Journal of Information Security and Applications 40 (2018) 217–235

Contents lists available at ScienceDirect

Journal of Information Security and Applications

journal homepage: www.elsevier.com/locate/jisa

A review of image steganalysis techniques for digital forensics

Konstantinos Karampidis [a,*], Ergina Kavallieratou [a], Giorgos Papadourakis [b]

[a] Department of Information & Communication Systems Engineering, University of the Aegean, 83200 Karlovasi, Samos, Greece
[b] Department of Informatics Engineering, Technological Educational Institute of Crete, 71410 Heraklion, Crete, Greece

ARTICLE INFO

Article history:

Keywords:
Image steganalysis
Steganography
Digital forensics
Universal steganalysis
Deep learning

ABSTRACT

Steganalysis and steganography are the two different sides of the same coin. Steganography tries to hide messages in plain sight while steganalysis tries to detect their existence or even more to retrieve the embedded data. Both steganography and steganalysis received a great deal of attention, especially from law enforcement. While cryptography in many countries is being outlawed or limited, cyber criminals or even terrorists are extensively using steganography to avoid being arrested with encrypted incriminating material in their possession. Therefore, understanding the ways that messages can be embedded in a digital medium –in most cases in digital images-, and knowledge of state of the art methods to detect hidden information, is essential in exposing criminal activity. Digital image steganography is growing in use and application. Many powerful and robust methods of steganography and steganalysis have been presented in the literature over the last few years. In this literature review, we will discuss and present various steganalysis techniques – from earlier ones to state of the art- used for detection of hidden data embedded in digital images using various steganography techniques.

1. Introduction

Steganography is the art of covered or hidden messaging. It is far different from Cryptography which is the art of making something inevitable to understand, unless the cryptography key is known. Steganography hides a message in a medium -which is in plain sight-, but no one understands hidden message's existence unless he is aware of it. It is an ancient technique and the etymology of the word comes from ancient Greek words: steganos (cover) + grapho (write). This technique of hiding messages is very common now days since cryptography in many countries is forbidden or limited by law [1]. In many cases, if a suspect maintains cryptographic content and refuses to reveal the cryptography key, the authorities automatically consider him as guilty.

Steganalysis is the opposite procedure of steganography. Primarily, we try to detect the existence of steganographic content in a digital device and secondly discover the hidden message. From this point of view, steganalysis can be classified into two major categories i.e. passive or active. Passive steganalysis tries to classify a cover medium as stego and identify the steganographic embedding algorithm, while active steganalysis additionally tries to estimate the embedded message length and ideally extract it from the cover medium.

Digital forensics is a relative new field in Computer Science and focuses on the acquisition, preservation and analysis of digital evidence. As Palmer said, digital forensics are "the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations."[2].

Both steganography and cryptography intend to hide information and often both are used together. Though cryptographic messages are easily detectable while they are meaningless, steganography messages appear to be normal at first sight. Based on knowledge of the actual message, the availability of the original cover file and the steganography tool, the following types of technical steganalysis can be distinguished [3]:

- Stego only attack - only the stego object is available for analysis.
- Known cover attack - the cover and the stego object are both available for analysis.
- Known message attack - the message is known and can be compared with the stego object.
- Chosen stego attack - the stego object and the stego tool (algorithm) are available for analysis.

* Corresponding author.
E-mail addresses: karampidis@aegean.gr (K. Karampidis), kavallieratou@aegean.gr (E. Kavallieratou), papadour@cs.teicrete.gr (G. Papadourakis).

https://doi.org/10.1016/j.jisa.2018.04.005
2214-2126/

- Chosen message attack - the steganalyst generates stego-media from some steganography tool or algorithm from a known message. The goal in this attack is to determine corresponding patterns in the stego-media that may point to the use of specific steganography tools or algorithms.
- Known stego attack - the steganography tool (algorithm) is known and both the original and stego-object are available.

Cover medium can be an image file, an audio file, a video file, a network packet or even a text file. It is obvious that as more elements are known to a digital forensics examiner, the more effective steganalysis will be. Furthermore, steganalysis becomes more complex when moving from detection only, to detecting and deciphering the embedded message i.e. moving from passive to active steganalysis. As steganography becomes more widely available and data either on digital devices or internet increases, the detection of steganographic content by digital forensics examiners becomes highly important. Theoretically, this concerns any type of digital objects, but practically -in most cases- audiovisual files are more frequently met. This literature review will deal with image steganography and analyze state of the art methods of steganalysis.

More than one hundred methods extended to any type of image steganalysis were recorded and presented. Two major approaches were adopted by scientists. The first one refers to extraction of statistical features from stego and clean images. These statistical features are compared then, in order to discriminate clean from stego images.

