Pattern Recognition 43 (2010) 405–415

Contents lists available at ScienceDirect

Pattern Recognition

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

Steganalysis and payload estimation of embedding in pixel differences using neural networks

Vajiheh Sabeti[a], Shadrokh Samavi[a,b,*], Mojtaba Mahdavi[a], Shahram Shirani[b]

[a]Department of Electrical and Computer Engineering, Isfahan University of Technology, Iran
[b]Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada

ARTICLE INFO

Article history:
Received 13 August 2008
Received in revised form 1 May 2009
Accepted 11 June 2009

Keywords:
Steganography
Steganalysis
Pixel-value differencing
Artificial neural network (ANN)
Perceptrons

ABSTRACT

In this paper a steganalysis technique is proposed for pixel-value differencing method. This steganographic method, which is immune against conventional attacks, performs the embedding in the difference of the values of pixel pairs. Therefore, the histogram of the differences of an embedded image is different as compared with a cover image. A number of characteristics are identified in the difference histogram that show meaningful alterations when an image is embedded. Five distinct multilayer perceptrons neural networks are trained to detect different levels of embedding. Every image is fed to all networks and a voting system categorizes the image as stego or cover. The implementation results indicate 88.6% success in correct categorization of the test images that contained more than 20% embedding. Furthermore, using a neural network an estimator is presented which gives an estimate of the amount of the MPVD embedding in an image. Implementation of the estimator showed an average accuracy of 88.3% in the estimation of the amount of embedding.

1. Introduction

Steganography is an art of sending a secret message under the camouflage of a carrier content. The carrier content appears to have normal ("innocent") meanings. The goal of steganography is to mask the very presence of communication, making the true message not discernible to the observer [1]. Steganography is different from classical encryption, which conceals the contents of secret messages. Steganography is about hiding the very existence of the secret messages [2].

Three different aspects in information hiding systems contend with each other: capacity, security, and robustness. Steganographic methods strive for high security and capacity and usually robustness is not of main concern [3]. While different communication mediums, such as text, audio and video, can be used for steganography, embedding data in images have been the subject of many studies. In Fig. 1, a block diagram of a secret communication link through steganography is illustrated. The purpose of this system is to send out a secret message. The message is a string of bits. We emphasize that the compression and the encryption steps are required before the steganography can be implemented. The compression process eliminates the redundancy in the contents of the message and hence it enhances the utilization of the capacity of the steganographic algorithm. The compressed message is encrypted to further increase the security of the link [4].

The carrier image in steganography is called the "cover image" and the image which has the embedded data is called the "stego image". The embedding process is usually controlled using a secret key shared between the communicating parties. This key ensures that only recipients who know the corresponding key will be able to extract the message from a stego image.

With an increase in popularity of steganographic methods a new field of steganalysis is established which is amid at distinguishing the presence of steganography. A successful attack is one which detects the presence of hidden data in a stego image. Most steganalytic methods use changes in the statistical characteristics of an image to detect the presence of embedded data. Steganalitic methods, in terms of application, can be categorized into two general groups of special purpose and blind. Special steganalysers aim to identify the presence of secret message embedded by a specific algorithm, while blind steganalysers are intended to detect a wide range of steganographic algorithms including previously unknown ones [5].

There are two kinds of image steganography techniques: spatial-domain- and transform domain-based methods. Spatial-domain-based methods embed messages directly in the intensity of pixels of images [6,7]. For transform domain-based techniques, images are

* Corresponding author at: Department of Electrical and Computer Engineering, Isfahan University of Technology, Iran.
E-mail address: samavi@mcmaster.ca (S. Samavi).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.06.006

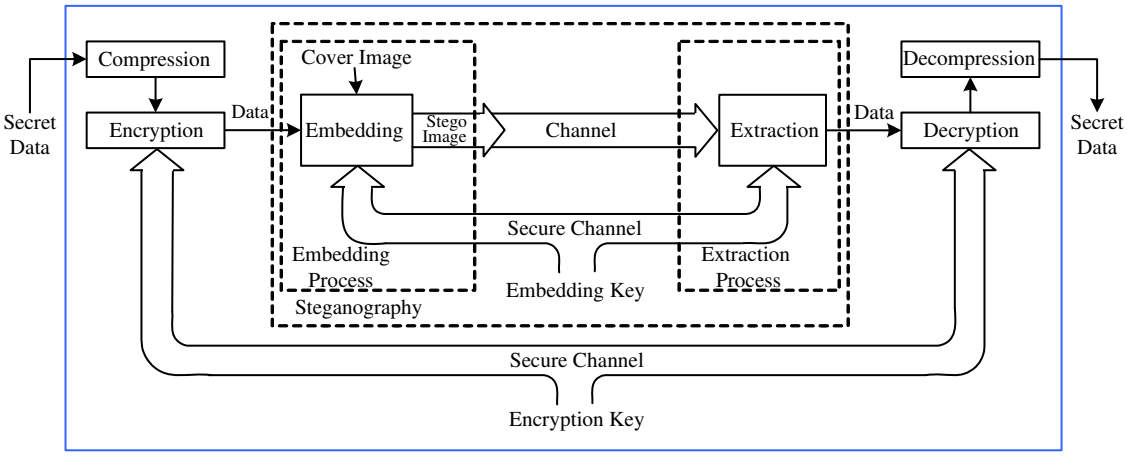

Fig. 1. Block diagram of a secret data link which uses steganography.

first transformed to another domain (such as the frequency domain), and messages are then embedded in transform coefficients [8,9].

