International Journal of Digital Crime and Forensics
Volume 8 • Issue 2 • April-June 2016

On More Paradigms of Steganalysis

Xianfeng Zhao, State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

Jie Zhu, State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

Haibo Yu, State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

ABSTRACT

Up to now, most researches on steganalysis concentrate on one extreme case. Typically, the a priori knowledge of the embedding way and cover-media is assumed known in the classifier training and even feature design stage. However, the steganalysis in the real world is done with different levels of such knowledge so that there can be various paradigms for doing it. Although some researchers have addressed the situations, there is still a lack of a systematic approach to defining the various paradigms. In this paper, the authors give such an approach by first defining four extreme paradigms, and then defining the rest among them. Each paradigm is related with two sets of assumed known a priori knowledge respectively about the steganographic algorithm and cover-media, and each paradigm corresponds to a particular case of steganalysis. Also we will see that different paradigms can have very different aims so that the designs may be various.

KEYWORDS

Information Security, Paradigm, Steganalysis, Steganalytic Paradigm, Steganography

1. INTRODUCTION

Steganalysis, intuitively, is for recognizing stego-media. So it is not strange that the methods from pattern recognition are widely used in this area. In the most studied model of steganalysis (Chen & Shi, 2014; Pevný, Bas, & Fridrich, 2010; Davidson & Jalan, 2010; Pevný & Fridrich, 2007; Shi, Chen, & Chen, 2007; Fridrich & Kodovský, 2012; Kodovský & Fridrich, 2012; Kodovský, Fridrich, & Holub, 2012), steganalytic features are first generated and selected. Then, their values, which are computed over the samples containing both covers and stego-media, are used to train a classifier which will be the final tool to recognize stego-media. Notably, in such a dominant model, supervised learning is adopted. The stego-media used in the training are produced by the targeted steganography, of which the embedding method except for the secret key is assumed known to the steganalyst. Moreover, the media used in the training and testing often have a set of uniform parameters. For instance, they come from one database with images of the same size and quality factor (QF).

Many researchers have noticed that the a priori knowledge of the embedding way, primarily including the embedding algorithm and embedding rate, plays an important role in steganalysis. If the knowledge is available in designing features which are applicable to only the targeted steganography, the steganalysis is called a specific scheme; otherwise, if the features are suitable for detecting more steganography, it is regarded as universal (Shi, Chen, & Chen, 2007) or blind (Davidson & Jalan,

DOI: 10.4018/IJDCF.2016040101

2010). However, such universal schemes finally become specific because the training samples, in supervised learning, are produced by the targeted steganography. Apparently, it is impractical to assume such a priori knowledge be available to a steganalyst in the real world. Possibly for this reason, universal steganalysis has a few other constructions. Ker et al. (Ker, Bas, Böhme, Cogranne, Craver, Filler, Fridrich, & Pevný, 2013) defined two types of universal steganalysis, that is, supervised and unsupervised. Untraditionally, only the schemes based on one-class classifier (Pevný & Fridrich, 2008-2; Lyu & Farid, 2006) are categorized into the supervised type. Such steganalysis only models the covers, and classifies each medium which does not resemble a cover into the stego-media. For the unsupervised schemes (Ker & Pevný, 2011; Ker & Pevný, 2012), the property of no need of training has been exploited to make the steganalysis universal in finding a steganographer. Then, one may wonder how the steganalysis based on the most common supervised learning can has any practical usage. Kodovský and Fridrich (Kodovský & Fridrich, 2008) pointed out that besides being used as specific attacks, such steganalysis can be used as an oracle for designing steganography, where an oracle is often a theoretical tool for proving the security of a cryptosystem (Mao, 2004; Schneier, 1996).

More recently, some researchers begun to think that the a priori knowledge of media plays another important role in steganalysis. In most existing schemes, the media used in training and testing, often from one database, have a set of uniform parameters such as the same size and QF. However, some works do not made the assumption and deal with the problem of mismatch between the training and testing media. A method proposed by Pevný & Fridrich (2008-1) can recognize the double-compressed JPEG files and let them be steganalyzed by a special classifier. The forensic-aided steganalysis proposed by Barni, Cancelli, & Esposito (2010) has 2 classifiers, one for computer graphics images and the other for camera-generated images. It also has a pre-classifier used to differentiate the types of images at first. Amirkhani & Rahmati (2011) used the images of different content types to train different classifiers responsible for analyzing the corresponding types of content. Images can also be divided into the uncompressed and those compressed with different QFs to train the corresponding classifiers (Hou, Zhang, Xiong, & Wan, 2012). Moreover, in the training, they can be classified by joint image characteristics including both size and quantization factor (Deng, Guan, Zhao, Zhu, & Cao, 2015), or by camera sources (Kodovský, Sedighi, & Fridrich, 2014).

Although some paradigms of steganalysis have been proposed, nonetheless, there is still a lack of a systematic approach to defining them. For example, one may still ask whether there are more paradigms or what their relations and functionalities are. In this paper, a paradigm is defined as a typical way of designing and performing steganalysis under a specific circumstance in which different level of a priori knowledge is available. For instance, the most studied model of steganalysis, which we can call mainstream or dominant paradigm, means that a steganalyst has all a priori knowledge of steganography and cover-media. We prefer using the term 'paradigm' instead of 'model' because 'model' has been extensively used as a way of designing features. In our view, current research on steganalysis also lacks adequate discussion on the paradigms other than the dominant one. Ker, Bas, Böhme, Cogranne, Craver, Filler, Fridrich, & Pevný (2013) discussed the difference between the dominant paradigm and the expected real-world paradigm, and think that the most studied steganalysis is only the techniques in laboratory. Very likely, the existence of a dominant paradigm and the lack of research on the others have greatly simplifies the research.

In this paper, we will give an approach to systematically defining the paradigms of steganalysis by first defining four extreme paradigms and then defining the rest ones among them. The paper is organized as follows. In Section 2, the current mainstream paradigm will be reviewed with its actual aim being discussed. Also a comparative overview on the research models of steganalysis and cryptanalysis will be given. In Section 3, the four extreme paradigms will be defined and the rest ones will be defined accordingly. In Section 4, some typical instances of the steganalytic paradigms will be constructed and experimented. Finally, the conclusions will be drawn in Section 5.