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

Breaking Steganography:
Slight Modification with Distortion Minimization

Zhenxing Qian, School of Computer Science, Fudan University, Shanghai, China

Zichi Wang, Shanghai Institute for Advanced Communication and Data Science, Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai, China

Xinpeng Zhang, School of Computer Science, Fudan University, Shanghai, China

Guorui Feng, School of Communication and Information Engineering, Shanghai University, Shanghai, China

ABSTRACT

This article describes how to overcome the shortage of steganalysis for small capacity-based embedding. A slight modification method is proposed to break steganography. For a given image, traditional steganalysis methods are first used to achieve a preliminary result. For the "clear" image judged by steganalysis, it is still suspicious because of the incompleteness of steganalysis for small capacity. Thus, slight modifications are made to break the possibility of covert communication. The modifications are made on the locations with minimal distortion to guarantee high quality of the modified image. To this end, a proposed distortion minimization based algorithm using slight modification. Experimental results show that the error rate of secret data extraction is around 50% after implementation, which indicates that the covert communication of steganography is destroyed completely.

KEYWORDS

Distortion Minimizing, Modification, Steganalysis, Steganography

INTRODUCTION

Digital image steganography aims to transmit data secretly by embedding secret data into cover image (Zhang, 2016). Nowadays, the plentiful images transmitted over the social network provide convenience for steganography. How to break steganography is becoming a troublesome issue. Steganalysis attempts to reveal the presence of the embedded data. As shown in Figure 1, however, because of the missing detection error of steganalysis, the images judged as "clear" by steganalysis are still possible carried small amount of secret data. Therefore, some slight modifications should be made to prevent the still possible covert communication. On the other hand, the images judged as "stego" by steganalysis are still possible innocent because of the false alarm error. There are many kind of image processing operations frequently used on the images transmitted over social network, such as denoising, recompression, and beautification. So, the false alarm error may be caused by these

DOI: 10.4018/IJDCF.2019010109

Figure 1. The combat against steganography

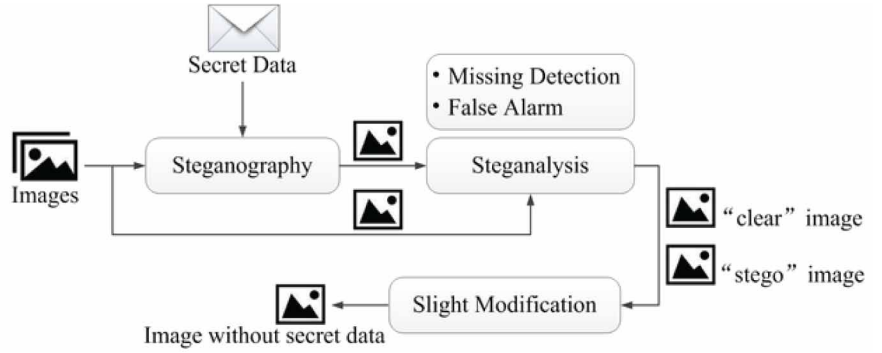

image processing operations instead of steganography. In this case, it is inappropriate to intercepte all the "stego" images. So, these images can also be processed as same as the "clear" image to stop the quite possible covert communication if the evidence is not enough to declare the guilty of the images. In this way, the combat against steganography is victorious, and meanwhile, the innocent images are transmitted over the social network as usual.

In addition, for early steganographic methods which increase the security performance by decreasing the quantity of embedding changes (Frdrich & Soukal, 2006;Zhang & Wang, 2006; Zhang, Zhang & Wang, 2008), current machine learning based steganalytic methods (Kodovsky, Fridrich & Holub, 2012; Fridrich & Kodovsky, 2012; Holub & Fridrich, 2014; Denemark, Sedighi, Holub, Cogranne & Fridrich, 2014; Song, Liu, Yang, Luo & Zhang, 2015) perform excellent detectability. But for modern steganographic methods (Holub & Fridrich, 2013, Li, Wang, Huang & Li, 2014, Sedighi, Fridrich & Cogranne, 2015, Guo, Ni, Su, Tang & Shi, 2015; Wang, Zhang & Yin, 2016) which improve security performance by minimize the additive distortion between a given cover object and its stego version (Filler, Judas & Fridrich, 2011, Filler & Fridrich, 2010), steganalysis becomes powerless to verdict the presence of secret data especially for the case of small capacity. Recently, adaptive steganalysis (Denemark, Boroumand & Fridrich, J. 2016; Yu, Li, Cheng, Zhang, 2016; Tang, Li, Luo & Huang, 2016) improves the detectability observably. In adaptive steganalysis, different weights are assigned to different cover elements in feature extraction. For the elements with high modifying probabilities, larger weights are assigned since these elements contribute more to steganalysis and vice versa. For small capacity, however, these methods still not perform satisfactory detectability. The detection for small capacity is still a to be resolved problem. In other words, the approach to break steganography is still undiscovered.

Actually, to break the covert communication of steganography, steganalysis is not the only choice. For a suspicious image, the possibly existing secret data can be destroyed by modifying the image although it is difficult to judge whether the image is stego or not. In this way, there is no secret data can be transmitted via the modified image. Thus, the threat from steganography is disappeared.

This paper proposes a slight modification method to break steganography. Some slight modifications are made on a suspicious image to break the extracting of secret data. And these modifications are made on the locations with minimal distortion to guarantee high quality of the modified image. Experiment results show that the secret data cannot be extracted from the modified image, and the modified image keeps a high quality comparing with the given image.