ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Πρόγραμμα Μεταπτυχιακών Σπουδών
Τεχνοοικονομική Διοίκηση & Ασφάλεια Ψηφιακών Συστημάτων

Κατεύθυνση: **Ασφάλεια Ψηφιακών Συστημάτων**

Διπλωματική Εργασία

Μέθοδοι διερεύνησης, ανάκτησης και απαλοιφής μεταδεδομένων σε γνωστά λειτουργικά περιβάλλοντα

Χρήστος Αλέξης

Επιβλέπων: Επίκουρος Καθηγητής Χρήστος Ξενάκης

Πειραιάς, Απρίλιος 2014

[Κενή Σελίδα]

Περίληψη

Στην παρούσα διπλωματική εργασία γίνεται μελέτη-ανάλυση μεταδεδομένων διαφόρων αρχείων σε πλατφόρμες των Windows και Linux.

Αρχικά, προσδιορίζεται η έννοια των μεταδεδομένων και παρουσιάζεται η χρησιμότητά τους στην εγκληματολογική έρευνα.

Στη συνέχεια, αναλύονται τα είδη μεταδεδομένων σε προγράμματα της Adobe Reader's, του Open Office, αρχείων ήχου και εικόνας, κ.λπ., και το είδος των πληροφοριών που αποκαλύπτουν.

Επίσης, γίνεται αναφορά σε εργαλεία που υπάρχουν ενσωματωμένα σε διάφορες εφαρμογές, για την ανάκτηση καθώς και αφαίρεση μεταδεδομένων με παραδείγματα και οδηγίες.

Παρατίθενται μερικά από τα δημοφιλέστερα εμπορικά εργαλεία προβολής και απαλοιφής μεταδεδομένων, όπως επίσης και κάποια εργαλεία με άδεια ελεύθερης χρήσης.

Λέξεις κλειδιά: Linux, Windows, Μεταδεδομένα, Open Office, pdf, MP3, Digital Forensics.

Abstract

In the present thesis for the acquisition of a Master of Science Degree, we study the metadata in Microsoft Office, PDF and other types of files, in various operating systems, mainly the MS Windows and Linux.

Initially, we define the meaning of metadata and we present their usefulness in forensic analysis.

Subsequently, we analyze the types of metadata in programs like the Microsoft Office suite, Open Office, Adobe Reader and the kind of information they reveal.

Afterwards, we discuss incorporated tools, which exist in different applications, for the acquisition and removal of metadata.

Finally, we study the cases of metadata extraction and deletion, with the help of various commercial and open-source tools.

Keywords: Linux, Windows, Metadata, Digital Forensics, Open Office, PDF.

Ευχαριστίες

Η διπλωματική εργασία εκπονήθηκε στα πλαίσια του Μεταπτυχιακού Προγράμματος Σπουδών «Τεχνοοικονομική Διοίκηση & Ασφάλεια Ψηφιακών Συστημάτων», του Τμήματος Ψηφιακών Συστημάτων, του Πανεπιστημίου Πειραιώς, υπό την επίβλεψη του Επίκουρου Καθηγητή κ. Χρήστου Ξενάκη. Θα ήθελα να ευχαριστήσω θερμά τον κύριο Χρήστο Ξενάκη, για την ευκαιρία που μου έδωσε να ασχοληθώ με ένα τόσο ενδιαφέρον αντικείμενο, το οποίο ανταποκρίνεται απολύτως στα επιστημονικά μου ενδιαφέροντα, καθώς και για την αμέριστη συμπαράστασή του καθ' όλη την διάρκεια της εκπόνησης.

Ευχαριστώ θερμά τον κ. Χριστόφορο Νταντογιάν – επιστημονικό συνεργάτη του κ. Ξενάκη - για την αποτελεσματική συνεργασία, τις πολύτιμες υποδείξεις και τη συμβολή του στην ολοκλήρωση της παρούσας εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω από καρδιάς, τους εκλεκτούς συναδέλφους, συμφοιτητές και φίλους, Χριστόδουλο και Αντώνη, αφιερώνοντάς τους το παρακάτω απόφθεγμα του Νίκου Καζαντζάκη:

"Το πρώτο χρέος σου εκτελώντας τη θητεία σου στη ράτσα είναι να νιώσεις μέσα σου όλους τους προγόνους.

Το δεύτερο να φωτίσεις την ορμή τους και να συνεχίσεις το έργο τους.

Το τρίτο σου χρέος είναι να δώσεις στο γιό σου τη μεγάλη εντολή να σε ξεπεράσει"

[Κενή Σελίδα]

Πίνακας Περιεχομένων

Κατάλογος Εικόνων

Κατάλογος Πινάκων

1 Εισαγωγή

Αυτό που «μαρτυρά» ένα έγγραφο για το άτομο το οποίο το δημιούργησε, είναι – σχεδόν – τόσο σημαντικό όσο και ο επιδιωκόμενος σκοπός του ίδιου του εγγράφου.

Σε ένα έγγραφο με αδιαμφισβήτητα στοιχεία, είναι εξίσου σπουδαίο να μάθουμε ποιος και πότε το έγραψε.

Η εξαγωγή και μόνο του εγγράφου και η χρήση του ως πειστήριο εγκλήματος δεν επαρκούν για τη διεξαγωγή μιας λεπτομερούς ανάλυσης. Θα πρέπει – με κάποιον τρόπο – να υπάρχει ένας συνδετικός κρίκος μεταξύ πειστηρίου και υπόπτου. Ακριβώς στο σημείο αυτό υπεισέρχεται η ανάλυση ψηφιακών πειστηρίων, βασισμένη στα μεταδεδομένα.

Στην παρούσα εργασία αναλύονται τα εργαλεία τα οποία θα μας βοηθήσουν να στην ανάλυση διαφόρων εγγράφων, όπως του Microsoft Office και της Adobe Systems (αρχεία pdf), καθώς είναι τα πλέον χρησιμοποιούμενα αυτή τη στιγμή.

Περιγράφεται λεπτομερώς η διαδικασία εύρεσης των μεταδεδομένων τους, καθώς επίσης και η δυνατότητα απαλοιφής μέρους ή όλου αυτών.

2 Δομή συστημάτων αρχείων

2.1 Αρχεία (NTFS)

Τα δεδομένα που αποθηκεύονται στον σκληρό δίσκο, είναι διαιρεμένα σε αρχεία. Τα περισσότερα λειτουργικά συστήματα επιτρέπουν στον χρήστη να οργανώσει τα αρχεία του σε «δέντρα» με καταλόγους, όπου ένας κατάλογος μπορεί να περιέχει αρχεία ή άλλους καταλόγους. Έτσι, θα πρέπει να βρεθεί ένας τρόπος να οργανώσουμε τα αρχεία, ώστε να μπορούμε ανά πάσα στιγμή να τα ανακτούμε εύκολα και γρήγορα.

Υπάρχουν πολλοί τρόποι διάταξης των περιεχομένων ενός αρχείου στο σκληρό δίσκο. Έτσι μπορούμε να χρησιμοποιήσουμε:

- Συνεχόμενη διάταξη
- Διάταξη συνδεδεμένης λίστας
- Διάταξη με δείκτη
- Διάταξη με i-κόμβους (i-node)

Στην περίπτωση της συνεχόμενης διάταξης, το αρχείο βρίσκεται σε μία ακολουθία από clusters.

Οι περισσότερες από τις αδυναμίες του FAT προέρχονται από την απλοϊκή και ξεπερασμένη αρχιτεκτονική που ακολουθούσε. Κατά την σχεδίασή του, δεν είχαν προβλεφθεί δομές για στοιχεία ασφάλειας και αξιοπιστίας και φυσικά ήταν απαγορευτικά δύσκολο να προστεθούν εκ των υστέρων τέτοιες δυνατότητες. Η αρχιτεκτονική του NTFS όχι μόνο περιλαμβάνει τέτοιου είδους δυνατότητες, αλλά χρησιμοποιεί απλές θεμελιώδεις αρχές, που κάνει εύκολο να προστεθούν νέα χαρακτηριστικά στο μέλλον με μικρές αλλαγές.

Η πολύ καλή αρχιτεκτονική που χρησιμοποιεί το NTFS μπορεί να φανεί και από τον τρόπο που αποθηκεύονται οι πληροφορίες στο δίσκο. Κάθε δομή στο NTFS είναι ένα αρχείο, συμπεριλαμβανομένου και των δομών διαχείρισης των partitions, διατήρησης στατιστικών και των πληροφοριών ελέγχου του ίδιου του partition.

Οι πληροφορίες ελέγχου αποθηκεύονται σε «ειδικά» αρχεία που δημιουργούνται όταν το NTFS δημιουργείται. Αυτά τα αρχεία ονομάζονται *Αρχεία Μεταδεδομένων (MetaData Files)* και περιλαμβάνουν τις λίστες των αρχείων του partition, πληροφορίες του τόμου του δίσκου, τις διατάξεις των αρχείων κ.λ.π. Η μοναδική εξαίρεση στη φιλοσοφία «τα πάντα είναι ένα αρχείο» είναι ο Αρχικός Τομέας του Δίσκου (Volume Boot Sector), ο οποίος προηγείται των υπολοίπων αρχείων μεταδεδομένων και ελέγχει τις πιο βασικές λειτουργίες του NTFS, όπως να «φορτώσει» το λειτουργικό σύστημα. Το ίδιο απλό θεμελιώδες μοντέλο χρησιμοποιείται και για τα αρχεία και την διαχείριση αυτών σε εσωτερικό επίπεδο. (1)

2.1.1 Αρχεία Μεταδεδομένων και Πίνακας Κυρίου Αρχείου (MFT)

Το NTFS αποθηκεύει όλα τα δεδομένα, δεδομένα του χρήστη και δεδομένα εσωτερικής διαχείρισης, σε μορφή αρχείων. Τα πιο σημαντικά από αυτά είναι μία ομάδα από ειδικά αρχεία συστήματος τα οποία λέγονται επίσης και *Αρχεία Μεταδεδομένων*. Τα μεταδεδομένα είναι αρχεία που περιέχουν δεδομένα για τα δεδομένα, δηλαδή περιέχουν πληροφορίες (δεδομένα), για τα «πραγματικά» δεδομένα που αποθηκεύονται από το NTFS.

Τα αρχεία μεταδεδομένων δημιουργούνται αυτόματα από το σύστημα όταν «μορφοποιείται» (formatted) ένας δίσκος σε NTFS, και τοποθετούνται στην αρχή του partition. Για να γίνει αντιληπτή η τοποθεσία αυτών των αρχείων, θα πρέπει να εισαχθεί και μία άλλη θεμελιώδης δομή, ο Master File Table (MFT).

Ο MFT είναι ένα αρχείο μεταδεδομένων, που περιέχει περιγραφές άλλων αρχείων μεταδεδομένων και σε μερικές περιπτώσεις ολόκληρα αρχεία μεταδεδομένων.

Παρακάτω, φαίνεται ο πίνακας του MFT, όπου αναφέρονται οι πρώτες 16 εγγραφές του, δηλαδή τα πρώτα 16 αρχεία. (2)

Το σύστημα αποθηκεύει τα αρχεία ή τους καταλόγους στις εγγραφές του MFT, δίνοντάς τους την μορφή χαρακτηριστικών. Επειδή όμως το μέγεθος της εγγραφής του MFT είναι περιορισμένο, υπάρχουν διάφοροι τρόποι που το NTFS αποθηκεύει τα χαρακτηριστικά ενός αρχείου. Έτσι, μπορεί να τα αποθηκεύσει είτε ως *resident* χαρακτηριστικά, δηλαδή ότι όλα *τα χαρακτηριστικά θα αποθηκευθούν μέσα στην εγγραφή του MFT*, είτε ως *non-resident*, δηλαδή ότι *τα χαρακτηριστικά θα αποθηκευθούν σε άλλες εγγραφές του MFT ή σε επεκτάσεις που βρίσκονται εκτός MFT*.

