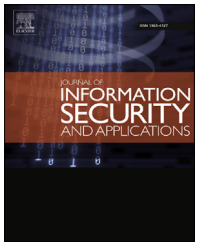
Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jisa



PDF steganography based on Chinese Remainder Theorem

Stéphane Gael R. Ekodeck [a,b,c], *René Ndoundam* [a,b,c,*]

[a] *LIRIMA, Team GRIMCAPE, Cameroon CETIC, University of Yaounde I, P.O. Box 812 Yaounde, Yaounde, Cameroon*

[b] *IRD, UMI 209, UMMISCO, IRD France Nord, F-93143, Bondy, France*

[c] *Univ. Paris 06, UMI 209, UMMISCO, Sorbonne Unversités, F-75005, Paris, France*

ARTICLE INFO

Article history:
Available online 3 February 2016

Keywords:
Steganography
PDF files and readers
Chinese Remainder Theorem

ABSTRACT

We propose different approaches of PDF file based steganography, essentially based on the Chinese Remainder Theorem. Here, after a cover PDF document has been released from unnecessary characters of ASCII code A0, a secret message is hidden in it using one of the proposed approaches, making it invisible to common PDF readers, and the file is then transmitted through a non-secure communication channel. Each of our methods tries to ensure the condition that the number of inserted A0 is less than the number of characters of the secret message *s*.

1. Introduction

Information hiding is an emerging research area which encompasses many applications such as copyright protection for digital media, watermarking, fingerprinting, data embedding and steganography (Moulin and O'Sullivan, 2003).

Steganography (Wayner, 2009), as shown on Fig. 1, is one of the most important ones. It is derived from a work by Johannes Trithemus (1462–1516) entitled *Steganographia* and comes from the Greek name *steganos* (hidden or secret) and *graphy* (writing or drawing) and literally means *hidden writing* (Por and Delina, 2008; Richer, 2003). It is an ancient art of hiding information whose goal consists in hiding a secret message in a public media (video, text, sound, image, etc.) acting as a cover, in a way that sent through a non-secure communication channel; only the sender and the receiver are able to understand it, and anyone else cannot distinguish the existence of an hidden message.

Steganography has been many times used throughout history and a very famous example is a message which a German spy sent in World War II (Johnson and Jajodia, 1998):

> *"Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by products, ejecting suets and vegetable oils."*

Taking the second letter in each word reveals the following secret message:

> *"Pershing sails from NY June 1."*

Many steganography schemes throughout history were developed, allowing to divide it into two main subdomains as shown on Fig. 1, where Linguistic Steganography is defined by Chapman et al. (Por and Delina, 2008) as "*the art of using written natural language to conceal secret messages*", and Technical Steganography is defined as a structure, rather than a text, that

* *Corresponding author*. LIRIMA, Team GRIMCAPE, Cameroon CETIC, University of Yaounde I, P.O. Box 812 Yaounde, Yaounde, Cameroon. Tel.: 00(237)677495680; fax: 00(237)222221320.
E-mail address: ndoundam@yahoo.com (R. Ndoundam).

http://dx.doi.org/10.1016/j.jisa.2015.11.008

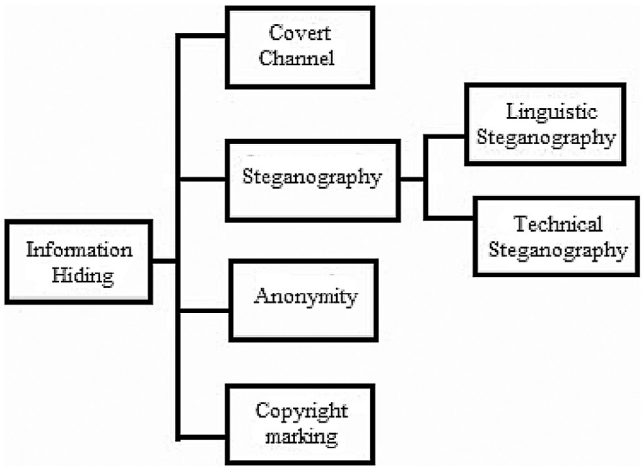


Fig. 1 – Classification of information hiding techniques.

can be represented by any physical means such as invisible inks, or microdots (Por and Delina, 2008). Most of the work in steganography has been done on images, video clips, music, sounds and texts. But text steganography is the most complex, due to the lack of redundant information in text files, whereas a lot of redundancy is present in image or sound files, leading to a high exploitation of those files in steganography (Govada et al., 2012; Li et al., 2008).

