Chapter 11
Steganography Encoding as Inverse Data Mining

Dan Ophir
Ariel University, Israel & Tel Aviv University, Israel

ABSTRACT

Supercomputers and cloud computing seem to be competing paradigms. Supercomputing focuses on increasing CPU speed, thus significantly increasing the speed of its associated memory access and its capacity. Conversely, cloud computing increases the computing throughput by parallel computing, spreading computing tasks over unused nodes and platforms. Steganography, the art of concealing a message within a message, is a type of encoding whose operations are required to remain secret. Steganography encoding requires data manipulation and is linked to data mining methodologies. Data mining reveals concealed data that is embedded in exposed data. Encoding by steganography is reverse data mining, hiding data among visible data. Conventionally, encryption methods are used to successfully hide the data. Cloud computing can take the data and disperse it in a way that even without any encryption, each individual packet of data is meaningless, thus hiding the message as like by steganography. This chapter explores steganography encoding as inverse data mining.

INTRODUCTION: PICTURE OF STEGANOGRAPHY

This paper will analyze the current implications of cloud computing on steganography (e.g. Hayati P., et al. 2005) that is based on signal and image processing algorithms. Steganography, which means concealed writing, is the art and science of writing a message in a way that no one except the intended recipient suspects that there is any hidden message at all. The concealed information can be images, text, or any type of binary data. This work will focus on uses of steganography which include the following:

Data Hiding

Images or any other types of data can be concealed in another image, leaving the manipulated image as visually similar as possible to the initial image (Figure 1).

DOI: 10.4018/978-1-4666-7461-5.ch011

Figure 1. An example of steganographic manipulation: image (a), the original image, has been overlapped with image (b). In order to see this superimposed image, the observer has to look at the picture at a distance of about 50 cm (the image is a private acquisition)

Analyzing and Detecting

An image can be analyzed for the existence of hidden data. If such hidden data are found, it can be extracted and saved externally.

There are two modes of steganography:

- **Hiding:** This mode's purpose is to hide the information in the image. The hidden data may be any binary information.
- **Decoding:** The action of interpreting the hidden information.

Data Mining vs. Inverse Data Mining

In order to understand the term "inverse data mining", the better known "data mining" term will be explained.

Data mining, similarly to mineral mining, is the art of extracting objects that are different by some set of properties from their surroundings. In most cases, both in mineral mining and data mining, the valuable material is a very small percentage of the overall volume.

The different properties of the valuable material are exploited to separate them from their environment. For example, iron is separated from surrounding material by either using its magnetic properties or lower melting point (smelting). The genome (Watson J. D. 2003) investigation (Ophir, 2013) is an example of data-mining, where specific sequences, for example ternary tracts, are separated from the whole genome. This operation is very CPU-time consuming and requires the use of supercomputers.

Data mining places, sometimes, even greater challenges than the mineral mining before the potential miner. This statement is expressed in the fact that the data miner doesn't know what to look for, whereas the mineral miner knows what is he looking for. The data-miner generally looks for something different than the surroundings. This is usually the starting point for most data mining.

Data mining is looking for exceptional data, or conversely, looking for data properties that have common denominators with the whole or part of the investigated data collection. This common