Στο NTFS, δεν υπάρχει διαφοροποίηση ανάμεσα στα δεδομένα του αρχείου και στα χαρακτηριστικά που περιγράφουν το αρχείο. Το γεγονός αυτό είχε μια πολύ ενδιαφέρουσα εφαρμογή κατά την δημιουργία μικρών αρχείων. Εάν ο χώρος που χρειάζεται το αρχείο για *όλα* τα χαρακτηριστικά του αρχείου, συμπεριλαμβανομένου και των (μετα)δεδομένων του, είναι μικρότερο από το μέγεθος του MFT, τότε το αρχείο θα αποθηκευθεί μέσα στο MFT. Τέτοια αρχεία δεν καταλαμβάνουν επιπλέον χώρο στο δίσκο και δεν απαιτείται ξεχωριστή διαδικασία ελέγχου του MFT και ανάγνωσης από το δίσκο, πετυχαίνοντας έτσι καλύτερη απόδοση. (3)

Αρχείο Μετα-Δεδομένων	Όνομα Αρχείου	Εγγραφή MFT	Περιγραφή
Master file table (MFT)	$Mft	0	Περιέχει βασικές πληροφορίες για το ίδιο το MFT, όπως το μέγεθός του.
Master file table 2	$MftMirr	1	Αντίγραφο ασφαλείας της πρώτης εγγραφής του MFT. Αυτό συμβαίνει λόγω της σημαντικότητας των περιεχομένων.
Log file	$LogFile	2	Περιέχει μία λίστα από βήματα που χρησιμοποιεί το NTFS για ανάκτηση του συστήματος σε περίπτωσης αστοχίας του συστήματος. Το μέγεθος τουLog file εξαρτάται από το μέγεθος του δίσκου και μπορεί να είναι μεγαλύτερο από τις υπόλοιπες εγγραφές.
Volume	$Volume	3	Περιέχει πληροφορίες του δίσκου, όπως την ετικέτα του τόμου του δίσκου και την έκδοση του τόμου του δίσκου.
Attribute definitions	$AttrDef	4	Πίνακας που περιέχει ορισμούς των χαρακτηριστικών.
Root file name index	$	5	Είναι ο δείκτης προς τον ριζικό κατάλογο.
Cluster bitmap	$Bitmap	6	Περιέχει έναν «χάρτη» του δίσκου, που δείχνει ποια cluster χρησιμοποιούνται και ποια είναι διαθέσιμα.
Boot sector	$Boot	7	Περιέχει τον Αρχικό Τομέα του Δίσκου ή δείκτη προς αυτόν.
Bad cluster file	$BadClus	8	Περιέχει μία λίστα με τα BAD cluster του δίσκου (δηλαδή όσα έχουν εντοπιστεί ότι παρουσιάζουν σφάλματα).
Security file	$Secure	9	Περιέχει δικαιώματα χρήσης όλων των αρχείων του δίσκου
pcase table	$Upcase	10	Περιέχει πληροφορίες μετατροπής των ονομάτων των αρχείων σε Unicode (16-bit). Χρησιμοποιείται όταν βάζουμε ένα αρχείο σε κάποιο κατάλογο, αλλά μόνο από το NTFS 5.0 και μεταγενέστερες εκδόσεις του.
NTFS extension file	$Extend	11	Χρησιμοποιείται για επιπλέον λειτουργίες του NTFS, όπως είναι τα Quota, τα Reparse Point (δηλαδή ειδικές λειτουργίες συνεργασίας του συστήματος αρχείων με τα αρχεία και τους φακέλους), κ.α.
		12–15	Ο χώρος αυτός έχει δεσμευτεί για μελλοντική χρήση

Πίνακας 1ος: Οι αρχικές εγγραφές του MFT

2.2 Αρχεία Ext4

Το Ext4 είναι το προεπιλεγμένο σύστημα αρχείων για τις νέες εγκαταστάσεις πολλών δημοφιλών διανομών Linux. Ενώ υπάρχουν αρκετές αναφορές που τεκμηριώνουν το σχεδιασμό του Ext4 ή επισημαίνουν ορισμένα χαρακτηριστικά, από την οπτική των προγραμματιστών των αρχείων συστήματος, ωστόσο, υπάρχει έλλειψη πληροφοριών αναφορικά με το κομμάτι της ψηφιακής εγκληματολογίας (digital forensics). (4)

Στις επόμενες παραγράφους αναφέρονται ορισμένες αλλαγές στην τοπολογία του συστήματος αρχείων και κάποιες σημαντικές δομές δεδομένων, όπως, για παράδειγμα, η υπερ-ομάδα συστήματος αρχείων (*file system superblock*), η απεικόνιση δομών δεδομένων (*inode data mapping*), κ.λπ.

Το *inode* είναι μία δομή δεδομένων, η οποία διατηρεί πληροφορίες για αρχεία, σε συστήματα αρχείων όπως είναι το Linux. Αποθηκεύει πληροφορίες για μεταδεδομένα σχετικά με ένα αρχείο, όπως αποτυπώσεις ημερομηνίας και ώρας (*timestamps*), δικαιώματα χρηστών και ομάδας, καθώς και δείκτες σε μπλοκ δεδομένων. Αφού κάθε αρχείο πρέπει να έχει ένα μοναδικό *inode*, ο αριθμός των αρχείων που ένα σύστημα μπορεί να υποστηρίξει περιορίζεται, όχι μόνο από το ποσό των ελεύθερων μπλοκ δεδομένων, αλλά και από το συνολικό αριθμό των *inodes*.

Το Ext4 δεσμεύσει κάποια *inodes* για συγκεκριμένες ενέργειες, όπως φαίνεται και στον παρακάτω πίνακα:

Inode	Σκοπός
0	(Δεν υπάρχει)
1	Λίστα ελαττωματικών *blocks*
2	Ριζικός κατάλογος (*Root Directory)*
3	Αναλογία χρήστη (*User quota)*
4	Αναλογία ομάδας (*Group quota)*
5	Φόρτωση εκκίνησης (*Boot loader*)
6	Επαναφορά διαγραμμένων αρχείων καταλόγου (*Undelete directory*)
7	Δεσμευμένη ομάδα περιγραφέων (*Reserved Group Descriptors*)
8	Ημερολόγιο (*Journal inode*)
9	"Exclude" inode
10	Αντίγραφο (*Replica inode*)
11	(Συνήθως) Κατάλογος που χάθηκε και βρέθηκε

Πίνακας 2[05]: Δομή δεδομένων inode

Το ExtX, και πιο συγκεκριμένα, το *superblock* του ExtX, αποτελεί μία αποθηκευτική μονάδα για τα δεδομένα σε ολόκληρο το σύστημα αρχείων. Εμπεριέχει πληροφορίες για το συνολικό αριθμό των blocks και των inodes στο σύστημα αρχείων, τον αριθμό των διαθέσιμων blocks στο σύστημα αρχείων, καθώς, επίσης και το πρώτο διαθέσιμο *inode*.

Ένας περιγραφέας ομάδας (*group descriptor)* περιέχει μετα-πληροφορίες σχετικά με μία συγκεκριμένη ομάδα μπλοκ, όπως το φάσμα των *inodes* στην ομάδα, τα μπλοκ

(blocks) που περιλαμβάνονται στην ομάδα (*groups*) και τη μετατόπιση των βασικών μπλοκ στην ομάδα.

2.2.1 Μεταδεδομένα στο Ext4

Σε αντίθεση με τα μεταδεδομένα στο JFS (σύστημα αρχείων 64-bit στο Linux), το Ext4 έχει αρκετά στατικές δομές μεταδεδομένων του συστήματος αρχείων. Τα inodes αποθηκεύονται σε πίνακες παρόμοιους με αυτούς του ExtX και όχι σε ένα ειδικό αρχείο. Η τοποθέτηση των δομών, όπως οι πίνακες των inodes, τα λογότυπα, τα bitmaps και οι περιγραφείς ομάδων, ορίζονται όταν δημιουργείται μια ευέλικτη ομάδα (flex group). Ωστόσο, το σύστημα αρχείων είναι ικανό για περιορισμένη ανάπτυξη λόγω της χρήσης περιγραφέων ομάδων αυξητικών μπλοκ (*group descriptor growth blocks*).
Παρά το γεγονός ότι το inode παραμένει ο κύριος αποθηκευτικός χώρος για τα μεταδεδομένα αρχείων, τα μεταδεδομένα μπορούν να αναμιχθούν με κανονικά μπλοκ δεδομένων, όπως ακριβώς και στο ExtX, όταν δημιουργούνται τα λεγόμενα δένδρα έκτασης (*extent trees*).

Ωστόσο, η χρήση των κεφαλίδων έκτασης (*extent headers)* μπορεί να λειτουργήσει προς όφελος των προσπαθειών ανάκτησης δεδομένων, σε περίπτωση «σκαλίσματος» μεταδεδομένων. (5)

3 Μεταδεδομένα

Τα μεταδεδομένα βρίσκονται στο επίκεντρο του ενδιαφέροντος το τελευταίο διάστημα. Αποδεικνύεται άλλωστε και από τη φράση: *«Don't worry, we don't collect your "Data" we only collect "MetaData"»,* την οποία επανέλαβε δημόσια ο Πρόεδρος των Η.Π.Α Barack Obama, αναπαράγοντας την φράση κάποιων υψηλόβαθμων στελεχών της NSA, ενώ είχε ξεσπάσει το σκάνδαλο Snowden, στα μέσα του Ιουνίου 2013 που αφορούσε σε τηλεφωνικές παρακολουθήσεις.

Από τη στιγμή που έγινε αντιληπτό ότι υπάρχει κάτι περισσότερο σε ένα αρχείο, πέρα από την πληροφορία που είναι ορατή στον κοινό χρήστη, οι ερευνητές επέδειξαν ιδιαίτερο ενδιαφέρον να λάβουν γνώση αυτής της «επιπλέον πληροφορίας», προκειμένου να τη χρησιμοποιήσουν στις έρευνές τους.

Με απλά λόγια, τα μεταδεδομένα είναι δεδομένα που αναφέρονται σε δεδομένα. Είναι πληροφορία που περιγράφει ή τοποθετεί τα δεδομένα σε γενικότερο πλαίσιο, χωρίς να είναι κομμάτι των δεδομένων που αποτελεί το πρωταρχικό σημείο εστίασης του χρήστη, π.χ., το γραφικό μέρος μιας JPEG φωτογραφίας. (6)

Τα μεταδεδομένα είναι κατ' ανάγκη διαφορετικά για πολλά ατομικά πληροφοριακά πεδία (fields) και τομείς (domains). Έτσι, τα μεταδεδομένα αρχείων

είναι διαφορετικά από αυτά που βρίσκονται στις ιστοσελίδες, αλλά και τα δύο περιγράφουν, ως ένα σημείο, τα χαρακτηριστικά των δεδομένων που αναπαριστούν. Για παράδειγμα, η αποτύπωση της ώρας (time-stamp) σε μία φωτογραφία αποτελεί μέρος των μεταδεδομένων, αφού μας μαρτυρά την ώρα που αυτή ελήφθη.

Σε λειτουργικό περιβάλλον των MS Windows, τα μεταδεδομένα διαιρούνται σε δύο κατηγορίες: α) στα μεταδεδομένα συστήματος αρχείων και β) στα μεταδεδομένα εφαρμογών.

3.1 Σύστημα Αρχείων (File System Metadata)

Πρέπει να γίνει σαφής η διάκριση μεταξύ των μεταδεδομένων συστήματος αρχείων, τα οποία φιλοξενούν τα διαχειριστικά δεδομένα του συστήματος αρχείων, και των μεταδεδομένων αρχείων, τα οποία περιέχουν πληροφορίες σχετικά με τα περιεχόμενα αρχείων, χωρίς όμως να είναι στην πραγματικότητα το περιεχόμενο του αρχείου.

Παρόλο που τα μεταδεδομένα συστήματος αρχείων, όπως τα δικαιώματα των αρχείων, η κατάσταση αρχείου (τα ενεργά σε αντίθεση με τα διαγραμμένα) και οι πληροφορίες σχετικά με το εάν ένα αρχείο είναι παραμένον (resident file) ή μη, μπορεί να είναι χρήσιμα σε σωστό πλαίσιο εφαρμογής.

Η πτυχή των μεταδεδομένων που συνήθως τραβά περισσότερο την προσοχή είναι οι πληροφορίες σχετικά με την αποτύπωση ημερομηνίας και ώρας (date-time stamp). Το περιβάλλον των Windows, αλλά και τα περισσότερα εργαλεία ψηφιακών πειστηρίων, αποτυπώνουν την ημερομηνία και την ώρα που υπάρχει μέσα στο χώρο του SIA (Standard Information Attribute) και του MFT (Master File Table), κατά κύριο λόγο, γιατί αυτές είναι οι αποτυπώσεις οι οποίες ενημερώνονται όταν ένα αρχείο ή ένας φάκελος αντιγράφεται, μετακινείται ή εγγράφεται (σε αντίθεση με την αποτύπωση ημερομηνίας και ώρας που υπάρχει στην εγγραφή FNA, η οποία καταχωρείται όταν το αρχείο πρωτο-δημιουργείται στον τόμο και γενικότερα, ύστερα από αυτό δεν μεταβάλλεται). (7)

Η συμπεριφορά των Windows, σχετικά με την αποτύπωση ημερομηνίας και ώρας, μπορεί να ποικίλλει κατά πολύ, βασισμένη στην ακριβή ενέργεια που λαμβάνει χώρα σε ένα αρχείο ή φάκελο. Για παράδειγμα, μετακινώντας ένα αρχείο από έναν τόμο σε έναν άλλον, κάνοντας χρήση της δυνατότητας drag-and-drop ή μέσω της γραμμής εντολών, θα επιτευχθεί η ενημέρωση της ημερομηνίας δημιουργίας του αρχείου. Αντίθετα, εάν επιλεγεί η ενέργεια του Cut-and-Paste, το αρχείο δε θα μεταβάλει την ημερομηνία δημιουργίας του. Επίσης, πρέπει να ληφθούν υπόψη ειδικές περιπτώσεις που επηρεάζουν την αποτύπωση ημερομηνίας και ώρας. Έτσι, τα αρχεία τα οποία εξάγονται (extract) από ένα αρχείο, (π.χ., αρχεία .ZIP ή .RAR), μπορεί να φέρουν μαζί τους και αποτυπώσεις ημερομηνίας και ώρας από το σύστημα στο οποίο πρωτο-αρχειοτέθησαν.