A popular digital steganography technique is the so-called least significant bit (LSB) replacement. With the LSB replacement technique, the two parties in communication share a private secret key that creates a random sequence of samples of a digital signal. The encrypted secret message is embedded in the LSBs of those samples of the sequence. This digital steganography technique takes the advantage of random noise present in the acquired images [1]. Many reliable steganalytic methods have been devised for LSB flipping technique. The production of Pair of Value (PoV) in the histogram of a stego image is the main weak point of the LSB flipping. The presence of PoV has allowed many steganalysis methods, such as $\chi^2$and RS, to successfully attack LSB flipping [10,11].

Wu and Tsai [12] presented a steganographic method based on pixel-value differencing (PVD). They divide the cover image into a number of non-overlapping two-pixel blocks. Each block is categorized according to the difference of the gray values of the two pixels in the block. A small difference value indicates that the block is in a smooth area and a large one indicates that it is in an edged area. The pixels in edged areas may tolerate larger changes of pixel values than those in the smooth areas. Therefore, it is possible to embed more data in edged areas than in the smooth areas. All possible difference values are classified into a number of intervals. The number of bits that are to be embedded in a pixel pair depends on the width of the interval that the difference value belongs to. PVD is immune against the attacks that scrutinize changes in spatial domain [11] or changes in the histogram [10].

In [13] another method based on PVD is proposed which tries to increase the embedding capacity of PVD. They use LSB embedding for smooth regions and PVD embedding for edged areas. We refer to this method as the enhanced PVD as opposed to the basic PVD of [10]. Sabeti et al. [14] successfully attacked both the basic and the enhanced version of PVD. In that attack a *substitute image* is formed from the original image. The formation of this substitute image is different when attacking the basic PVD as compared to when the enhanced PVD is being attacked. When attacking the basic PVD, the image is traversed in a zigzag manner and the 2-pixel differences are taken as the pixel values of the new substitute image. In case of the enhanced PVD steganalysis we are faced with two types of blocks. For the blocks that have been embedded with LSB flipping the values of the block pixels are directly used in the substitute image. But if the block has been embedded by PVD embedding then the pixel difference is used to form one pixel in the substitute image. If the substitute image is constructed from a stego-image we can notice the POV phenomenon in the histogram of the substitute image. The presence of the POV phenomenon can be revealed with the application of the χ^2 attack. Hence, the embedding by the basic PVD or its enhanced version can be detected.

In a paper by Zhang [15] PVD was successfully attacked. This was done by analysis of the histogram of the stego-image. Zhang has also proposed a modified version of PVD where the shortcomings of the basic PVD are alleviated. We refer to this work as the modified PVD (MPVD). So far, no attack has been devised against MPVD. Also, recently Lu et al. [16] have proposed a steganographic method which is similar to MPVD but performs the embedding in the difference of nibbles. Lin et al. [17] also use pixel differencing in groups of three pixels to embed data.

Generally, blind steganalytic methods use classifiers. These classifiers are trained using a set of cover and stego images obtained from a number of steganographic methods. Classification is based on alterations in a number of characteristics of natural images due to the embedding process. Blind steganalysis based on high order statistics [18] and blind steganalysis for JPEG images [19] are two such methods. The goal of these methods is to discover a number of embedding algorithms. There are also methods that use classifiers and are devised to attack a specific steganographic method [20].

In this paper we propose a steganalysis technique which attacks and successfully identifies the existence of MPVD embedding. Actually we are offering a strategy to reveal data embedding in pixel differences. In this attack, we have used an artificial neural network (ANN) for classification purposes and used a second ANN to estimate the amount of embedding. The organization of the paper is as follows. In Section 2 the working of PVD and MPVD are reviewed. Statistical changes that occur in an image after MPVD embedding are analyzed in Section 3. The proposed steganalysis technique is detailed in Section 4. The implementation results of the proposed method are presented in Section 5. Conclusions appear in Section 6.

2. Pixel-value differencing steganography

The cover images used in the PVD method are supposed to be 256 gray-valued ones. A difference value d is computed from every non-overlapping block of two consecutive pixels, say p_i and p_{i+1} of a given cover image. Partitioning the cover image into two-pixel blocks runs through all the rows of each image in a zigzag manner. In Fig. 2, an example of an image is shown. The two-pixel blocks that are constructed by zigzag scanning of the example image is shown in Fig. 3.