Batch Steganography and Pooled Steganalysis

Andrew D. Ker

Oxford University Computing Laboratory, Parks Road, Oxford OX1 3QD, England
adk@comlab.ox.ac.uk

Abstract. Conventional steganalysis aims to separate cover objects from stego objects, working on each object individually. In this paper we investigate some methods for pooling steganalysis evidence, so as to obtain more reliable detection of steganography in large sets of objects, and the dual problem of hiding information securely when spreading across a batch of covers. The results are rather surprising: in many situations, a steganographer should *not* spread the embedding across all covers, and the secure capacity increases only as the square root of the number of objects. We validate the theoretical results, which are rather general, by testing a particular type of image steganography. The experiments involve tens of millions of repeated steganalytic attacks and show that pooled steganalysis can give very reliable detection of even tiny proportionate payloads.

1 Introduction

The classic definition of *steganography* involves an actor (Steganographer) aiming to communicate with a passive conspirator over an insecure channel, and an eavesdropper (or Warden) monitoring the channel. The Steganographer hides his communication inside some other medium by taking a seemingly-innocent *cover object* and making changes, hopefully imperceptible to the Warden, which convey the secret information to the recipient. The Warden's aim is not to decode the hidden information but merely to deduce its presence. This is *steganalysis*: the creation of hypothesis tests which can distinguish cover objects from so-called stego objects in which a payload has been embedded. Such language assumes that each cover object is treated in isolation by both the embedder and the eavesdropper, and in the literature the focus is almost exclusively on single cover objects (usually individual digital images, but also sometimes audio files, movies, or more unusual digital objects). In this paper we begin to ask about large groups of cover objects, and how the methods for both embedding into, and steganalysis of, individual pieces can be applied to the groups as a whole.

There are two good reasons for doing so. First, we contend that practical applications of steganalysis inevitably will involve multiple objects: the Warden will surely have intercepted more than one communication from the Steganographer, and the Steganographer will surely have access to more than one cover. Second, even given state-of-the-art steganalysis and weak steganography, very high reliability steganalysis (in which false positive rates are as low as, say,

J. Camenisch et al. (Eds.): IH 2006, LNCS 4437, pp. 265–281, 2007.

10^{-5}) is simply not possible with the small amount of evidence obtained from a single cover (except for deeply flawed steganography which leaves a particular signature, or perhaps enormous objects such as entire digital movies).

In this paper we assume that an imperfect method of statistical steganalysis already exists for individual cover objects, and investigate how the set of detection statistics computed over a group can be combined by the Warden into an overall detector for steganography for the whole group. This gives information on the opposite problem, where the Steganographer has to decide how best to spread secret information amongst a batch of covers. The answers to this latter question, at least for some of the pooled detectors suggested here, are rather surprising. There seems to be little literature on this problem: Trivedi [1] has used sequential hypothesis tests to repeat steganalysis, but only in the context of locating a hidden message embedded sequentially within a single image.

In Sect. 2 we formulate more precisely the competing aims of *batch steganography* and *pooled steganalysis*. In this work, which only scratches the surface of what appears to be a complex topic, we allow certain assumptions (which are not implausible) about the steganalysis methods for individual objects which we aim to combine; they are discussed in Sect. 2. In Sect. 3 we suggest three possible pooling strategies for the Warden, analysing them for performance and deriving the Steganographer's best tactic to avoid detection. In Sect. 4 we move away from the abstract nature of the first part of the paper, and focus on Least Significant Bit Replacement in digital images, a well-studied problem; for this embedding method, and a popular detection algorithm, we perform millions of simulations to benchmark the strategies of Sect. 3, confirming the theoretical results. Briefly, we return to our assumptions about steganalysis response – there is a sting in the tail here. Finally, Section 5 suggests avenues for further work.

2 Problem Formulation

The scenario we have in mind, which motivates this paper, is the following. Suppose that a criminal wishes to hide information on his computer, deniably, using steganography. He already has a large number of innocent cover pictures on his hard disk. To be quite sure of hiding his secret information well, he might split it into many small pieces and hide a little in each of a selection of the pictures, believing that this is more secure than the alternative of filling a smaller number of images to maximum capacity.

When the authorities impound his computer, they are faced with a dilemma: how do they know which pictures to examine? Even possessing state-of-the-art steganalysis, they still observe fairly large false positive rates, and so if they test every picture on his computer they will inevitably turn up a lot of positive diagnoses – even if he is not a steganographer at all. They must run their statistical detector on every picture individually, and then find some way to combine the detection statistics into an overall "pooled" steganalysis for the presence of data, possibly spread across all the images.