Έχοντας κατά νου όλα τα προηγούμενα, η αποτύπωση ημερομηνίας και ώρας στην εγγραφή (record) του SIA έχει την εξής σημασία:

3.1.1 Δημιουργία αρχείου (File Created)

Αυτή η αποτύπωση ημερομηνίας και ώρας, συνήθως δηλώνει πότε δημιουργήθηκε ένας φάκελος ή ένα αρχείο και αποτελεί μία από τις πιο πολύτιμες πληροφορίες για τους περισσότερους ερευνητές. Για παράδειγμα, όταν ένα υπάρχον αρχείο αντιγράφεται (copy), η αποτύπωση ημερομηνίας και ώρας δημιουργίας του νέου αρχείου ορίζεται στην τρέχουσα ώρα (το νέο αντίγραφο διατηρεί την αποτύπωση ημερομηνίας και ώρας της τελευταίας εγγραφής (Last Written) και της τροποποίησης καταχώρησης (Entry Modified), του αρχικού αρχείου.

Ομοίως, εάν ένα αρχείο μετακινηθεί (Move) σε έναν διαφορετικό τόμο, κάνοντας χρήση της γραμμής εργαλείων (command line) ή της λειτουργίας drag-and-drop, η αποτύπωση ημερομηνίας και ώρας του File Created του νέου αντίγραφου (New Copy) τίθεται στην τρέχουσα ώρα (μετακινώντας ένα αρχείο εντός του ίδιου τόμου δεν μεταβάλλεται η αποτύπωση ημερομηνίας και ώρας του File Created).

Ωστόσο, εάν ένα αρχείο μετακινηθεί σε διαφορετικό τόμο, κάνοντας χρήση της λειτουργίας Cut-and-Paste, η αποτύπωση ημερομηνίας και ώρας του File Created παραμένει αμετάβλητη (οι αποτυπώσεις ημερομηνίας και ώρας στο Last Accessed και Entry Modified είναι πολύ πιθανό να μεταβληθούν) .

3.1.2 Τροποποίηση (Modified ή Last Written)

Αυτή η αποτύπωση ημερομηνίας και ώρας αναπαριστά την τελευταία φορά που μεταβλήθηκε η ιδιότητα $DATA ενός αρχείου. Για παράδειγμα, αν ένας χρήστης άνοιξε ένα .BAT αρχείο, επεξεργάσθηκε το περιεχόμενο και ξανα-αποθήκευσε (re-save) το αρχείο, ο χρόνος της Last Written – κατά πάσα πιθανότητα – θα ενημερωθεί (update) επίσης. Αντιθέτως, εάν ο χρήστης άνοιγε το ίδιο αρχείο, διάβαζε το περιεχόμενο και έκλεινε το αρχείο, χωρίς να κάνει κάποια αλλαγή, ο χρόνος στην αποτύπωση Last Written δε θα ενημερωνόταν.

3.1.3 Τελευταία προσπέλαση (Last Accessed)

Η αποτύπωση ημερομηνίας και ώρας αναπαριστά τον πιο πρόσφατο χρόνο, που ένας φάκελος ή αρχείο προσπελάσθηκε από το σύστημα. Αυτή η αποτύπωση ημερομηνίας και ώρας δεν αναπαριστά αναγκαία ότι ένα αρχείο ανοίχθηκε από το χρήστη. Και μόνο η τοποθέτηση του mouse πάνω από το όνομα του αρχείου, στον Windows Explorer, μπορεί να ενημερώσει το χρόνο της τελευταίας προσέγγισης. Επίσης, αν και η ενημέρωση της αποτύπωσης ημερομηνίας και ώρας *μπορεί* να είναι αποτέλεσμα μιας πράξης του χρήστη (π.χ., να ανοίξει και να διαβάσει το αρχείο), θα μπορούσε, επίσης, να είναι το αποτέλεσμα των αυτόματων, αβλαβών ενεργειών ενός συστήματος (π.χ., σάρωση ενός antivirus). Άρα, η πραγματική τιμή της χρονικής αποτύπωσης Last Accessed θα πρέπει να αξιολογηθεί προσεχτικά, σε συνάρτηση και με άλλα στοιχεία του συστήματος αρχείων. Επίσης, το σύστημα NTFS καθυστερεί να

ενημερώσει την χρονική αποτύπωση Last Accessed κατά μία ώρα, περίπου, από τη στιγμή που το αρχείο προσπελάσθηκε.

3.1.4 Τροποποίηση (SIA Modified ή Entry Modified)

Αυτή η αποτύπωση ημερομηνίας και ώρας αναπαριστά την τελευταία φορά που τροποποιήθηκε κάποια ιδιότητα στον MFT ενός αρχείου ή φακέλου. Λόγοι μιας τέτοιας ενημέρωσης μπορεί να αποτελέσουν: η αλλαγή τοποθεσίας ενός αρχείου σε ένα δίσκο, προσθήκη μίας επιπλέον ροής δεδομένων (data stream) σε ένα αρχείο, αλλαγή στο όνομα ενός αρχείου.

3.3 Εφαρμογές (Application metadata)

Αντίθετα με τα μεταδεδομένα συστήματος αρχείων, τα μεταδεδομένα εφαρμογών βρίσκονται μέσα σε αρχεία στα οποία αναφέρονται (όπως τα αρχεία του Open Office, τα αρχεία PDF, τις ψηφιακές φωτογραφίες, κ.λπ.). Αυτού του είδους τα μεταδεδομένα μπορούν να δώσουν πολύτιμες πληροφορίες στους ερευνητές ψηφιακών πειστηρίων.

Τα μεταδεδομένα εφαρμογής δημιουργούνται αυτόματα από το λογισμικό εφαρμογής και ενσωματώνονται σε κάθε αρχείο που δημιουργείται ή υφίσταται επεξεργασία από αυτό το λογισμικό. Τα λειτουργικά συστήματα που ελέγχουν προσωπικούς υπολογιστές, εξυπηρετητές (servers) και συστήματα επικοινωνίας δημιουργούν μεταδεδομένα συστήματος (system metadata), τα οποία αποδίδουν πεδία πίνακα καταχώρησης αρχείων (όνομα αρχείου, δημιουργία, μέγεθος και χρήση) σε όλα τα αρχεία που αποθηκεύονται στο σύστημα ώστε το λειτουργικό σύστημα να τα ταυτοποιεί και εντοπίζει για μελλοντική χρήση. Τα μεταδεδομένα συστήματος βρίσκονται στο μητρώο (registry) συστήματος του υπολογιστή ή server που χρησιμοποιήθηκε για να προσπελάσει και να αποθηκεύσει το αρχείο.

Τα μεταδεδομένα αποκαλούνται το «ηλεκτρονικό ισοδύναμο του DNA» και έχουν την ικανότητα να ρίξουν φως στην προέλευση, στο περιεχόμενο, στην αυθεντικότητα και στο διαμοιρασμό ηλεκτρονικών στοιχείων.

Αν και είναι πολύ χρήσιμα για τη συνεργασία μεταξύ πολλών ανθρώπων, ένα πρόβλημα είναι ότι πολλοί χρήστες δεν είναι καλά ενημερωμένοι για το ποια πληροφορία αποθηκεύεται στα έγγραφά τους όταν τα κοινοποιούν και τα διαμοιράζονται. Αν δεν αφαιρεθούν στοχευμένα από το χρήστη, τα μεταδεδομένα σε ένα έγγραφο, φύλλο εργασίας ή παρουσίαση του Microsoft Office ή του OpenOffice.org, θα αποθηκευτούν ταυτόχρονα με κάθε κείμενο που δημιουργεί ο χρήστης. Επομένως, αυτός ο τύπος δεδομένων μπορεί να αποκαλύψει ακούσια ευαίσθητες πληροφορίες, αλλά και να χρησιμεύσει στην ψηφιακή εξέταση υπολογιστών.

Τα μεταδεδομένα μπορούν να βοηθήσουν στην ταυτοποίηση ανθρωπίνων ή συστημικών ενεργειών σε πληροφοριακά συστήματα και μπορούν να χρησιμοποιηθούν για την έρευνα και εξακρίβωση απάτης, κατάχρησης, λαθών ή αποτυχιών συστήματος και να βοηθήσουν στην στοιχειοθέτηση κατηγορητηρίων, στη

χρονολόγηση και εκτίμηση του εύρους γνώσης, ή ως τεκμήριο αθωότητας, που αποτελούν υποθέσεις εγκληματικών ή πολιτικών δικαστικών αγωγών. Το δικαστικό σύστημα συνολικά αναγνωρίζει τη χρησιμότητα των μεταδεδομένων και την βιωσιμότητα και αποδοχή τους ως στοιχεία.

Τα μεταδεδομένα που περιέχονται σε έγγραφα μπορεί να είναι σημαντική πηγή πληροφορίας για τους ερευνητές. Παρόλα αυτά, η εγκληματολογική ανάλυση υπολογιστών και των δεδομένων που περιέχουν, μπορεί να είναι μια εξαιρετικά δύσκολη και χρονοβόρα διαδικασία.

4 Διερεύνηση εγγράφων

Τα έγγραφα είναι αναμφισβήτητα μία από τις πιο σημαντικές περιοχές, όπου μπορούμε να βρούμε μεταδεδομένα. Ο γρήγορα αναπτυσσόμενος τομέας της ηλεκτρονικής ανακάλυψης, έχει αποδειχθεί «χρυσορυχείο» μεταδεδομένων.

Η παρακάτω λίστα περιγράφει τους βασικούς τύπους μεταδεδομένων, τα οποία βρίσκονται σε ένα κοινό αρχείο (π.χ., αρχείο pdf). (8)

- **Δημιουργός**: Ανεξάρτητα αν αυτή η πληροφορία προέρχεται από το λειτουργικό σύστημα ή την ίδια την εφαρμογή, ένα όνομα – πάντοτε – ενσωματώνεται ως μέρος του αρχείου, προκειμένου να γίνει από όλους ορατό.

- **Φορέας ή Οργανισμός**: Σε αυτήν την περίπτωση, η ενσωμάτωση τέτοιου είδους πληροφορίας λαμβάνει χώρα όταν καταχωρίζεται ένα όνομα (του φορέα ή οργανισμού) κατά την εγκατάσταση του λειτουργικού συστήματος.

- **Αναθεωρήσεις**: Ως μέρος του ημερολογίου καταχώρησης (*revision log*), μπορεί να καταχωρηθούν οι προηγούμενοι δημιουργοί, όπως επίσης και η διαδρομή όπου αποθηκεύθηκε το αρχείο.

- **Προγενέστεροι δημιουργοί:** Τα έγγραφα, συνήθως, έχουν ένα ιστορικό από χρήστες, οι οποίοι επεξεργάστηκαν με οποιονδήποτε τρόπο το έγγραφο αυτό.

- **Πρότυπο (*Template*):** Αυτό το κομμάτι δεδομένων αναφέρεται στο πιο πρότυπο είναι ενσωματωμένο μέσα στο έγγραφο.

- **Όνομα υπολογιστή:** Αυτό το όνομα συνδέει το έγγραφο με τον υπολογιστή στον οποίο γράφτηκε.

- **Σκληρός δίσκος:** Αυτά τα δεδομένα συμπεριλαμβάνουν το όνομα του σκληρού δίσκου και τη διαδρομή όπου εγκαταστάθηκε το αρχείο.

- **Εξυπηρετητής δικτύου:** Μία επέκταση των πληροφοριών του σκληρού δίσκου. Αν ένα αρχείο αποθηκευθεί σε έναν *server* δικτύου, τα μεταδεδομένα ανακλούν το όνομα διαδρομής του δικτύου.

- **Χρόνος:** Αυτός ο τύπος μεταδεδομένων (συχνά) υποδεικνύει το χρονικό διάστημα που το έγγραφο ήταν ανοιχτό προς επεξεργασία.