The second general approach is by employing machine learning techniques. Thus, features are extracted from images (both clean and stego), a classifier is trained, and finally unseen images are presented to the model for evaluation. Typical paradigms of the utilized classifiers are mostly Support Vector Machines (SVM) and artificial neural networks. In both approaches an interesting subject discussed widely in each paper – and a critical step for achieving best results- is feature extraction and selection. Many techniques were used for this, such as statistics (mean, kurtosis, skewness, histogram analysis etc.), covariance matrix, similarity measures between pixels etc. Apart from the two prementioned approaches, modern methods employ deep learning techniques such as convolutional neural networks or deep autoencoders, where feature extraction and selection is made in an almost automatic way.

The performance and the quantitative analysis of the techniques discussed in the following sections has also been given, by using metrics such as the detection rate, the error rate and ROC curves in specific embedding rates. In appendix we also provide tables (Table 1 to Table 5) for each steganalysis category. These tables besides basic information (i.e. author, date, method in brief) also indicate the evaluation metric, dataset and number of images used, in order to make the comparison between methods from the same steganalysis category more distinct.

In [4,5], authors propose a different taxonomy of steganalysis i.e. specific and statistical while in [6] and [7]. Paper in [8] only refers to steganalysis methods applicable to jpeg images, while in [9] authors only refer to methods for universal (blind) detection for image steganography. Our review provides a detailed reference from earlier steganalysis methods to state of the art, refers to all steganalysis categories and not only to specific ones (such as jpeg, universal etc.) and it is up to date including current trends like deep learning techniques.

All presented papers were retrieved by Google Scholar. Primarily, the search term “image AND steganalysis” was given and 5590 results were retrieved. In order to reduce the number of given papers, we searched again under the following constraints: a) the search term should be part of paper’s title, so that the presented papers by Google Scholar should be more relevant to our subject b) only papers from year 2000 since today are acceptable c) patents were not included d) books were not included. Search queries used were: i) allintitle: "steganalysis", which resulted to 2080 results ii) allintitle: "image steganalysis", which resulted to 344 results iii) allintitle: "lsb", which resulted to 2150 results iv) allintitle: "lsb matching" which resulted to 159 results v) allintitle: "universal steganalysis" which resulted to 75 results. Combining all the above search results from the different given search terms and by reading the abstract of each paper to determine if the paper was relevant to our research, we ended up having more than 100 papers which are presented in the following sections.

The rest of the paper is organized as follows. In Section 2 the taxonomy of steganalytic techniques is presented. Section 3 examines visual steganalysis, while Section 4 presents signature steganalysis techniques. Statistical steganalytic techniques are discussed in Section 5. Spread Spectrum Steganalysis is discussed in Section 6, while in Section 7 the Transform Domain Steganalysis techniques are presented. Finally in Section 8, Universal (blind) Steganalysis methods are examined, and Section 9 presents the conclusions derived from this review.

2. Taxonomy of steganalysis techniques

The simplest method of steganography is by embedding a message after the end of file (EOF) or by embedding hidden information into exif header. Both methods are simple and fast, but they are vulnerable to steganalysts. Even by looking the file with a hex editor, the message -if unencrypted- can be revealed. This simple technique is effective for people with little or none steganalysis knowledge, but it is very easy for digital forensic examiners to detect and retrieve the hidden information from the cover medium. Consequently, new steganography techniques were developed and new steganalytic approaches were proposed. Depending on the attack method a forensic examiner uses, six major categories are introduced:

- visual steganalysis
- signature or specific steganalysis
- statistical steganalysis
- spread spectrum steganalysis
- transform domain steganalysis
- universal or blind steganalysis

3. Visual steganalysis

Visual attacks are the simplest form of steganalysis. A visual attack is the examination of the suspicious image with the naked eye to identify any noticeable discrepancies. This turns to be very difficult, since the alterations made to an image when a message is embedded, do not result in quality degradation. Most steganographic algorithms create stego objects that are similar to their cover medium. However, when unaltered parts of a stego image are removed, it is possible to observe signs of manipulation. Hence, if a steganalyst can identify those features of the image that characterize it as stego, a visual attack may reveal the existence of a hidden message. The most common form of a visual attack concerns Least Significant Bit (LSB) steganography. The image is converted to its binary form and the bits in the LSB plane are retrieved. In an image usually, there are as many even values as there are odd, typically saying that there are approximately as many 1’s as there are 0’s in its LSB plane. When text is converted to binary however, there are often more 0’s than 1’s. This indicates a visual inconsistency and helps the forensic examiner to classify the image as stego. However, this steganalytic technique is successful only when a poor steganographic algorithm was used to produce the stego image. Typical software paradigms following that embedding