IGI PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com

ITJ 4563

This paper appears in the publication, International Journal of Digital Crime and Forensics,Volume 1, Issue 1
edited by **Chang-Tsun Li © 2009, IGI Global**

Locally Square Distortion and Batch Steganographic Capacity

Andrew D. Ker, Oxford University Computing Laboratory, UK

ABSTRACT

A fundamental question of the steganography problem is to determine the amount of data which can be hidden undetectably. Its answer is of direct importance to the embedder, but also aids a forensic investigator in bounding the size of payload which might be communicated. Recent results on the information theory of steganography suggest that the detectability of payload in an individual object is proportional to the square of the number of changes caused by the embedding. Here, we follow up the implications when a payload is to be spread amongst multiple cover objects, and give asymptotic results about the maximum secure payload. Two embedding scenarios are distinguished: embedding in a fixed finite batch of covers, and continuous embedding in an infinite stream. The steganographic capacity, as a function of the number of objects, is sublinear and strictly asymptotically lower in the second case. This work consolidates and extends our previous results on batch and sequential steganographic capacity.

Keywords: *batch steganography; hidden information; information theory; source coding; steganographic capacity*

INTRODUCTION

We consider the following question: given a set of cover objects, how much data could be hidden in them? Although there is much literature on embedding and detection of steganographic payload, it is usual to consider only single cover objects, whereas this article is concerned with embedding in a finite or infinite stream of objects, deriving capacity bounds and optimal methods. We posed the questions about embedding and detection in a fixed number of covers in Ker (2006), where it was called the *batch steganography* problem, and the question is now also extended to infinite streams; we call this *sequential steganography*.

A key assumption, here, will be that the detectability of payload in a single object is (either exactly or locally for small payloads) proportional to the *square* of the number of changes caused by the embedding. Results of this nature have recently arisen in a number of theoretical steganalysis papers (Ker, 2007b, 2007c, 2007d) and the phenomenon has also been observed experimentally (Ker, Pevný, Kodovský, & Fridrich, 2008). Assuming that the same holds in general, we examine the implications for an embedder when a large payload is to be spread amongst multiple cover objects. The choice of how to split payload between multiple covers is called an *embedding strategy* and the aim is to find the optimal strategies implied by the square law. There is some recent related work (Ker, 2006, 2007a) where optimal embedding strategies were found, but only in the context of highly restricted detection frameworks; in this article we do not assume knowledge of the steganalyst's behaviour.

The structure of this article is as follows. In the *Problem Formulation* section we will present the problems of batch steganography and sequential steganography; we will make and justify a series of assumptions about how steganalysis evidence accumulates. Evidence is not generated by payload itself—it is found as changes in the cover object, caused by the embedding process—so we must also relate embedding changes to payload transmitted and, with adaptive source coding methods, these are not always proportional (Fridrich & Soukal, 2006; Bierbrauer & Fridrich, 2008). In the *Analysis of the Batch Steganography Problem* section we will apply the theory to the batch steganography problem, deriving optimal embedding strategies and maximum undetectable payload, and in the *Analysis of the Sequential Steganography Problem* section to the sequential steganography problem; there is no optimal strategy in this case, but bounds can be derived, and strategies exist which come arbitrarily close to the bounds. It will be shown that the asymptotic payload, as a function of the number of covers, must be strictly lower in the sequential than the batch setting. Finally in the concluding *Discussion* section we will discuss the significance and limitations of the results.

An early version of some of this work has appeared in conference proceedings without any mathematical proofs (Ker, 2008b). In this work we have changed focus to concentrate on the embedding changes—this reduces the algebraic complexity—and are able to widen the applicability and weaken the assumptions. In particular, the square evidence law need hold only locally as payloads tend to zero.

Before continuing, we review some asymptotic notation. We write $f(n) = O(g(n))$ if there are constants c and N such that $f(n) \leq cg(n)$ for all $n \geq N$. The analogous *strict* bound is $f(n) = o(g(n))$, which means that $f(n)/g(n) \rightarrow 0$. We write $f(n) = \Theta(g(n))$ if there are positive constants c, d and N such that $cg(n) \leq f(n) \leq dg(n)$ for all $n \geq N$. The most precise condition on growth is $f(n) \sim g(n)$, which means that $f(n)/g(n) \rightarrow 1$.

PROBLEM FORMULATION

It is rather plausible to suppose that a steganographer has access to multiple covers among which the payload can be spread, and that a steganalyst is presented with a large number of objects for steganalysis. We formulated (Ker, 2006) the competing aims of *batch steganography*, in which it is assumed that a fixed set of N covers is available to a steganographer who spreads payload amongst some or all of them, and *pooled steganalysis*, in which a steganalyst attempts to pool the evidence of N objects to determine whether some payload is present (without knowing which or how many do contain payload). Only the former will concern us here: we want to determine, subject to some assumptions about accumulation of evidence and a maximum acceptable risk of detection, how much payload can be embedded. In some cases we will also be able to identify the optimal strategies for the steganographer.

We also tackle a more difficult problem, dubbed *sequential steganography*. In the sequential setting we no longer suppose that the number of covers N is fixed in advance of embedding (this differs materially from the batch problem, because optimal strategies require advance knowledge of N). In the sequential setting, we want to establish a strategy for an infinite stream of communications, with transmission of as much payload as possible over time. We will see that, although the steganographer is forced to reduce the payload *rate* over time, an infinite payload can still be transmitted in an infinite amount of time. However it will be shown that