- **Διαγραμμένο κείμενο:** Ορισμένα μεταδεδομένα καταχωρίζουν κείμενο το οποίο έχει διαγραφεί.

- **Αντικείμενα αντικειμενοστραφούς προγραμματισμού (*Visual Basic Objects*):** Αντικείμενα που χρησιμοποιούνται και δημιουργούνται από τη VB, είναι συνήθως κομμάτι μακρο-εκτέλεσης και αποθηκεύονται και αποκρύπτονται από το χρήστη.

- **Αποτύπωση ημερομηνίας και ώρας (*Timestamps*):** Αυτός ο τύπος δεδομένων βασίζεται συνήθως στην αποτύπωση ημερομηνίας και ώρας του λειτουργικού συστήματος και καλύπτει τις εξής αποτυπώσεις: Δημιουργία, Πρόσβαση και Τροποποίηση (εγγράφου).

- **Εκτύπωση:** Τα μεταδεδομένα συνήθως μαρτυρούν την τελευταία εκτύπωση ενός εγγράφου.

4.1 Αρχεία σύνδεσης και ενσωμάτωσης αντικειμένων (OLE Files)

Αν και μπορεί να μην είναι ευρέως γνωστός ο τύπος αυτού του αρχείου, ωστόσο, είναι σε όλους γνωστά τα αρχεία τα οποία κάνουν χρήση αυτού του προτύπου.

Τα αρχεία αυτά είναι τα, γνωστά σε όλους, αρχεία του Microsoft Office 1997-2003 (Αρχεία doc, παρουσιάσεις σε Power point, φύλλα εργασίας σε Excel).

Τα αρχεία OLE είναι πραγματικά μικροσκοπικά, αποκλειστικής λειτουργίας, φορητά συστήματα αρχείων. Όπως τα παραδοσιακά συστήματα αρχείων, περιέχουν δεδομένα με δομημένο τρόπο και, επίσης, περιέχουν μεταδεδομένα.

Τα αρχεία OLE έχουν δύο κυρίως αποθηκευτικές σημασίες: αντικείμενα αποθήκευσης (*storage objects*) και αντικείμενα ροής (*stream objects*). Τα αντικείμενα αποθήκευσης εκτελούν τις ίδιες λειτουργίες όπως ο κατάλογος σε ένα τυπικό σύστημα αρχείων. Έτσι, όπως ένας κατάλογος, μπορεί να περιέχει επιπλέον

αντικείμενα ελέγχου, τα οποία λειτουργούν ως υποκατάλογοι. Τα αντικείμενα ροής είναι ακολουθίες από τομείς (*sectors*) δεσμευμένοι για ένα διακριτό κομμάτι δεδομένων. Έτσι, σε ένα σύστημα αρχείων, τύπου OLE, οι ροές (*streams*) αναλαμβάνουν το ρόλο των αρχείων. (9)

Ένα αρχείο OLE, αποτελείται από ένα αντικείμενο αποθήκευσης ρίζας (*root storage object*), παρόμοιο με τον κατάλογο ρίζας (*root directory*) και τουλάχιστον ένα αντικείμενο ροής, το οποίο αναπαριστά τα προεπιλεγμένα δεδομένα για το αρχείο.

Για παράδειγμα, σε ένα έγγραφο Word 2003, αυτό το προεπιλεγμένο αντικείμενο ροής, θα περιέχει την πλειονότητα του αληθινού περιεχομένου αρχείου. Πέραν αυτού του αντικείμενου ροής, το αντικείμενο αποθήκευσης μπορεί να εμπεριέχει οποιονδήποτε αριθμό επιπρόσθετων αντικειμένων αποθήκευσης, κάθε ένα από τα οποία μπορεί να περιέχει μία ή περισσότερες ροές (*streams*).

Περίπτωση 1η: BTK (Bind-Torture-Kill)

Ο Dennis Rader (γνωστός ως BTK) στοίχισε τη ζωή σε 10 θύματα, στο διάστημα από το 1974 μέχρι το 1991 και διέφευγε τη σύλληψη για 30 χρόνια. Τελικά, ήταν τα ενσωματωμένα μεταδεδομένα σε ένα έγγραφο του Word, αυτά που οδήγησαν στη σύλληψή του. (10)

Η τελευταία επικοινωνία του δολοφόνου BTK με τα μέσα ενημέρωσης και την αστυνομία ήταν ένας φάκελος, ο οποίος έφτασε σε έναν τηλεοπτικό σταθμό, στην περιοχή Wichita, στις 16 Φεβρουαρίου 2005. Μία μοβ, 1.44-MB Memorex δισκέτα εσωκλειόταν στη συσκευασία. Η αστυνομία βρήκε – εντός δισκέτας – μεταδεδομένα που ήταν ενσωματωμένα σε ένα έγγραφο του Microsoft Word, που οδηγούσε στη «Λουθηρανική Εκκλησία του Χριστού». Σύμφωνα με τα ίδια μεταδεδομένα, η τελευταία τροποποίηση του εγγράφου είχε γίνει από κάποιο άτομο, ονόματι *Davis*. Μία απλή αναζήτηση στην ηλεκτρονική τοποθεσία της εκκλησίας, εμφάνισε τον *Dennis Rader* ως πρόεδρο του συμβουλίου εκκλησιάσματος. [2].

Περίπτωση 2η: Αρχείο «Iraq.doc»

Ίσως ορισμένες φορές να είναι σημαντικό να γνωρίζουμε ποιος είναι ο προγενέστερος συγγραφέας ή ακόμα αν κάποιοι άνθρωποι έχουν τροποποιήσει ένα έγγραφο. Η Βρετανική Κυβέρνηση έμαθε με σκληρό τρόπο πόσες πληροφορίες μπορούν να προσφέρουν τα μεταδεδομένα και γιατί αυτές οι πληροφορίες μπορεί να είναι σημαντικές. (11)

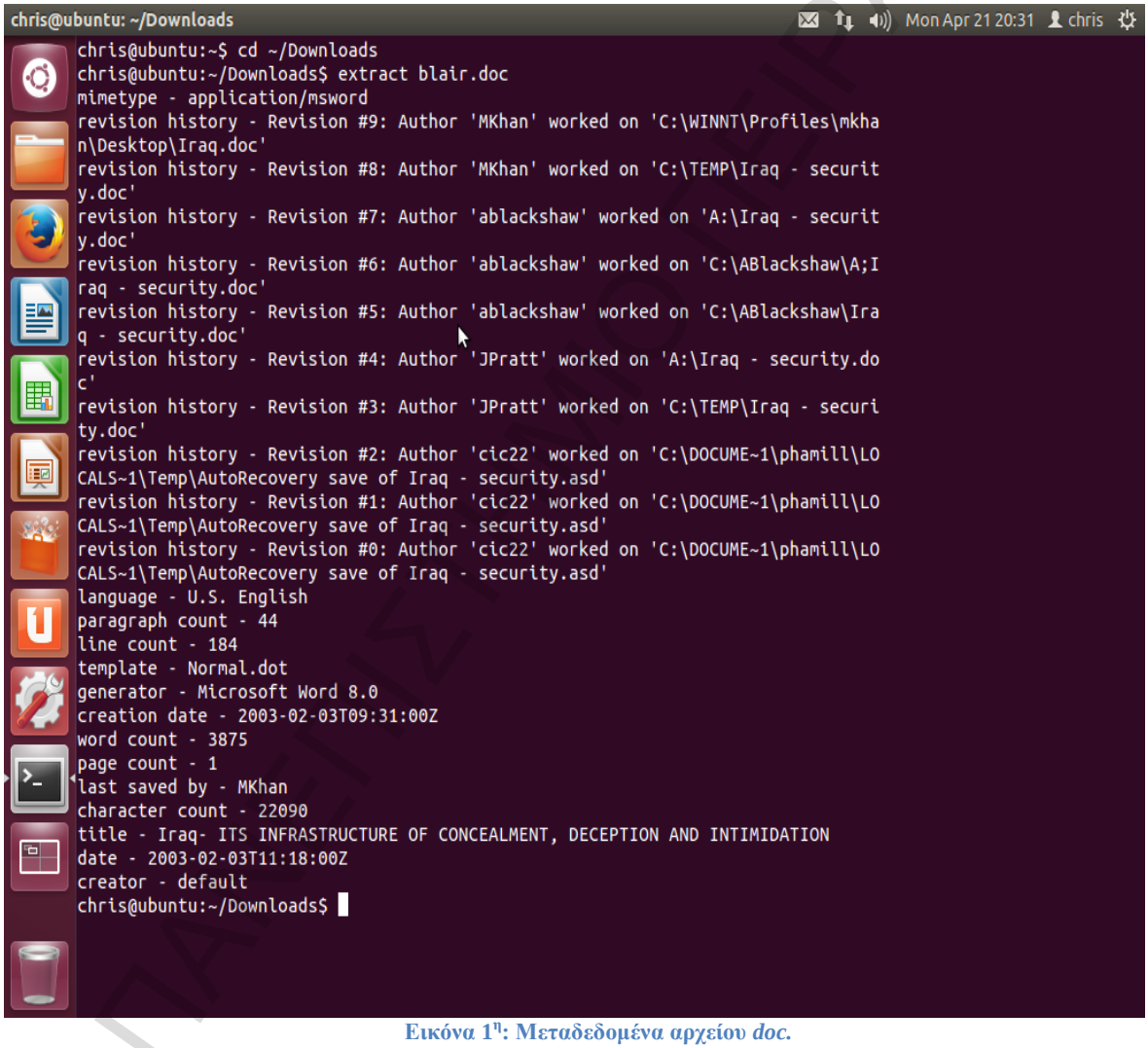
Ένας φάκελος σχετικά με τη στρατιωτική κατάσταση στο Ιράκ το 2003, ο οποίος είχε συνταχθεί από τη βρετανική κυβέρνηση, στην πραγματικότητα αντιγράφηκε από την έρευνα ενός μεταπτυχιακού φοιτητή (ονόματι Richard M. Smith), με θέμα το Ιράκ.

Ένας υπάλληλος της εταιρείας «ComputerBytesMan.com», εξήγαγε το ακόλουθο ημερολόγιο αναθεώρησης, το οποίο δείχνει την πρόοδο στις αναθεωρήσεις ή τα αντίγραφα του εγγράφου, συμπεριλαμβανομένου και του δήθεν αντιγράφου, το οποίο είχε προετοιμασθεί για τον τότε Υπουργό Colin Powell (Αναθεώρηση 5), για την

παρουσίασή του στα Ηνωμένα Έθνη. Το ημερολόγιο αναθεώρησης είχε τόσο λεπτομερείς πληροφορίες, που οι χρήστες του Διαδικτύου μπορούσαν να εντοπίσουν όχι μόνο τους συντάκτες, αλλά και σε ποια θέση της βρετανική κυβέρνησης υπηρετούσαν, τη στιγμή που η ιστορία αυτή πρωτοεμφανίσθηκε στο διαδίκτυο.