Text steganography integrates everything that goes from reformatting a text document to the change of words in the text, to the generation of random character sequences or using context free grammar to generate readable texts (Agarwal, 2013). The structure of text documents is identical with respect to that observed, while in regard to other types of documents such as images, the document structure is different from what is observed. Therefore, in such documents, one can hide information by introducing changes in the structure of the document without making any noticeable changes in the document preview output (Agarwal, 2013; Govada et al., 2012; Li et al., 2008). However, in text files, even a letter or punctuation can also be noticed by the reader (Agarwal, 2013).

Text steganography can be classified in three basic categories (Bennett, 2004; Por and Delina, 2008): format based, random and statistical generation and linguistic method, as shown in Fig. 2.

Format-based methods physically alter the format of a text in order to conceal information. Inserting spaces, deliberate sins through text, and resizing fonts are just some among all those used as a format based method in text steganography. However, these methods have a defect; if they can fool the human eye, this is not the case for a computer (Agarwal, 2013; Por and Delina, 2008).

Random and statistical generation allows the generation of coverage of texts based on statistical properties. It relies on sequences of characters and words. One approach would be to hide information in a sequence of characters that looks to be random. And a second approach would be to use the statistical properties of sizes of words and frequencies of letters to create words that have the same statistical properties as common words in a given language (Agarwal, 2013; Por and Delina, 2008).

Linguistic methods particularly consider the linguistic properties of the generated and modified text and, in many cases, uses linguistic structure as a place where to insert the hidden messages. In fact, steganographic data can be hidden within the syntactic structure itself (Agarwal, 2013; Por and Delina, 2008).

In this paper, different steganographic methods using text documents as cover media, more precisely PDF files, are proposed. These techniques are inspired by the work of Lee and Tsai (2010), who found in their study that the non-breaking space with American Standard Code for Information Interchange (ASCII) code A0 becomes invisible to common PDF readers and used that phenomenon to hide secret data, and thus proposed two methods where secret data are embedded at between-word or between-character locations in a PDF file.

As they stated in their work that an obvious disadvantage of the second method is that the resulting PDF file size would be higher than the original one, we focused on how to reduced that weight difference while increasing the embedding capacity of the cover PDF file. And to do so, we used in our approaches, in different ways, the Chinese Remainder Theorem (CRT) (Menezes et al., 1996; Shoup, 2008) to reduce the size of the secret data to embed in the cover PDF file and also to add to it randomness. We make use of the CRT because it provides benefits in computing, mathematics and also in cryptography, where the algorithm provides relief in case of generating random numbers and also in case of modular computation (Shoup, 2008), which allows the restriction of the length of all intermediate results and final value.

In the sequel, we present some PDF file studies found in the literature with focus on the work of Lee and Tsai in Section 2, then our contribution in Section 3 followed by the presentation of the CRT in Section 4. After that, we show how to prepare a cover PDF file in Section 5, before our different approaches embed secret data in it in Section 6, and we present in Sections 7 and 8, respectively, experimental results and discussion related to these methods. And finally a conclusion is presented in Section 9.

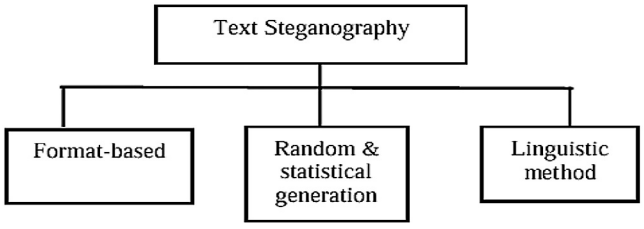


Fig. 2 – Basic categories of text steganography.

2. PDF file based steganography

PDF, created by Adobe Systems (Adobe Systems Incorporated, November 2006) for document exchange, is a fixed-layout format for representing documents in a manner independent of the application software, hardware, and operation system. PDF files are frequently used nowadays and this fact makes it possible to use them as cover documents in information hiding. Studies