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

Tesi di Dottorato in Informatica

Steganographic Methods for Information Hiding in MS-Office Files

---

Abstract

Il più semplice contenitore di informazioni digitali è "il file" e, tra la vasta gamma attualmente disponibile, i più utilizzati sono probabilmente i file creati con MS-Office. Microsoft, dalla suite MS-Office 2007, ha recepito lo standard denominato "Office Open XML" (OOXML). A differenza delle versioni precedenti, che utilizzavano file binari denominati "Microsoft Document Format" (MCDFF) spesso utilizzati per nascondere dati segreti, il nuovo formato, tra le altre potenzialità, ha ridotto notevolmente il rischio di perdita di informazioni.

Nel presente lavoro, partendo dalla classificazione di "information hiding" proposta da Bauer, sono state analizzate diverse tecniche di steganografia, utili per nascondere dati nei file conformi allo standard OOXML, e sono state scoperte quattro nuove metodologie. La prima, "*Data Hiding by Different Compression Algorithm in ZIP*", si basa sulla caratteristica che i nuovi documenti MS-Office sono dei comuni file compressi. La seconda, "*Data Hiding by Office Macro*", per memorizzare messaggi segreti, utilizza le macro opportunamente modificate. La terza, "*Data Hiding by Zero Dimension Image*", trova il proprio punto di forza nel fatto che un'immagine risulta invisibile all'occhio umano se si impostano i valori della larghezza e dell'altezza entrambi uguali a zero. L'ultima, "*Data Hiding by Revision Identifier Value*", memorizza informazioni nascoste in alcuni attributi degli elementi XML (rsid). Tutti i metodi presentati possono essere combinati tra loro per aumentare la quantità di informazione che si può nascondere utilizzando un singolo "cover file".

L'analisi effettuata su un campione di circa 50.000 file di MS-Office, mostra come le tecniche avanzate possono essere utili in applicazioni reali. Sono stati poi classificati e confrontati tutti i metodi proposti, stimando la quantità di informazioni che può essere nascosta e valutando i limiti delle suddette tecniche. E' stato infine verificato che il "*Document Inspector*", strumento automatico introdotto da Microsoft per rilevare informazioni nascoste o dati personali, non individua alcuna anomalia in file cui si applicano tali metodologie.