Βασισμένο στα μεταδεδομένα, το βρετανικό έγγραφο φάνηκε ότι αντιγράφηκε και στη συνέχεια, τροποποιήθηκε για να δοθεί έμφαση. Βάσει των αναθεωρήσεων, έγινε δυνατή η εύρεση του αρχικού συντάκτη του εγγράφου, του μεταπτυχιακού φοιτητή, ονόματι *Ibrahimal-Marashi.*

Το παραπάνω έγγραφο βρέθηκε σε δικτυακό χώρο και, κάνοντας χρήση του εργαλείου *Extract*, σε περιβάλλον Linux, εξήχθησαν (επιβεβαιώθηκαν) τα εξής μεταδεδομένα:



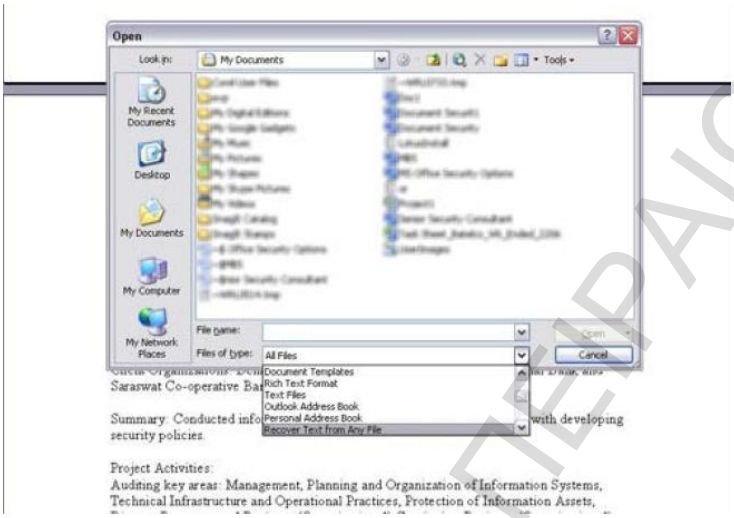
Εικόνα 1η: Μεταδεδομένα αρχείου *doc.*

4.1.1 Εμφάνιση μεταδεδομένων χωρίς χρήση εργαλείων

Από προεπιλογή τα έγγραφα του MS Office 2003 περιέχουν κρυμμένα δεδομένα, η ανακάλυψη των οποίων δεν είναι τόσο δύσκολη. Για παράδειγμα ένα χαρακτηριστικό του Word επιτρέπει το άνοιγμα κατεστραμμένου εγγράφου με προβολή του κειμένου

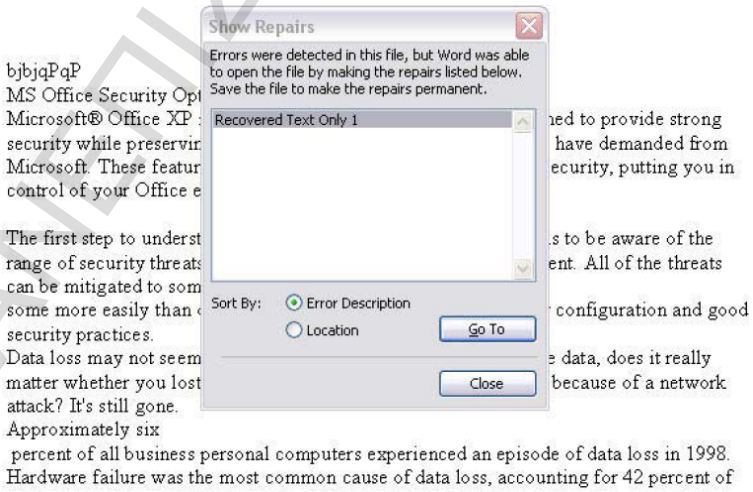
χωρίς μορφοποίηση. Μπορεί επίσης να χρησιμοποιηθεί για την προβολή ορισμένων μεταδεδομένων που σχετίζονται με ένα έγγραφο, με τα εξής βήματα: (12)

Αρχείο → "Άνοιγμα" → λίστα Αρχεία Τύπου → "Ανάκτηση Κειμένου από Οποιοδήποτε Αρχείο" → εντοπισμός ενός αρχείου Word *.doc → "Άνοιγμα".



Εικόνα 2η: Ανάκτηση Κειμένου από Οποιοδήποτε Αρχείο

Το έγγραφο ανοίγει χωρίς καμία μορφοποίηση. Μετά την κύλιση μέσα στο έγγραφο, μπορεί κανείς να δει πληροφορία, τέτοια όπως το όνομα του συγγραφέα του εγγράφου, τη διαδρομή του αποθηκευμένου εγγράφου κ.ο.κ. Η πληροφορία που φαίνεται μπορεί να μην είναι στο περιεχόμενο. Θα πρέπει να είναι κανείς προσεκτικός για να διακρίνει ποια πληροφορία από αυτή που βλέπει είναι μέρος του κειμένου του εγγράφου ή των μεταδεδομένων που προστέθηκαν στο έγγραφο. (13)

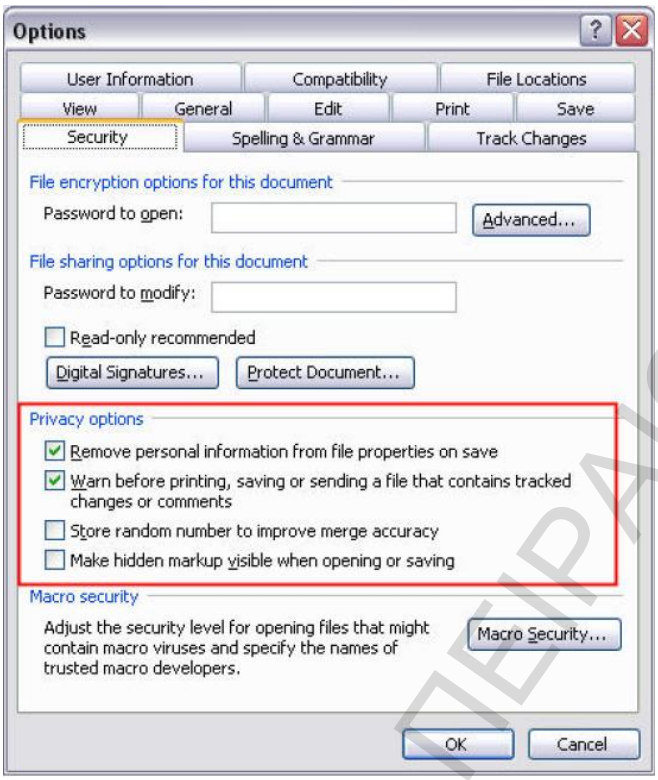


Εικόνα 3η: Πλαίσιο διαλόγου μετά την ανάκτηση αρχείου

Times New Roman
Symbol
Symbol
Tahoma
Tahoma
MS Office Security Options
MS Office Security Options
User 2
User 2
MS Office Security Options
User 2
Normal
Microsoft Office Word
MS Office Security Options
Root Entry
1Table
1Table
WordDocument
WordDocument
SummaryInformation
SummaryInformation
DocumentSummaryInformation
DocumentSummaryInformation
CompObj
CompObj
Microsoft Office Word Document
MSWordDoc
Word.Document.8

Εικόνα 4η: Δείγμα περιεχομένων ανακτηθέντος αρχείου

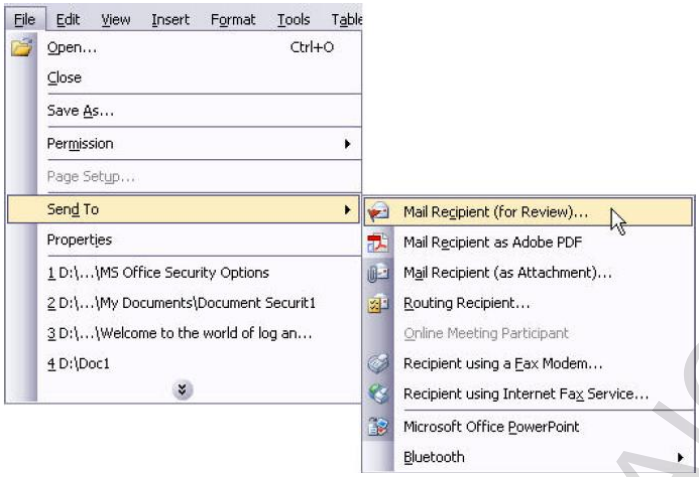
Αν κανείς χρησιμοποιεί την "Παρακολούθηση Αλλαγών" από το μενού "Εργαλεία", "Εκδόσεις" από το μενού "Αρχείο", ή την επιλογή "Να επιτρέπεται η γρήγορη αποθήκευση" από την καρτέλα "Αποθήκευση του πλαισίου διαλόγου "Επιλογές", διαθέσιμο στο μενού "Εργαλεία", θα πρέπει να κοιτάξει να αφαιρέσει όποια κρυμμένη ή διαγραμμένη πληροφορία που μπορεί να παραμένει στο έγγραφο.



Εικόνα 5[η]: Απομάκρυνση Προσωπικών Πληροφοριών από Ιδιότητες Αρχείου

Το Microsoft Office 2003 έχει επίσης την ικανότητα να διενεργεί ταχείες αποθηκεύσεις (Fast saves), σε αντιστοιχία με την Αυτόματη Αποθήκευση των μεταγενέστερων εκδόσεων. Αν ο χρήστης συνθέσει ένα έγγραφο, ενώ η "Ταχεία Αποθήκευση" είναι ενεργοποιημένη, και διαγράψει κάποιο κείμενο, αυτό το διαγραμμένο κείμενο είναι πιθανόν να παραμείνει με το έγγραφο σε όλη τη διάρκεια ζωής του. Κάποιο κείμενο απλά δε διαγράφεται ποτέ. (14)

Αν το έγγραφο έχει σταλεί μέσω του Outlook, όταν ο παραλήπτης ανοίξει το έγγραφο και δει τις ιδιότητες του εγγράφου, προβάλλονται εγγραφές όπως _TentativeReviewCycleID και _ReviewCycleID, _EmailSubject, _AuthorEmail, και _AuthorEmailDisplayName. Οι πληροφορίες αυτές αποθηκεύονται επίσης στο σύστημα του παραλήπτη σε ένα αρχείο που ονομάζεται ‘Adhoc.rcd’ ή ‘Review.rcd’ (ανάλογα με την έκδοση του MS Office που χρησιμοποιείται, π.χ. ‘Review.rcd’ είναι στην περίπτωση του Office 2003). Συνήθως βρίσκεται μέσα στο φάκελο Έγγραφα και Ρυθμίσεις Χρήστη>Δεδομένα Εφαρμογών\Microsoft\Office.



Εικόνα 6η: Αποστολή Εγγράφου για Αναθεώρηση

4.1.2 Φυσική Διεύθυνση

Ανοίγοντας ένα έγγραφο του Word, σε κάθε έκδοση του Office 97 και αργότερα (μέχρι, μη συμπεριλαμβανομένης όμως, του Office XP) σε ένα πρόγραμμα επεξεργασίας κειμένου, που δεν είναι του Word, θα εμφανισθούν – όπως αναφέρθηκαν λεπτομερώς πιο πάνω – τα πρώτα δυαδικά δεδομένα στο έγγραφο. Αν προβούμε σε αναζήτηση αυτού του εγγράφου για τη λέξη-κλειδί PID_GUID, ύστερα από την εμφάνιση κάποιων άλλων δεδομένων, φαίνεται κομμάτι κειμένου Unicode, μέσα αγκύλες, π.χ.:

{ 1 0 4 A 8 A 2 2 - 6 2 3 B - 1 1 D 4 - 8 8 D D - 0 0 D 0 B 7 1 B 0 4 C 4 }

Η συμβολοσειρά *0 0 D 0 B 7 1 B 0 4 C 4* είναι η διεύθυνση MAC (Media Access Control) ή η διεύθυνση υλικού MAC, που έχει παραχωρηθεί σε μια κάρτα Ethernet. Η MAC Ethernet είναι σημαντική, διότι μας επιτρέπει να συνδέσουμε τη δημιουργία ενός εγγράφου στο σύστημα ενός συγκεκριμένου χρήστη. (15)

Αυτό είναι ισχυρή απόδειξη, που δείχνει ότι ένας χρήστης συμμετείχε ή ήταν άμεσα υπεύθυνος για κάποια δραστηριότητα.

4.1.2.1 Περίπτωση ιού Melissa

Ο ιός Melissa ξεκίνησε στις αρχές του 1999 . Ήταν ένας μικρο-ιός, που εστάλη μέσω e-mail στις πρώτες διευθύνσεις που βρήκε στο Microsoft Outlook Address Book. Αρχικά είχε κυκλοφορήσει σε μια συγκεκριμένη ομάδα συνδρομητών (alt.sex). Ο ιός μακροεντολών γρήγορα εξαπλώθηκε το Μάρτιο του 1999.

Ο Melissa εξαπλώθηκε σε ένα αρχείο του Microsoft Word, με το όνομα *List.doc*, που αποτελείτο από κωδικούς πρόσβασης σε ιστοσελίδες πορνογραφικού περιεχομένου. Αν κάποιος χρήστης «κατέβαζε» το αρχείο και να το άνοιγε σε Microsoft Word, εκτελείτο μια μακροεντολή μέσα στο έγγραφο και έστελνε με e-mail το αρχείο *List.doc* σε άλλους χρήστες. (16)

Λόγω του τεράστιου αντίκτυπου, η ομοσπονδιακή κυβέρνηση προσπάθησε να αναζητήσει και να προσαγάγει το δημιουργό του ιού. Ήταν πολύ δύσκολο έργο, αλλά

υποστηρίχθηκε από δύο ανεξάρτητους εμπειρογνώμονες, οι οποίοι ανακάλυψαν κάποια καταστροφική πληροφορία.

Οι εμπειρογνώμονες εξέτασαν τα μεταδεδομένα του εγγράφου που περιείχε τον ιό και εμφάνισαν το GUID (Globally Unique Identifier). Πραγματοποίησαν περισσότερη έρευνα και τελικά βρήκαν μια ιστοσελίδα που ανήκε σε έναν κακόβουλο *hacker*, στην οποία εκτίθεντο και άλλα έγγραφα με το ίδιο GUID. Στο τέλος, οι πληροφορίες που ανακαλύφθηκαν διαβιβάστηκαν στο ZDNet, το οποίο δημοσίευσε πολλά άρθρα, καθώς και στις ομοσπονδιακές αρχές. Αυτή η πολύτιμη πληροφορία συντέλεσε στον εντοπισμό και σύλληψη του David Smith, ως δημιουργού του ιού.

