case their integrity is disputed[5], the authentication of critical documents (e.g., passports), and the authentication of news photos distributed by a news agency. In this context, the watermarking techniques can either signal an authentication violation even when the digital item is slightly altered or tolerate certain manipulations (e.g., valid mainstream lossy content compression) and declare an item as nonauthentic only when “significant” alterations have occurred (e.g., content editing). Certain watermarking methods used for authentication can provide tampered region localization, e.g., can detect the image regions that have been modified/edited.

F. Persistent item identification.

According to this concept, watermarking is used for associating an identifier with a digital item in a way that resists certain content alterations. This identifier can be used, in conjunction with appropriate databases, to convey various information about the digital item. Depending on the related information, persistent identification can be the vehicle for some of the applications presented above, e.g., owner identification, or usage control. Furthermore, the attached information can be used both for carrying copyright information and for enhancing the host data functionalities, e.g., by providing access to free services and products, thus, implicitly, discouraging the user fromremoving the watermark or illegally distributing the item and thus losing the added value provided by the watermark. Persistent association is dealt with in the MPEG-21 standard.

G. Enhancement of legacy systems.

Data embedded through watermarking can be used for the enhancement of information or functionalities carried/provided by legacy systems while ensuring backwards compatibility. For example, using techniques capable of generating watermarks that are robust to analog to digital and digital to analog conversion, one can embed in a digital image URLs that are related to the depicted objects. When such an image is printed (e.g., in a magazine) and then scanned by a reader, the embedded URL can be used for connecting her automatically to the corresponding webpage. Digital data embedding in conventional analog PAL/SECAM signals is another application in this category. In a more “futuristic” scenario, one can envision that information capable of enabling stereoscopic viewing to stereo-enabled receivers could be embedded through watermarking in conventional digital TV broadcasts. Using such an approach, conventional TV receivers would continue to receive the conventional signal with—hopefully—nonperceptible degradations.

III. CLASSIFICATION OF WATERMARKING ALGORITHMS

Various types of watermarking techniques each with their own distinct properties and characteristics can be found in the watermarking literature. In the following, we will review the basic categories of watermarking schemes[4, 7] and provide descriptions for the properties that distinguish each class from the rest. A first classification of watermarking schemes can be organized on the basis of their resistance to host medium modifications. Such modifications can either be the result of common signal processing operations (e.g., lossy compression) or be specifically devised and applied in order to render the watermark undetectable or affect the credibility and reliability of a watermarking system in other ways. Such modifications are usually referred to as *attacks*. Depending on the level of robustness offered, one can distinguish between the following categories of watermarking techniques:

A. Robust.

In this class, the watermarks are designed so as to resist host signal manipulations and are usually employed in IPR protection applications. Obviously, no watermarking scheme can resist all types of modifications, regardless of their severity. As a consequence, robustness refers to a subset of all possible manipulations[6] and up to a certain degree of host signal degradation.

B. Fragile.

In this case, the watermarks are designed to be vulnerable to all modifications, i.e., they become undetectable by even the slightest host data modification. Fragile watermarks are more easy to devise than robust ones and are usually applied in authentication scenarios.

C. Semifragile.

This class of watermarks provides selective robustness to a certain set of manipulations which are considered as legitimate and allowable, while being vulnerable (fragile) to others. Such watermarks can also be used in authentication cases instead of fragile ones. In practice, all robust watermarks are essentially semifragile, but in the former case, the selective robustness is not a requirement imposed by the system designer but rather something that cannot be avoided.

In order to achieve a sufficient level of security, watermark embedding and detection are usually controlled by a (usually secret) key *K*. In a way analogous to cryptographic systems, the watermarking schemes can be distinguished in two classes on the basis of whether the same key is used during embedding and detection:

D. Symmetric or privatekey.

In such schemes, both watermark embedding and detection are performed using the same key *K*.

E. Asymmetric or publickey.

In contrast to the previous class, these watermarks can be detected with a key that is different than the one that was used in the embedding stage. Actually, a pair of keys is used in this case: a private key to generate the watermark for embedding, and a public one for detection. For each private key, many public keys may be produced. Despite their advantages over their symmetric counterparts, asymmetric schemes[8] are much more difficult to devise. In terms of the information taken into account during embedding, the watermarking
methods can be broadly classified in two categories:

F. Blind embedding schemes.

Schemes belonging to this category consider the host data as noise or interference. Therefore, these techniques essentially treat watermarking like the classical communications problem of signal transmission over a noisy channel, the only difference being that, in the case of watermarking, restrictions on the amount of distortions imposed on the channel (i.e., the host medium) by the signal (the watermark) should be taken into consideration. In most cases, these methods rely implicitly or explicitly on a certain degree of knowledge of the host signal statistics, thus leading to the subclass of "known host statistics" methods. Essentially all methods developed in the first years of watermarking research belong to this category, most of them revolving around the spread spectrum principle where the watermark signal[9] consists of a pseudorandom sequence embedded, usually in an additive way, in the host signal.

