Lightweight Steganalysis Based on Image Reconstruction and Lead Digit Distribution Analysis

Alexandros Zaharis, University of Thessaly, Greece

Adamantini Martini, SIEMENS SA, Greece

Theo Tryfonas, University of Bristol, UK

Christos Ilioudis, ATEI of Thessaloniki, Greece

G. Pangalos, Aristotle University of Thessaloniki, Greece

ABSTRACT

This paper presents a novel method of JPEG image Steganalysis, driven by the need for a quick and accurate identification of stego-carriers from a collection of files, where there is no knowledge of the steganography algorithm used, nor previous database of suspect carrier files created. The suspicious image is analyzed in order to identify the encoding algorithm while various meta-data is retrieved. An image file is then reconstructed in order to be used as a measure of comparison. A generalization of the basic principles of Benford's Law distribution is applied on both the suspicious and the reconstructed image file in order to decide whether the target is a stego-carrier. The authors demonstrate the effectiveness of the technique with a steganalytic tool that can blindly detect the use of JPHide/JPseek/JPHSWin, Camouflage and Invisible Secrets. Experimental results show that the steganalysis scheme is able to efficiently detect the use of different steganography algorithms without the use of a time consuming training step, even if the embedding data rate is very low. The accuracy of the detector is independent of the payload. The method described can be generalized in order to be used for the detection of different type images which act as stego-carriers.

Keywords: *Benford's, Data Hiding, JPEG, Staganalysis, Steganography*

INTRODUCTION

Hidden data retrieval has always been a major part of Computer Forensics. Many cases have been solved after analyzing files that seemed of no interest for a case but had important evidence hidden in them. Data hiding in an information system can be performed for various reasons including potential malware attacks, hiding data for later use in a compromised environment by an attacker or exchanging secret information via the Internet. Steganography has always been a popular method of exchanging information in plain sight especially through the internet. Its

DOI: 10.4018/jdcf.2011100103

popularity grew along with new techniques of hiding information in different carrier files with image files being the most popular amongst them. With the rapid growth of steganography on image files came the great need for Forensic investigators to analyze large volumes of images in order to detect possible hidden evidence. Different tools have been developed to computerize the process of locating suspect carrier files of different file types using visual, protocol compatibility or statistic analysis attacks. Most of these techniques concentrate and actually work against specific steganography algorithms/tools and are usually time consuming. In order to speed up the process of Steganalysis without sacrificing high detection rates, we are going to present a universal technique of detecting image steganography carrier files. Our method concentrates on reconstructing (Nosratinia, 2001) an 'original' image in order to use it as a comparison measure against the original possibly stego-carrier file. Our work concentrates on:

1. Benford's Law, and the reasons why choosing this kind of metric as a detection schema.
2. The presentation of the process of creating a reconstructed image, resembling the data structure of the original image file before embedding any hidden data in it.
3. The design and usage of a custom, lightweight forensic tool utilizing the above mentioned technique to blindly detect image carrier files.
4. Hit ratio results along with time analysis of the detection process compared with other image steganalysis tools.

The contribution of this paper to the forensics community concentrates on the presentation of a lightweight steganalytic technique/ tool that minimizes computation time by implementing a well known statistical analysis method (Benford, 1938). This tool can be extended in order to be applicable to other image file types while complying with the known computer forensic standards.

LITERATURE REVIEW

There are numerous techniques that can be used in order to hide data from potential interception but we are going to focus on steganography (Anderson et al., 1998; Kessler, 2004) on image files. This technique has been well described, and is well known to forensic investigators. Different tools have been developed to computerize the process of locating suspect carrier files of different file types using visual, protocol compatibility or statistic analysis attacks (Fridrich & Goljan, 2002). Most of these techniques concentrate and actually work against specific steganography algorithms/tools. While others that are used for universal blind steganalysis need a training step for agents to be more efficient in locating statistic anomalies on carrier files (Barbier et al., 2007). These techniques are of great performance when the training step includes a large number of true positive carrier files to be examined but can be very time consuming. On the other hand the above techniques mentioned have low hit rate for no training step. In order to speed up the process of steganalysis without sacrificing high detection rates, we are going to present a less common technique of detecting image steganography carrier files. Our technique is going to focus on speed detection, based only on image reconstruction and simple comparing of file structure in order to detect possible steganography leading to the creation of a tool that can be used by Forensics practitioners.

STEGANOGRAPHY CONCEPTS AND TOOLS

In our work we are going to distinguish four image file types:

1. The **original file**, which in our case would be a JPEG image file created/saved with MS Paint.
2. The **carrier file**, which in our case is going to be the result of steganography applied