4.2 Αρχεία Open Office – Μεταδεδομένα

Το OpenOffice.org σκιαγραφεί στις προδιαγραφές της μορφής των XML αρχείων του τη δυνατότητα αποθήκευσης πολλών μεταδεδομένων μέσα στα έγγραφα, λογιστικά φύλλα και στις παρουσιάσεις του: (17)

- Πηγή
- Τίτλος
- Περιγραφή
- Θέμα
- Λέξεις κλειδιά
- Αρχικός δημιουργός
- Δημιουργός
- Εκτυπώθηκε από
- Ημερομηνία και ώρα δημιουργίας
- Ημερομηνία και ώρα τροποποίησης
- Ημερομηνία και ώρα εκτύπωσης
- Πρότυπο εγγράφου
- Γλώσσα
- Μεταδεδομένα ορισμένα από το χρήστη
- Στατιστικά εγγράφου

```
<office:meta>
  <dc:title>Title</dc:title>
  <dc:description>Comment</dc:description>
  <dc:language>en-US</dc:language>
  <meta:initial-creator>Name</meta:initial-
creator>
  <meta:creation-date>X</meta:creation-date>
  <dc:creator>Name</dc:creator>
  <meta:generator>NeoOffice</meta:generator>
  <dc:creator>Name</dc:creator>
  <meta:keywords>
    <meta:keyword>First</meta:keyword>
    <meta:keyword>Second</meta:keyword>
  </meta:keywords>
  <dc:date>X:</dc:date>
  <dc:subject>Subject</dc:subject>
  <meta:printed-by>Name</meta:printed-by>
  <meta:print-date>X</meta:print-date>
  <meta:duration-time>X</meta:editing-
duration>
  <meta:editing-cycles>4</meta:editing-cycles>
  <meta:editing-duration>X</meta:editing-
duration>
 </office:meta>
```

Εικόνα 7η: Δείγμα μεταδεδομένων σε OpenOffice.org Writer έγγραφο

Σύμφωνα με το έγγραφο των προδιαγραφών, όλες οι παραπάνω κατηγορίες αποθηκεύονται ή έχουν τη δυνατότητα να αποθηκευτούν με κάθε αρχείο που δημιουργείται στο OpenOffice.org. Αυτή είναι σε μεγάλο βαθμό πληροφορία του είδους που αποθηκεύεται κι από το Microsoft Office, παρόλα αυτά το OpenOffice.org δεν κάνει υποθέσεις για το τι προσωπική πληροφορία θα ήθελε να συσχετίσει ο χρήστης με το αρχείο.

Για παράδειγμα, αν ένας χρήστης δεν εισάγει ένα όνομα στο πλαίσιο πληροφοριών χρήστη της εφαρμογής, το πρόγραμμα δε θα προσπαθήσει να λάβει την πληροφορία εγγραφής από το λειτουργικό σύστημα προκειμένου να ανακαλύψει το όνομα ή τα αρχικά του χρήστη.

Όπως και το Microsoft Office, το OpenOffice.org έχει τη δυνατότητα να αποθηκεύει πληροφορία για να διευκολύνει τη χρήση της λειτουργίας "Παρακολούθηση" και τα χαρακτηριστικά σχολιασμού.

Το OpenOffice.org παρέχει ευκολότερη πρόσβαση στα μεταδεδομένα ενός αρχείου κάτι που διευκολύνει την ανεύρεσή τους από το χρήστη και τον εξεταστή, δεδομένου ότι αυτοί ξέρουν για το τι ψάχνουν. Το OpenOffice.org αποθηκεύει τα μεταδεδομένα που σχετίζονται με το συνολικό έγγραφο, φύλλο εργασίας ή την παρουσίαση σε ένα ξεχωριστό απλό αρχείο κειμένου για ευκολότερη πρόσβαση.

4.2.1 Παγκοσμίως Μοναδικά Αναγνωριστικά (GUID)

Τα Microsoft Office Word, Excel, και PowerPoint αποθηκεύουν ένα μοναδικό αναγνωριστικό που θα μπορούσε να αναγνωρίσει το σύστημα και την εγκατάσταση στην οποία το αρχείο δημιουργήθηκε.

Κατά την εγκατάσταση δημιουργείται το ακόλουθο δευτερεύον κλειδί μητρώου:

HKEY_LOCAL_MACHINE\Software\Microsoft\Windows\CurrentVersion\Uninstall

ή σε έκδοση 64-bit των Windows:

HKEY_LOCAL_MACHINE\Software\WOW6432Node\Microsoft\Windows\Current Version\Uninstall

Η λογική της Microsoft για την ενσωμάτωση αυτής της πληροφορίας στο αρχείο ήταν η διευκόλυνση συνεργασίας με προγράμματα τρίτων. Η Microsoft δήλωσε περαιτέρω ότι ο αριθμός GUID που ορίζεται στο RFC 4122, δεν θα μπορούσε να χρησιμοποιηθεί για να «αναγνωρίσει το συγγραφέα ενός εγγράφου χωρίς γνώση ενδόμυχων του προσωπικού υπολογιστή στον οποίο το έγγραφο δημιουργήθηκε αρχικά. Παρόλα αυτά, η μέθοδος που χρησιμοποιήθηκε για τη δημιουργία του GUID δεν συνέπιπτε εντελώς με τις δηλώσεις αυτές και αποδείχθηκε πολύτιμη για διάφορες υποθέσεις, όπως κατά την έρευνα για τον συγγραφέα του ιού Melissa.

Αυτό που έκανε τον αριθμό GUID τόσο αποκαλυπτικό σε πρώιμες εκδόσεις του Office, όπως οι 97 και 2000, ήταν ότι αποτελείτο από τη διεύθυνση Ethernet Media Access Control (MAC). Δεν υπάρχουν δυο ολόιδιες MAC και η αποθήκευση αυτού του τύπου πληροφορίας μπορούσε να οδηγήσει τους ερευνητές με υποστηρικτικά στοιχεία απευθείας στον υπολογιστή όπου το έγγραφο δημιουργήθηκε, για αυτό πλέον δεν περιλαμβάνεται αυτή η πληροφορία. (18)

Το μεγάλο μειονέκτημα της χρήσης αυτού του αναγνωριστικού σε έρευνες ήταν ότι αποθηκεύεται μόνο όταν το αρχείο δημιουργείται. Επομένως, αν το αρχείο τροποποιηθεί από άλλο χρήστη, το αναγνωριστικό GUID του νέου χρήστη δε θα αποθηκευτεί στο αρχείο και θα περιέχει μόνο αυτό του αρχικού δημιουργού.

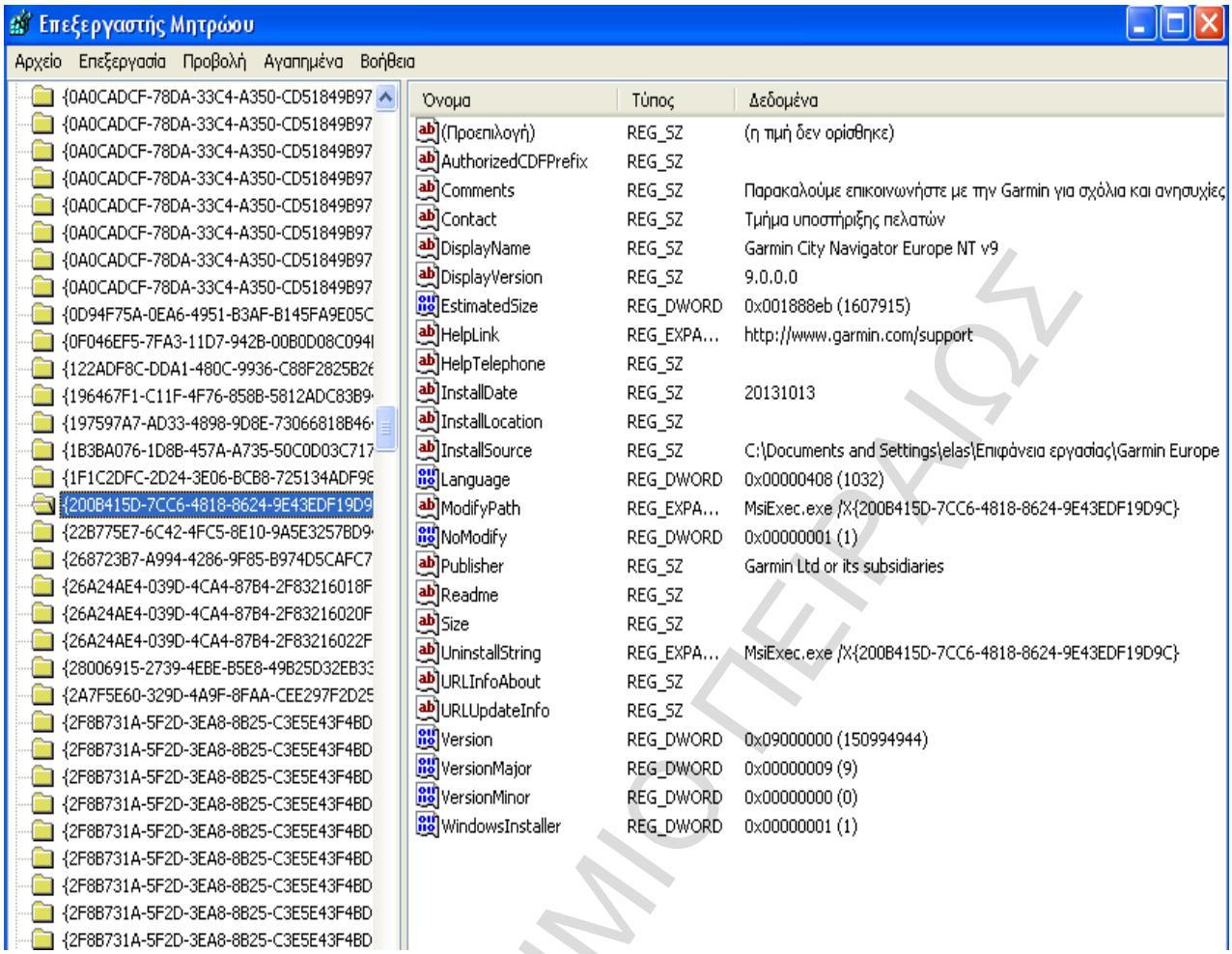
Για προβολή των αναγνωριστικών GUID για τις οικογένειες προγραμμάτων Microsoft Office:

- Εκτέλεση (Run) των Windows και πληκτρολόγηση *regedit*
- Εντοπισμός του ακόλουθου δευτερεύοντος κλειδιού:

HKEY_LOCAL_MACHINE\Software\Microsoft\Windows\CurrentVersion\Uninstall

ή για έκδοση 64-bit των Windows:

HKEY_LOCAL_MACHINE\Software\WOW6432Node\Microsoft\Windows\Current Version\Uninstall



Εικόνα 8η: Αναγνωριστικά GUID σε Windows XP

Τα ονόματα των GUID, ξεκινούν με ένα άγκιστρο ({}). Επομένως, τα αναγνωριστικά GUID είναι τα πρώτα στοιχεία που αναγράφονται στη θέση Uninstall.

Κάθε GUID χρησιμοποιεί την ακόλουθη μορφή:
{BRMMmmmm-PPPP-LLLL-p000-D000000FF1CE}

Ακολουθεί η σημασία και οι δυνατές τιμές των χαρακτήρων από τους οποίους αποτελείται:

Χαρακτήρες	**Ορισμός**	**Δεκαεξαδικές τιμές**
B	Τελική έκδοση	0-9, A-F
R	Τύπος έκδοσης	0-9, A-F
MM	Κύρια έκδοση	0-9
mmmm	Δευτερεύουσα έκδοση	0-9
PPPP	Αναγνωριστικό προϊόντος	0-9, A-F
LLLL	Αναγνωριστικό γλώσσας	0-9, A-F
p	0 για x 86, 1 για x 64	0-1
000	Δεσμευμένο για μελλοντική χρήση,	

	προς το παρόν 0	0
D	1 για το πρόγραμμα εντοπισμού σφαλμάτων	0-1
000000FF1CE	Αναγνωριστικό οικογένειας προγραμμάτων του Office	0-9

Κάθε παράθυρο δεδομένων GUID περιέχει πολλές τιμές. Αυτές οι τιμές περιλαμβάνουν τις παρακάτω πληροφορίες:

Όνομα τιμής	**Περιγραφή**
DisplayName	Το όνομα του προϊόντος που εμφανίζεται στο πλαίσιο διαλόγου "Προσθαφαίρεση προγραμμάτων"
InstallDate	Η ημερομηνία εγκατάστασης του προϊόντος
Product ID	Το Αναγνωριστικό προϊόντος
InstallSource	Η προέλευση της εγκατάστασης
RegCompany	Η καταχωρημένη εταιρεία
RegOwner	Το όνομα χρήστη που έχει καταχωρηθεί

Επίσημη έκδοση

Οι τιμές έκδοσης καθορίζουν το επίπεδο της κυκλοφορίας, όπως μια έκδοση beta ή μια έκδοση στην παραγωγή (RTM).