IV. USES OF STEGANOGRAPHY

The use of steganography undeniably connotes dishonest activity, but there is a peaceful aspect to consider. Steganography is used in map making, where cartographers add a nonexistent street or lake in order to detect copyright offenders. Or similarly, fictional names are added to mailing lists to catch unauthorized resellers. Modern techniques use steganography as a watermark to inject encrypted copyright marks and serial numbers into electronic mediums such as books, audio, and video. For instance, DVD recorders detect copy protection on DVDs that contain embedded authorizations. Potential uses of steganography are undoubtedly vast[11]. Companies could advertise public Web pages containing private, hidden text that only internal members could intercept. An attempt to decipher the hidden text would be unwarranted since no encryption (or code) was used. Steganography could also be used to hide the existence of sensitive files on storage media. This would entail a cover folder and an embedded hidden folder. 1. Steganography can be a solution which makes it possible to send news and information without being censored and without the fear of the messages being intercepted and traced back to us. 2. It is also possible to simply use steganography to store information on a location. For example, several information sources like our private banking information, some military secrets, can be stored in a cover source. When we are required to unhide the secret information in our cover source, we can easily reveal our banking data and it will be impossible to prove the existence of the military secrets inside. 3. Steganography can also be used to implement watermarking. Although the concept of watermarking is not necessarily steganography, there are several steganographic techniques that are being used to store watermarks in data. The main difference is on intent, while the purpose of steganography is hiding information[12], watermarking is merely extending the cover source with extra information. Since people will not accept noticeable changes in images, audio or video files because of a watermark, steganographic methods can be used to hide this.

V. PROTECTION AGAINST STEGANOGRAPHY

The proliferation of networks has added intensity to both noble and ignoble purposes of steganography. Network security analysts face an insidious foe for sure. But how is steganography detected, and why should network security analysts be alarmed and cautious? Nearly all steganography programs in use leave behind traces or fingerprints that indicate something is not right. Based on research conducted over the years, organized crime[12, 15], terrorists, and various other groups operating worldwide commonly use steganography to operate via public forums, Web sites, etc. Software programs that detect steganography do exist, and enhanced iterations are under development. Neil Johnson, a graduate student at George Mason University, is developing a *stego detector[13]*. The program, he describes, is designed to search hard drives for electronic fingerprints that typically result from steganography applications. Similar to a virus scanner, this stego detector identifies signatures. As Johnson explains: Different authors have different ways to hide information to make it less perceptible. The author may come up with ideas that nobody else is using. That tool may have a special signature. Once that signature is detected, it can be tied to a tool. Johnson and other law enforcement agencies use software to locate signatures by studying the native structure of files including image, voice, text, and video files, and known software tools that implement steganography.

VII. SCIENTIFIC AND COMMERCIAL APPLICATIONS of STEGANOGRAPHY

A. Medical imagery.

Hospitals and clinical doctors can put together patient's exams, imagery, and their information. When a doctor analyzes a radiological exam, the patient's information is embedded in the image, reducing the possibility of wrong diagnosis and/or fraud. Medical-image steganography requires extreme care when embedding additional data within the medical images: the additional information must not affect the image quality[14].

B. Strong watermarks.

Creators of digital content are always devising techniques to describe the restrictions they place on their content. These technique can be as simple as the message ―Copyright 2007 by someone, as complex as the digital rights management system (DRM) devised by Apple Inc. in its iTunes store's contents, or the watermarks in the contents of the Vatican Library .

C. Military agencies.

Militaries' actions can be based on hidden and protected communications. Even with crypto-graphed content, the detection of a signal in a modern battlefield can lead to the rapid identification and attack of the involved parties in the communication. For this reason, military-grade equipment uses modulation and spread spectrum techniques in its communications[12, 14] .

D. Document tracking tools.

We can use hidden information to identify the legitimate owner of a document. If the document is leaked, or distributed to unauthorized parties, we can track it back to the rightful owner and perhaps discover which party has broken the license distribution agreement .

E. Document authentication.

Hidden information bundled into a document can contain a digital signature that certifies its authenticity .

F. Digital elections and electronic money.

Digital elections and electronic money[14, 15] are based on secret and anonymous communications techniques.

G. Radar systems.

Modern transit radar systems can integrate information collected in a radar base station, avoiding the need to send separate text and pictures to the receiver's base stations.

H. Remote sensing.

Remote sensing can put together vector maps and digital imagery of a site, further improving the analysis of cultivated areas, including urban and natural sites, among others.

