International Journal of Digital Crime and Forensics
Volume 11 • Issue 1 • January-March 2019

Detecting and Distinguishing Adaptive and Non-Adaptive Steganography by Image Segmentation

Jie Zhu, SKLOIS, Institute of Information Engineering, Chinese Academy of Sciences (CAS), Beijing, China
Xianfeng Zhao, SKLOIS, Institute of Information Engineering, Chinese Academy of Sciences (CAS), Beijing, China
Qingxiao Guan, SKLOIS, Institute of Information Engineering, Chinese Academy of Sciences (CAS), Beijing, China

ABSTRACT

This article describes how blind steganalysis aiming at uncovering the existence of hidden data in digital images remains an open problem. Conventional spatial image steganographic algorithms hide data into pixels spreading evenly in the entire cover image, while the content-adaptive algorithms prefer the textural areas and edge regions. In this article, the impact of image content on blind steganalysis is discussed and a practical and extensible approach to distinguish the different types of steganography and construct blind steganalytic detector is proposed. Through the technique of image segmentation, the images are segmented into sub-images with different levels of texture. The classifier only cares for the sub-images which can help modeling the statistical detectability and is trained on sub-images instead of the entire image. Experimental results show the authors' scheme can recognize the type of steganographic methods reliably. The further steps to improve capacity of blind steganalysis based on image segmentation are also mentioned and achieve better performance than ordinary blind steganalysis.

KEYWORDS

Blind Steganalysis, Ensemble Classifier, Image Segmentation, Information Security, Steganography, Texture Complexity

1. INTRODUCTION

Steganography and staganalysis have become important topics in the field of information security, and acquired a great deal of attention by researchers all over the world. While steganography embeds the secret to the media file which aims to hide the fact of covert communication, staganalysis tries to find out the embedding modification by the adversary. Digital image files are the usual carriers of secret message due to the simple approach of acquirement and convenient delivery and operation.

The Least significant bit (LSB) is the main embedding channel for digital image files. Two classical but typical embedding ways are LSB replacement and LSB matching which is also called ± 1 embedding (Sharp, 2001). LSB methods are commonly used among the many free steganography tools

DOI: 10.4018/IJDCF.2019010105

available on the internet. LSB replacement simply replaces the least significant bits of pixel selected to be modified by the message bits after encryption using a secret key (Ker, 2005). This results in the detection of LSB replacement is easy and achieves reliable performance. To improve the ability for resisting attacks, LSB matching is proposed which add or subtract one randomly from the original pixel value when the message bits differ from the least significant bits of pixels. It turns out that LSB matching is harder to attack by some useful detectors in steganalysis of LSB replacement (Ker, 2005). In recent years, adaptive steganography draws lots of attention since it affords high security. The framework of adaptive steganography is utilizing a coding scheme to minimize a well-designed distortion function which models the statistical changes caused by embedding (Vojtěch Holub & Fridrich, 2013). Adaptive Steganography tends to heuristically embed secret data into the complex areas of an image to avoid causing perceptual artifacts and the changes of statistical properties of images content. This also explains why it can outperform the non-adaptive Steganography.

On the other hand, the trend of modern steganalysis is to extract the statistical features and combine with ensemble classifier to construct binary classifier to distinguish cover images and stego images embedded with secret message. Zhao et al. (Zhao, Zhu, & Yu, 2016) summarized systematic and comprehensive definitions of all paradigms of steganalysis, covering not only laboratory research but also the real-world application. In targeted steganalysis it has different principles to detect different types of embedding strategies. For example, the structural analysis is effective to attack the LSB embedding, and features based on selection channel knowledge are applied to analysis up-to-date content-adaptive steganography algorithms (T. D. Denemark, Boroumand, & Fridrich, 2016). Meanwhile, blind steganalysis play a more important role in universal steganalysis. It requires the attacker to design universal features which are suitable for a wide range of steganography algorithm under circumstances of lack of knowledge about the embedding strategy and the embedding payload. As a result, the goal of the statistical features is to uncover the local and global measures sensitive to the pixels or coefficients modification. Although, cover source mismatch is another open problem since the performance of detector degrades dramatically when the distribution of testing images set does not match that of training set (Zhu, Guan, Zhao, Cao, & Chen, 2017).

Considering that features have a strong impact on the performance of blind detector, the approaches to design the features have been studied from different perspectives. The rich model is a combinatorial feature, which contains many submodels. The submodels are designed to capture a large number of different types of dependencies among neighboring pixels and ultimately assembled together to detect different kinds of embedding algorithms. However, the common characteristic of the feature above is treating the entire images as a whole and rarely taking the content block of diverse complexity into account. We argue that the statistical changes due to embedding are related to not only the embedding algorithm itself, but also the different blocks of cover image which exhibit different texture feature. This can be indicated by the fact that adaptive steganography assigns a higher cost value to those pixels in textural areas and along edges than in smooth areas, consequently the noise areas contribute more pixel modification.

Some literature has focused on steganalysis based on image content, which constructs several classifiers based on different level of complexity of image content and detect the image by fusing the prediction of trained classifiers. Amirkhani & Rahmati (2011) proposed a new framework which can accommodate all blind image steganalysis methods. Firstly, in the training phase, the input training images set are divided into classes according to an image content evaluation criterion and then the training of classifiers is specialized for each class. In the testing phase, a fuzzy approach is used to include the decision of different classes. Note that the framework classifies each of images in training set database into disjoint subclass. Cho et al. (Cho, Cha, Gawecki, & Kuo, 2013) decompose single image to smaller blocks of fixed size, then classify image blocks into multiple classes according to their feature vector and find a classifier for each class. Detection result of the whole image can be obtained by integrating results of all image blocks via decision fusion. Xiong et al. (Xiong, Ping, Zhang, & Hou, 2012) proposed to decompose images into several "textural detail subbands" by the