1

Introduction

Steganography is the art of communicating a secret message, from Alice to Bob, in such a way that Alice's evil sister Eve cannot even tell that a secret message exists. This is (typically) done by hiding the secret message within a non-sensitive one, and Eve should believe that the non-sensitive message she can see is all there is. Steganalysis, on the contrary, is Eve's task of detecting the presence of a secret message when Alice and Bob employ steganography.

1.1 Real Threat or Hype?

A frequently asked question is *Who needs steganalysis?* Closely related is the question of who is using steganography. Unfortunately, satisfactory answers to these questions are harder to find.

A standard claim in the literature is that terrorist organisations use steganography to plan their operations. This claim seems to be founded on a report in *USA Today*, by Kelley (2001), where it was claimed that Osama bin Laden was using the Internet in an 'e-*jihad*' half a year before he became world famous in September 2001. The idea of the application is simple. Steganography, potentially, makes it possible to hide detailed plans with maps and photographs of targets within images, which can be left on public sites like e-Bay or Facebook as a kind of electronic *dead-drop*.

The report in *USA Today* was based on unnamed sources in US law enforcement, and there has been no other evidence in the public domain that terrorist organisations really are using steganography to plan their activities. Goth (2005) described it as a hype, and predicted that the funding opportunities enjoyed by the steganography community in the early years of the millennium would fade. It rather seems that he was correct. At least the EU and European research councils have shown little interest in the topic.

Machine Learning in Image Steganalysis, First Edition. Hans Georg Schaathun.
© 2012 John Wiley & Sons, Ltd. Published 2012 by John Wiley & Sons, Ltd.

Steganography has several problems which may make it unattractive for criminal users. Bagnall (2003) (quoted in Goth (2005)) points out that the acquisition, possession and distribution of tools and knowledge necessary to use steganography in itself establishes a traceable link which may arouse as much suspicion as an encrypted message. Establishing the infrastructure to use steganography securely, and keeping it secret during construction, is not going to be an easy exercise.

More recently, an unknown author in *The Technical Mujahedin* (Givner-Forbes, 2007; Unknown, 2007) has advocated the use of steganography in the *jihad*, giving some examples of software to avoid and approaches to evaluating algorithms for use. There is no doubt that the technology has some potential for groups with sufficient resources to use it well.

In June 2010 we heard of ten persons (alleged Russian agents) being arrested in the USA, and according to the news report the investigation turned up evidence of the use of steganography. It is too early to say if these charges will give steganalysis research a new push. Adee (2010) suggests that the spies may have been thwarted by old technology, using very old and easily detectable stego-systems. However, we do not know if the investigators first identified the use of steganography by means of steganalysis, or if they found the steganographic software used on the suspects' computers first.

So currently, as members of the general public and as academic researchers, we are unable to tell whether steganography is a significant threat or mainly a brain exercise for academics. We have no strong evidence of significant use, but then we also know that MI5 and MI6, and other secret services, who would be the first to know if such evidence existed, would hardly tell us about it. In contrast, by developing public knowledge about the technology, we make it harder for criminal elements to use it successfully for their own purposes.

1.2 Artificial Intelligence and Learning

Most of the current steganalysis techniques are based on machine learning in one form or another. Machine learning is an area well worth learning, because of its wide applications within medical image analysis, robotics, information retrieval, computational linguistics, forensics, automation and control, etc. The underlying idea is simple; if a task is too complex for a human being to learn, let's train a machine to do it instead. At a philosophical level it is harder. What, after all, do we really mean by *learning*?

Learning is an aspect of intelligence, which is often defined as the ability to learn. Machine learning thus depends on some kind of artificial intelligence (AI). As a scientific discipline machine learning is counted as a sub-area of AI, which is a more well-known idea at least for the general public. In contrast, our impression of what AI is may be shaped as much by science fiction as by science. Many of us would first think of the sentient computers and robots in the 1960s and 1970s literature, such as Isaac Asimov's famous robots who could only be kept from world domination by the three robotic laws deeply embedded in their circuitry.

As often as a dream, AI has been portrayed as a nightmare. Watching films like *Terminator* and *The Matrix*, maybe we should be glad that scientists have not yet managed to realise the dream of AI. Discussing AI as a scientific discipline today, it may be more fruitful to discuss the different sub-disciplines. The intelligent and sentient computer remains science fiction, but various AI-related properties have been realised with great success and valuable applications. Machine learning is one of these sub-disciplines.

The task in steganalysis is to take an *object* (communication) and classify this into one out of two classes, either the class of steganograms or the class of clean messages. This type of problem, of designing an algorithm to map objects to classes, is known as *pattern recognition* or classification in the literature. Once upon a time pattern recognition was primarily based on statistics, and the approach was analytic, aiming to design a statistical model to predict the class. Unfortunately, in many applications, the problem is too complex to make this approach feasible. Machine learning provides an alternative to the analytic approach.

A learning classifier builds a statistical model to solve the classification problem, by brute-force study of a large number of statistics (so-called *features*) from a set of objects selected for training. Thus, we say that the classifier learns from the study of the training set, and the acquired learning can later be used to classify previously unseen objects. Contrary to the analytic models of statistical approaches, the model produced by machine learning does not have to be comprehensible for human users, as it is primarily for machine processing. Thus, more complex and difficult problems can be solved more accurately.

With respect to machine learning, the primary objective of this book is to provide a tutorial to allow a reader with primary interest in steganography and steganalysis to use black box learning algorithms in steganalysis. However, we will also dig a little bit deeper into the theory, to inspire some readers to carry some of their experience into other areas of research at a later stage.

1.3 How to Read this Book

There are no ‘don’t do this at home’ clauses in this book. Quite the contrary. The body of experimental data in the literature is still very limited, and we have not been able to run enough experiments to give you more than anecdotal evidence in this book. To choose the most promising methods for steganalysis, the reader will have to make his own comparisons with his own images. Therefore, the advice must be ‘don’t trust me; try it yourself’. As an aid to this, the software used in this book can be found at: **http://www.ifs.schaathun.net/pysteg/**.

For this reason, the primary purpose of this book has been to provide a hands-on tutorial with sufficient detail to allow the reader to reproduce examples. At the same time, we aim to establish the links to theory, to make the connection to relevant areas of research as smooth as possible. In particular we spend time on statistical methods, to explain the limitations of the experimental paradigms and understand exactly how far the experimental results can be trusted.

In this first part of the book, we will give the background and context of steganalysis (Chapter 2), and a quick introduction and tutorial (Chapter 3) to provide a test platform for the next part.

Part II is devoted entirely to feature vectors for steganalysis. We have aimed for a broad survey of available features, but it is surely not complete. The best we can hope for is that it is more complete than any previous one. The primary target audience is research students and young researchers entering the area of steganalysis, but we hope that more experienced researchers will also find some topics of interest.

Part III investigates both the theory and methodology, and the context and challenges of steganalysis in more detail. More diverse than the previous parts, this will both introduce various classifier algorithms and take a critical view on the experimental methodology and applications in steganalysis. The classification algorithms introduced in Chapters 11 and 12 are intended to give an easy introduction to the wider area of machine learning. The discussions of statistics and experimental methods (Chapter 10), as well as applications and practical issues in steganalysis (Chapter 14) have been written to promote thorough and theoretically founded evaluation of steganalytic methods. With no intention of replacing any book on machine learning or statistics, we hope to inspire the reader to read more.