Όνομα τιμής	**Έκδοση**
0	Οποιαδήποτε κυκλοφορία πριν από την Beta 1
1	Beta 1
2	Beta 2
3	Release Candidate 0 (RC0)
4	Release Candidate 1 (RC1) / OEM Προεπισκόπηση έκδοσης
5-8	Δεσμευμένες τιμές
9	RTM. Αυτή είναι η πρώτη έκδοση που αποστέλλεται (αρχική έκδοση).
A	Το Service Pack 1 (SP1). Αυτή η τιμή δεν χρησιμοποιείται εάν ο κωδικός προϊόντος δεν αλλάξει μετά την έκδοση RTM
B	Το Service Pack 2 (SP2). Αυτή η τιμή δεν χρησιμοποιείται εάν ο κωδικός προϊόντος δεν αλλάξει μετά την έκδοση RTM
C	Το Service Pack 3 (SP3). Αυτή η τιμή δεν χρησιμοποιείται εάν ο κωδικός προϊόντος δεν αλλάξει μετά την έκδοση RTM
D-F	Δεσμευμένες τιμές

Τελική έκδοση

Ο τύπος έκδοσης καθορίζει το ακροατήριο για την οικογένεια προγραμμάτων Office 2007, όπως Enterprise ή Retail.

Τιμή	Τύπος έκδοσης
0	Άδειας χρήσης βάσει ποσότητας
1	Λιανικής πώλησης/OEM
2	Δοκιμαστική έκδοση

Τύπος έκδοσης

Το Αναγνωριστικό προϊόντος είναι η έκδοση της οικογένειας προγραμμάτων του Office 2007 ή του προγράμματος, όπως το Microsoft Office Professional 2007 ή του Microsoft Office Standard 2007.

Product ID	ΑΠΟΘΗΚΕΥΤΙΚΗ ΜΟΝΑΔΑ
0011	Microsoft Office Professional Plus 2007
0012	Microsoft Office Standard 2007
0013	Microsoft Office Basic 2007
0014	Microsoft Office Professional 2007
0015	Microsoft Office Access 2007
0016	Microsoft Office Excel 2007
0017	Microsoft Office SharePoint Designer 2007
0018	Microsoft Office PowerPoint 2007
0019	Microsoft Office Publisher 2007
001A	Microsoft Office Outlook 2007
001B	Microsoft Office Word 2007
001C	Microsoft Office Access Runtime 2007
0020	Microsoft Office Compatibility Pack for Word, Excel, and PowerPoint 2007 File Formats
0026	Microsoft Expression Web
002E	Microsoft Office Ultimate 2007
002F	Microsoft Office Home and Student 2007
0030	Microsoft Office Enterprise 2007
0031	Microsoft Office Professional Hybrid 2007
0033	Microsoft Office Personal 2007
0035	Microsoft Office Professional Hybrid 2007
003A	Microsoft Office Project Standard 2007
003B	Microsoft Office Project Professional 2007
0044	Microsoft Office InfoPath 2007
0051	Microsoft Office Visio Professional 2007
0052	Microsoft Office Visio Viewer 2007
0053	Microsoft Office Visio Standard 2007
00A1	Microsoft Office OneNote 2007
00A3	Microsoft Office OneNote Home Student 2007
00A7	Calendar Printing Assistant for Microsoft Office Outlook 2007

00A9	Microsoft Office InterConnect 2007
00AF	Microsoft Office PowerPoint Viewer 2007 (English)
00B0	Microsoft Save as PDF add-in
00B1	The Microsoft Save as XPS add-in
00B2	Microsoft Save as PDF or XPS add-in
00BA	Microsoft Office Groove 2007
00CA	Microsoft Office Small Business 2007
10D7	Microsoft Office InfoPath Forms Services
110D	Microsoft Office SharePoint Server 2007
1122	Windows SharePoint Services Developer Resources 1.2
0010	SKU - Microsoft Software Update for Web Folders (English) 12

Αναγνωριστικό γλώσσας

Το αναγνωριστικό γλώσσας ή LCID, ποικίλλει από γλώσσα σε γλώσσα. Επειδή το LCID αποθηκεύεται στο αναγνωριστικό GUID σε δεκαεξαδική μορφή, ίσως χρειαστεί να μετατραπεί η τιμή LCID σε μια δεκαδική τιμή για να καθοριστεί η γλώσσα του αναγνωριστικού GUID. Για παράδειγμα, η δεκαεξαδική τιμή 0409 μετατρέπεται σε δεκαδική τιμή 1033. Αυτή η τιμή αντιπροσωπεύει τα Αγγλικά.

Για παράδειγμα τα πρώτα 16 ψηφία ενός GUID είναι 91120000-0014-0407. Αυτό το GUID δημιουργήθηκε από την αρχική έκδοση (9) λιανικής πώλησης ή OEM (1), έκδοση 120000, του Microsoft Office Professional 2007 (0014). Στην περίπτωση αυτή, η δεκαεξαδική τιμή 0407 μετατρέπεται σε τη δεκαδική τιμή 1031. Αυτή η τιμή αντιπροσωπεύει τα γερμανικά.

Δεν είναι όμως όλα τα μεταδεδομένα εύκολα προσβάσιμα μέσω της διεπιφάνειας χρήστη κάθε προγράμματος Office. Κάποια είναι προσβάσιμα μέσω ασυνήθιστων τρόπων, όπως ανοίγοντας το αρχείο σε ένα χαμηλού επιπέδου επεξεργαστή δυαδικών αρχείων.

4.2.2 Μακροεντολές και ταχείες αποθηκεύσεις

Οι μακροεντολές, οι οποίες χρησιμοποιούνται σε στοιχεία δεδομένων του office, μπορεί να είναι πολύ χρήσιμες και μπορεί επίσης να αυξήσουν την παραγωγικότητα σε πολλές περιπτώσεις. Ωστόσο, το Microsoft Office, επίσης, εφοδιάζει με ένα επιπλέον κομμάτι πληροφορίας κάθε μακροεντολή που χρησιμοποιείται στο έγγραφο, το υπολογιστικό φύλλο ή παρουσίαση - το όνομα του συγγραφέα της μακροεντολής. Η δυνατότητα της ταχείας αποθήκευσης δίνεται επίσης στο Microsoft Office , η οποία αποτελεί κατάλληλο τρόπο για τον χρήστη να βεβαιωθεί ότι εάν ο υπολογιστής «καταρρεύσει», ένα πρόσφατο αντίγραφο ασφαλείας είναι μόνο ένα κλικ μακριά. Αυτό μπορεί να δώσει σημαντικές πληροφορίες σε έναν ερευνητή ψηφιακών

πειστηρίων. Παρόμοια με τα άλλα μεταδεδομένα, οι αλλαγές που αποθηκεύτηκαν κατά τη διάρκεια μιας γρήγορης αποθήκευσης μπορεί να αποκαλύψουν ευαίσθητες πληροφορίες σε έναν ερευνητή, όταν ιδωθούν με τη χρήση π.χ., του hex -editor.

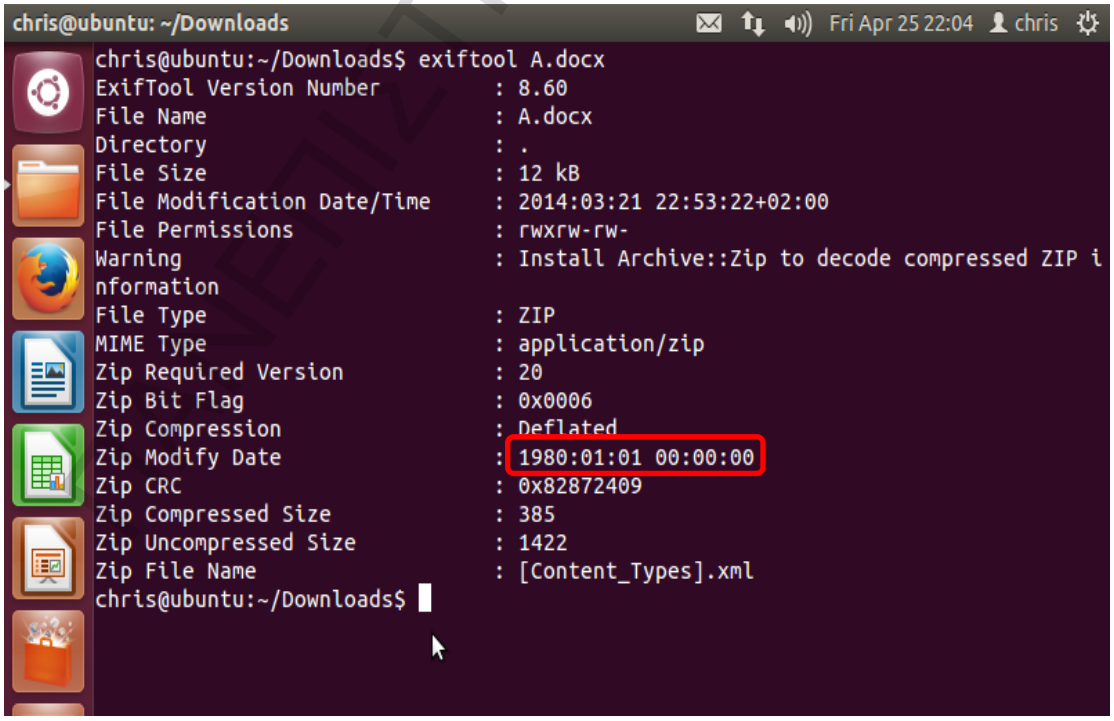
Στα ηλεκτρονικά αρχεία το κείμενο μπορεί να παραμείνει μετά τη διαγραφή. Σύμφωνα με την έρευνα του Gartner Group πάνω στα μεταδεδομένα στο Office "οι χρήστες δεν θυμούνται ότι υπάρχουν μεταδεδομένα όταν αποστέλλουν ένα αρχείο σε κάποιο άτομο. Ορισμένα μεταδεδομένα δεν είναι ποτέ εμφανή, όπως είναι τα τμήματα που διαγράφονται από τους χρήστες, αλλά δε διαγράφονται πραγματικά από το Microsoft Office, όταν η λειτουργία της «ταχείας αποθήκευσης» είναι ενεργοποιημένη".

4.2.3 Αποτυπώσεις χρόνου

Ο χρόνος είναι συχνά κρίσιμης σημασίας σε έρευνες ψηφιακών πειστηρίων. Τόσο το Open Office, όσο και το Office Open XML, περιέχουν πολλές εσωτερικές χρονοσφραγίδες, οι οποίες δηλώνουν τη στιγμή που τα έγγραφα δημιουργήθηκαν ή τροποποιήθηκαν. Χρονοσφραγίδες μπορούμε να βρούμε σε αρχείο ZIP, στα ενσωματωμένα αρχεία XML, και ενδεχομένως σε άλλα ενσωματωμένα αντικείμενα (για παράδειγμα, στην επικεφαλίδα EXIF των ενσωματωμένων αρχείων JPEG).

Δυστυχώς, δεν είναι όλες οι χρονοσφραγίδες ακριβείς. Στην περίπτωση του NeoOffice (μία θύρα του OpenOffice.org στο Macintosh) καθορίζει τις χρονοσφραγίδες των αρχείων ZIP να είναι το ίδιο με το ρολόι του συστήματος.

Το Microsoft Word και Excel, από την άλλη, καθορίζουν τη χρονοσφραγίδα στο αρχείο ZIP, την 1η Ιανουαρίου 1980, την εποχή του συστήματος αρχείων FAT της Microsoft. Αυτό φαίνεται στην εικόνα 9.



Εικόνα 9η: Λανθασμένη χρονοσφραγίδα σε αρχείο *.docx*

Αυτές οι χρονοσφραγίδες μπορεί να είναι σημαντικές σε μία έρευνα. Για παράδειγμα, οι χρονοσφραγίδες πιθανώς να εμφανίζουν πότε ένα αρχείο ODF ή Office Open XML τροποποιήθηκε με μία ODF / Office Open XML-aware εφαρμογή. Οι χρονοσφραγίδες μπορεί να υποδεικνύουν πολλαπλές περιόδους επεξεργασίας.