VIII. CONCLUSIONS

Steganography works as a technique to hide information in plain sight. Steganography is an instrument of security, but not exclusively secure. Almost like magic, images, executable programs, and text messages can hide in images. The cover image does not appear altered. The approach steganography offers reduces the chance of a message being detected by its *inadvertent* layer of cover. However, if the hidden message is discovered, it is easily readable. For this reason, combining encryption algorithms with steganography offers a much stronger encryption routine. Although this article discusses some applications of steganography, there are many more uses in voice, media applications (such as communication channels), audio, and text, to name a few. This article unveils potential exploits of steganography regarding network security. Although awareness of steganography applications today is limited, progress is unfolding to expose the hidden art. Unfortunately, in the information age, the old adage ―what you don't know can't hurt you‖ is not always accurate.

Digital Watermarking defines methods and technologies that hide information, for example a number or text, in digital media, such as images, copyright protection, tamper proofing, video or audio. In this paper, we have surveyed many of the techniques proposed for watermarking. Thus this paper may serve as a ready reference for any new researcher willing to explore the basics and foundation works in the area of digital watermarking since its evolution to the point where it started gaining prominence in the area of digital media control. This paper gives a base to understand the recent advances in digital watermarking which may have happened after the previous works described in this paper but not covered in this paper.

REFERENCES

[1] R. Anderson and F. Petitcolas. On the limits of steganography. *IEEE Journal of Selected Areas in Communications*, 16:474–481, may 1998.

[2] Sara V. Hart, John Ashcroft, and Deborah J. Daniels. Forensic examination of digital evidence: a guide for law enforcement. Technical Report NCJ 199408, U.S. Department of Justice – Office of Justice Programs, Apr 2004.

[3] Sheridan Morris. The future of netcrime now (1) – threats and challenges. Technical Report 62/04, Home Office Crime and Policing Group, 2004.

[4] Niels Provos and Peter Honeyman. Hide and seek: an introduction to steganography.IEEE Security & Privacy Magazine, 1:32–44,May 2003.

[5] Andreas Pfitzmann. Information hiding terminology. In *Proceedings of the First Intl.Workshop on Information Hiding*, Cambridge, UK, May 1996. Springer–Verlag.

[6] Fabien A. P. Petitcolas, Ross J. Anderson, and Markus G. Kuhn. Information hiding—A survey. *Proceedings of the IEEE*, 87:1062–1078, Jul 1999.

[7] Bruce Norman. Secret warfare, the battle of Codes and Ciphers. Acropolis Books Inc., first edition, 1980. ISBN 0-87491-600-3.

[8] Marcus G. Kuhn. The history of steganography. In *Proceedings of the First Intl. Workshop on Information Hiding*, Cambridge, UK, May 1996. Springer–Verlag.

[9] Jian Liu, Xiangjian He; "A Review Study on Digital Watermarking", Information and Communication Technologies, 2005. ICICT 2005. First International Conference, Page(s):337 – 341, 27-28 Aug. 2005.

[10] Cox, I.J., M.L. Miller, and J.A. Bloom, "Digital Watermarking.", 1st edition 2001, San Francisco: Morgan Kaufmann Publisher.

[11] Johnson, N.F., Duric, Z. and Jajodia, S.G. *Information Hiding: Steganography and Watermarking - Attacks and Countermeasure*s.Norwell (MA): Kluwer Academic Publishers, 2001.

[12] Johnson, N.F., Duric, Z. and Jajodia, S.G. *Information Hiding:Steganography and Watermarking - Attacks and Countermeasure*s. Norwell (MA): Kluwer Academic Publishers, 2001.

[13] Anderson, R., Needham, R., and Shamir, A. "The Steganographic File System." In: Aucsmith, D. (ed.). *Proc. of the Second International Workshop on Information Hiding (IH '98*), Portland, OR, April 1998. *Lecture Notes in Computer Scienc*e, Vol. 1525. New York: Springer-Verlag, 1998.

[14] Artz, D. "Digital Steganography: Hiding data within Data." *IEEE Internet Computin*g, May/June 2001. URL: http: //www.cc.gatech.edu/classes/AY2003/cs6262_fall/digital_steganography.pdf. Last accessed: 2003.

[15] McDonald, A.D. and Kuhn, M.G. "StegFS: A Steganographic File System for inux." In: Pfitzmann, A. (ed.). *Proc. of the Third International Workshop on Information Hiding (IH '99*), Dresden, Germany, Sept.-Oct. 1999. *Lecture Notes in Computer Scienc*e, Vol. 1768. New York: Springer-Verlag, 2000. URL: http://www.cl.cam.ac.uk/~mgk25/ih99-stegfs.pdf. Last accessed: 2003.