Chapter 74
Data Hiding for Text and Binary Files

Hioki Hirohisa
Kyoto University, Japan

ABSTRACT

This chapter presents an overview of text-based and binary-based data hiding methods. Text methods, through which secret information is embedded into innocent-looking textual data, are mostly used for steganography. Binary methods are applied to program binary codes: executables and libraries. In binary methods, information is embedded into a binary code so that its functionality is preserved. Data hiding methods for binary codes have been studied intensively to perform watermarking for protecting software from piracy acts. A message can also be embedded into a binary code in a steganographic manner. Another method is also introduced, which is proposed for enhancing the performance of an executable file.

INTRODUCTION

Data hiding methods are mainly applied to multimedia files such as image, video, and audio files to embed information imperceptibly to files. The purpose of data hiding includes discouraging unauthorized copying, concealing the existence of secret messages, and adding extra values to digital files. A data hiding technique for copyright protection is specifically called watermarking. Steganography refers to the technique of concealing secret messages. Multimedia files are indeed suitable for data hiding because they provide a fair amount of exploitable redundancies. For example, slight modification in pixel values of an image will not damage it severely. We can thus embed information into the image while preserving its appearance.

In contrast, text files or binary executable files contain fewer redundancies. Changing a single character in a text file document may be noticeable. Flipping even a single bit of an executable file can make it totally useless. However, it is still possible to use text files or binary executable files for data hiding.

Text data hiding methods date back to the pre-computer era. Secret messages are stealthily embedded into manually composed apparently innocuous texts in order to deceive enemies or even for fun. In our information network era, text steganographic methods are also applied to digital text files. Messages are embedded into texts by modifying their formats or contents on the basis of linguistic knowledge. Texts with hidden data are called "stego" texts and sometimes

DOI: 10.4018/978-1-4666-6042-7.ch074

can be even generated from scratch. Moreover, there are steganographic methods available for non-documental texts such as program sources.

With well-deliberated knowledge, we can transform a binary code to a functionally equivalent code, thereby embedding information. Data hiding methods for binary executable files have been intensively studied to perform watermarking for protecting software from piracy acts. A message can also be embedded into a binary code in a steganographic manner. Another method is also introduced, which is proposed for enhancing the performance of an executable file.

In this chapter, various text data hiding methods including classical ones are presented. Then, binary data hiding methods for watermarking, steganography, and system performance enhancement are described.

TEXT DATA HIDING METHODS

Text data hiding methods are mostly used for steganographic purposes. Text steganographic methods are not limited to digital texts. Steganographic methods have been studied from ancient times.

It has been reported that (Bennett, 2004) digital text steganographic methods include format-based and linguistic methods. In format-based methods, existing text is modified but retains visible characters. Modification is made independent of the textual content. Linguistic methods are those exploiting linguistic knowledge for embedding information. Embedding is done by either modifying the existing text with considering its content or by generating text that looks plausible to the reader.

In this section, before discussing the topics of digital text steganography, several classical steganographic methods are described. Then, the principles of format-based and linguistic methods used in digital steganographic methods are reviewed. Methods for non-documental texts–such as program sources and game logs–are also reviewed.

Classical Methods

Text-based steganographic methods date back to the pre-computer era, because text was the only medium that could be freely edited before computers became widely available. Classical methods include acrostic and null cipher. An acrostic is a text where a message is hidden typically in the initial letters of each line of the text. Null cipher can be regarded as a variant of acrostic. A null cipher hides a message in each word in a text.

An acrostic written by Lewis Carroll appears in his book *Through the Looking-Glass, and What Alice Found There* (Carroll, 1871), as shown in Listing 1. The initial letters of each line spell out "Alice Pleasance Liddell"–the real name of Alice.

The following is a null cipher crafted by a German spy during World War 1 (Kahn, 1996):

PRESIDENT'S EMBARGO RULING SHOULD HAVE IMMEDIATE NOTICE. GRAVE SITUATION AFFECTING INTERNATIONAL LAW. STATEMENT FORESHADOWS RUIN OF MANY NEUTRALS. YELLOW JOURNALS UNIFYING NATIONAL EXCITEMENT IMMENSELY.

Taking the initial letters of each word, the message "Pershing sails from N.Y. June 1" pops up. A check message was returned as another null cipher:

APPARENTLY NEUTRAL'S PROTEST IS THOROUGHLY DISCOUNTED AND IGNORED. ISMAN HARD HIT. BLOCKADE ISSUE AFFECTS PRETEXT FOR EMBARGO ON BYPRODUCTS, EJECTING SUETS AND VEGETABLE OILS.

The same message as the first one appears if we pick up the second letters of each word.

Francis Bacon (1561–1626) invented a steganographic method based on the bi-literarie alphabet (Kahn, 1996). In this method, each alphabet is encoded as a five-bit binary code. A message is hidden into an innocuous text. Each character of the text represents a bit. Two different fonts are