# UNIVERSITA' DEGLI STUDI DI SALERNO
## DIPARTIMENTO DI INFORMATICA "RENATO M. CAPOCELLI"

CORSO DI DOTTORATO IN INFORMATICA
XI CICLO - NUOVA SERIE

anno accademico 2011-2012

**Bonaventura D'Alessio**

Ph.D Thesis

# Steganographic Methods for Information Hiding in MS-Office Files

## Abstract

The simplest container of digital information is "the file" and among the vast array of files currently available, MS-Office files are probably the most widely used. These files, starting from MS-Office 2007, respect the standards "Office Open XML" (OOXML). One of the benefits of the new format is that it lowers the risk of information leakage. Infact, before MS-Office 2007 was used binary files, called "Microsoft Document File Format" (MCDFF), were often used to host secret information.

In this work, starting from the classification of information hiding adapted from Bauer, methods for embedding data into the OOXML file format are analyzed. It also includes four novel methods. The first one, "*Data Hiding by Different Compression Algorithm in ZIP*", is based on the characteristic that the new MS-Office documents are container of compressed files. The second one, "*Data Hiding by Office Macro*", uses modified macro to store secret messages. The third one, "*Data Hiding by Zero Dimension Image*", is based on the characteristic that it is possible to create invisible images, setting the width and height both equal to zero. The last one, "*Data Hiding by Revision Identifier Value*" hides data by storing it in some attributes of XML elements (*rsid*). The methods presented can be combined in order to extend the amount of data to be hidden in a single cover file.

Analyzing a scenario composed of about 50; 000 MS-Office files, we show how the proposed methods can be helpful in real applications. We have calculated the capacity of all stegosystem described, verifying the amount of information that can be hidden. After that, an evaluation of the limits of the proposed methods is carried out by comparing them with the tool introduced by Microsoft to sanitize MS-Office files. This tool, called Document Inspector, was projected to help the users to find and remove hidden data and personal information in MS-Office documents.