High Performance Genetic Algorithms for Steganalysis

Joseph Charles Bingham
Department of Mathematics
Iowa State University of Science and Technology
Ames, Iowa 50010, United States of America
jbingham@iastate.edu

April 21, 2018

Abstract

This research will outline a novel implementation of a genetic algorithm that leverages high performance parallelizations to detect steganaraphically embedded images. The two main components that are new to this project are the application of parallelization for genetic algorithms and the application of genetic algorithm for steganalysis. Typical steganalytic methods which use machine learning techniques require an unreasonable amount of pre-classified data and copious amounts of time for training the engine. The data needed for such operation usually must be lab generated, which can lead to the biases when compared to real world, and often is constrained to specific parameters, such as they must remain within either the spacial domain or the JPEG domain, must be the same pixel width, etc., making these engines limited to what image space they detect over. The need for all data to be used in training the engine, as well as the linear nature of the engines used precludes them from being parallelized in any meaningful fashion.

This new algorithm does not succumb to these shortfalls. By using solution sets of pixels as the data which the genetic engine trains over, and fixing the image(s) under suspicion, the engine can be parallelized, vastly increasing the efficiency. Since the algorithm searches for sets

solutions and then uses a fitness function to determine whether the findings were statistically significant, it does not require training data at all, mean that there are no biases introduced by lab generated data. The algorithms flexibility does not limit it to just one format of images, meaning that the same program can be used without the need to generate a new set of data or train a new engine.

1 Background Information

Steganography and Steganalysis Steganography is the practice of embedding information into photographs or audio media in such a way that it is not detectable to the average person. The form of steganography that the scope of this project will be most interested in is LSB (Least Significant Bit) embedding of images using random paths (1).

This, as the name suggests, is where the path of the embedding stream is random in nature. When a pixel from the image is selected to be altered, the least significant bit is overridden to be the same as the next bit in the embedding stream. Since it is just the least significant bit, very little visual alterations occur. This makes detection difficult, if not impossible without the aid of software.

Steganalysis is the discovery of the existence of hidden information; therefore, like cryptography and cryptanalysis, the goal of steganalysis is to discover hidden information and to break the security of its carriers (2).

Genetic Algorithms Genetic Algorithms are algorithms that follow a specific machine learning paradigm [3]. They are classified by their similarity to natural selection as found in evolution in the wild. They consist of three main parts: a cost function, a tester bot, and a builder bot.

The cost function determines the viability of each solution. Typically this function determines how well the solution completes the task that is trying to be accomplished, and is usually from $\mathbb{R}^n \mapsto [0, 1]$.

The tester bot tests each solution based on the cost function. Its job is to determine what the value of the test function and to rank each solution.

The builder bot takes each the best solutions, as ranked by the tester bot, generates more solutions based on their attributes. It generates the new solutions by randomly modifying the best solutions in hopes of descending the gradient.

2 Previous work

3 Results

4 Conclusions

(1) https://link.springer.com/article/10.1155/2010/876946
(2) https://www.sans.org/reading-room/whitepapers/stenganography/steganalysis-detecting-hidden-information-computer-forensic-analysis-1014
(3) https://link.springer.com/article/10.1007/BF00175354