Εναλλακτικά, θα μπορούσαν να υποδεικνύουν αλλοίωση ενός εγγράφου. Οι χρονοσφραγίδες θα μπορούσαν ακόμη να χρησιμοποιηθούν σε έναν *file carver* και να καθορίσουν ποια ανακτηθέντα κομμάτια ενός αρχείου ταιριάζουν με άλλα κομμάτια.

4.3 Αρχεία PDF

Τα αρχεία που δημιουργούνται σε πρότυπο PDF, χρησιμοποιούνται συνήθως ως διαδικασία διανομής αρχείων σε ένα καθολικό πρότυπο μιας αναγνώσιμης διασυστημικής πλατφόρμας. Τα χαρακτηριστικά που δίνονται από τα PDF αρχεία, παρουσιάζουν, επίσης, μία πρόκληση σε έναν ερευνητή ψηφιακών πειστηρίων και ένα συγκεκριμένο ποσοστό κινδύνου στο χρήστη.

Το πρόσθετο πρόγραμμα που περιλαμβάνεται στην Adobe μπορεί να λειτουργήσει άψογα με το OpenOffice.org και τα προϊόντα του Microsoft Office. Αυτό το άψογο χαρακτηριστικό (ή λειτουργία) είναι ένα τεράστιο όφελος για το χρήστη, αλλά μπορεί επίσης να παραγάγει μια πρόκληση όσον αφορά τα αποθηκευμένα μεταδεδομένα.

Όταν ένα έγγραφο μετατρέπεται σε ένα έγγραφο PDF, όλα τα μεταδεδομένα, τα οποία είχαν αποθηκευθεί μαζί με το πρωτότυπο έγγραφο, όπως το όνομα του δημιουργού, η έκδοση, καθώς και η πληροφορία σχετικά με την παρακολούθηση αλλαγών, αποθηκεύονται επίσης στο έγγραφο PDF. (19)

Το έγγραφο PDF αποθηκεύει μεταδεδομένα, επίσης, όπως προαναφέρθηκε , όταν ένα έγγραφο μετατραπεί σε έγγραφο PDF. Όλα τα μεταδεδομένα που ήταν αποθηκευμένα μαζί με το πρωτότυπο έγγραφο, όπως το όνομα του δημιουργού, η έκδοση, καθώς και η πληροφορία σχετικά με την παρακολούθηση αλλαγών, αποθηκεύονται επίσης στο έγγραφο PDF .

Το έγγραφο PDF έχει την ικανότητα να αποθηκεύει τις ίδιες περίπου πληροφορίες όπως και τα προϊόντα που περιγράφηκαν πιο πάνω. Αναλυτικότερα, περιλαμβάνονται οι εξής πληροφορίες :

- Τίτλος
- Συγγραφέας
- Θέμα
- Λέξεις-κλειδιά
- Δημιουργός
- Παραγωγός
- Ημερομηνία Δημιουργίας (*CreationDate*)
- Ημερομηνία Τροποποίησης (*ModDate*)

Η Adobe Systems έχει επίσης ενταχθεί στον κανονισμό που διέπει τα μεταδεδομένα που αποθηκεύονται σε έγγραφα . Αυτό το χαρακτηριστικό θα

διευκολύνει περισσότερο την αναζήτηση και την αποθήκευση πληροφοριών σε βάσεις δεδομένων, καθώς επίσης και σε *web-based* περιβάλλοντα.

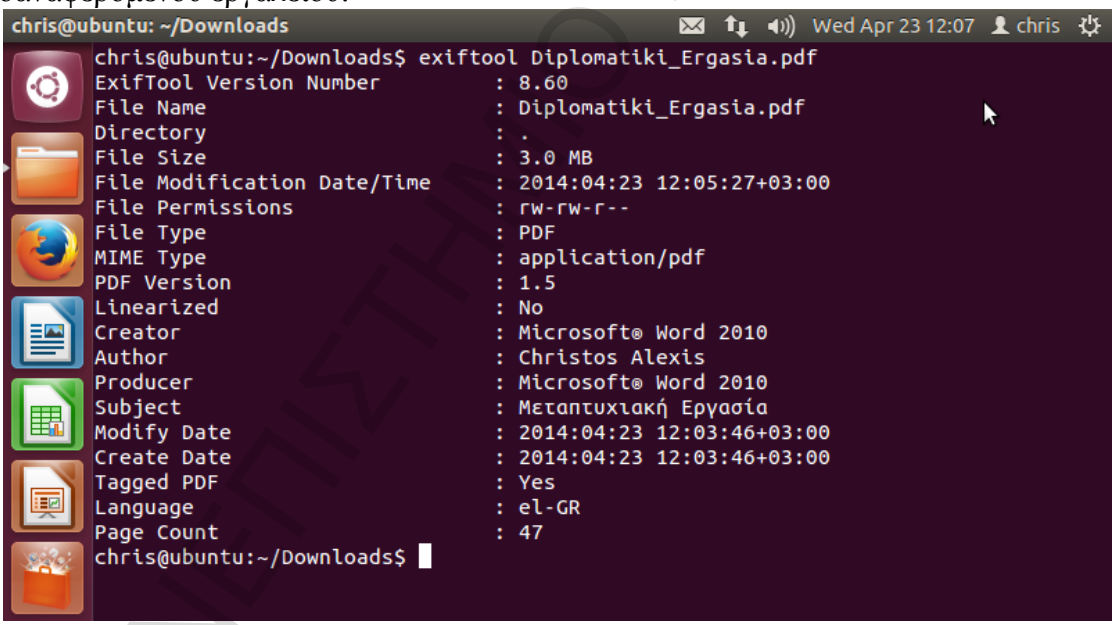
Στην πιο βασική του μορφή, ένα PDF αρχείο είναι ένα αρχείο αποδέκτης (*file container*), που περιέχει μια ακολουθία εντολών της γλώσσας PostScript, καθώς και ενσωματωμένες γραμματοσειρές και γραφικά. Εκτός από την απλή εμφάνιση δεδομένων ενός εγγράφου, τα αρχεία PDF μπορεί να περιέχουν διαδραστικά πεδία τύπου, όπως τα πεδία εισαγωγής κειμένου και τα πλαίσια ελέγχου (*checkboxes*) .

Με την πάροδο του χρόνου, ο τύπος του περιεχομένου που μπορεί να αποθηκευτεί σε ένα αρχείο PDF έχει μεγαλώσει και τώρα περιλαμβάνει συνδέσεις σε εξωτερικό περιεχόμενο, αντικείμενα JavaScript και Flash movie.

Τα αρχεία PDF μπορεί να περιέχούν δύο διαφορετικούς τύπους μεταδεδομένων. Ο Κατάλογος Πληροφορίας Εγγράφου (*Document Information Directory*) περιέχει ζεύγη κλειδιών/τιμών με πληροφορίες του συντάκτη, του τίτλου εγγράφου και χρονοσφραγίδες δημιουργίας/τροποποίησης. Τα σύγχρονα PDFs υποστηρίζουν την ευέλικτη πλατφόρμα μεταδεδομένων (*Extensible Metadata Platform, XMP*), ως μέθοδο αποθήκευσης μεταδεδομένων, η οποία, επίσης, χρησιμοποιείται για να την αποθήκευση μεταδεδομένων, σε ορισμένες μορφές γραφικών αρχείων.

Το ***ExifTool*** μπορεί να χρησιμοποιηθεί για την εξαγωγή μεταδεδομένων από τα αρχεία PDF.

Παρακάτω εμφαίνονται τα μεταδεδομένα ενός αρχείου pdf, με τη χρήση του προαναφερόμενου εργαλείου:



Εικόνα 10η: Μεταδεδομένα αρχείου pdf με *exiftool*

4.4 Αρχεία HTML

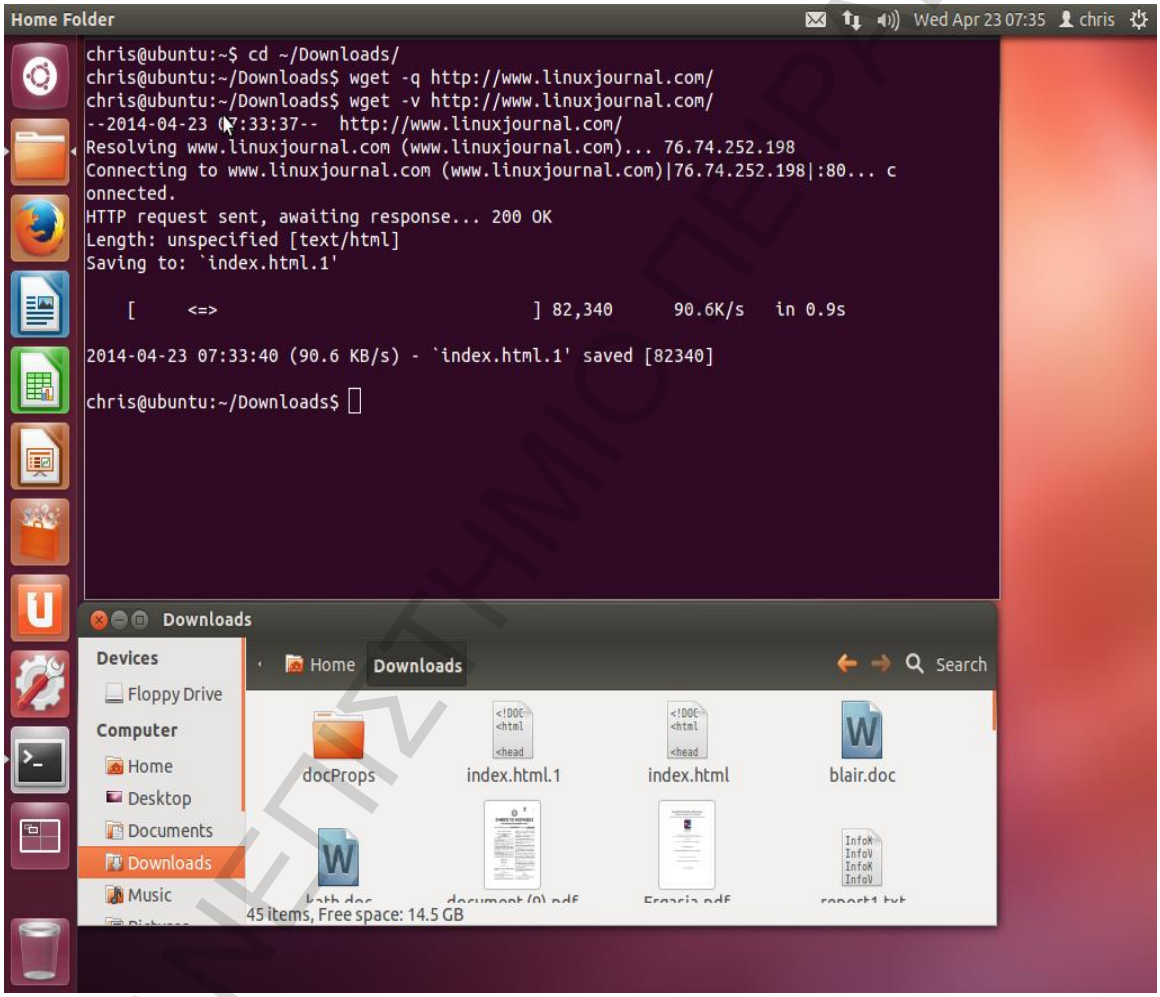
Τα αρχεία HTML είναι αρχεία κειμένου που περιέχουν την πληροφορία που βλέπουν οι χρήστες και ετικέτες που καθορίζουν πληροφορία μορφοποίησης σχετικά με το πώς η πληροφορία θα παρουσιαστεί προς προβολή. Η HTML μπορεί να χρησιμοποιηθεί για να αποθηκεύσει, διανείμει και παρουσιάσει έγγραφα Office και δεδομένα, σε μορφή που να μπορεί να προβληθεί από τους περισσότερους περιηγητές διαδικτύου, διατηρώντας το πλούσιο περιεχόμενο και τη λειτουργικότητα των εγγράφων Office.

4.4.1 Εξαγωγή μεταδεδομένων από HTML

Το εργαλείο που χρησιμοποιείται είναι το extract και η βιβλιοθήκη είναι η Libextractor. Λεπτομερέστερη αναφορά σχετικά με τη βιβλιοθήκη αυτή, αλλά και το εργαλείο, γίνεται στις επόμενες σελίδες της εργασίας.

Η βασικότερη εντολή προκειμένου να εξαγάγουμε μεταδεδόμενα από αρχεία που βρίσκονται σε μία συγκεκριμένη τοποθεσία στον παγκόσμιο ιστό είναι η *wget*.

Στο παράδειγμά μας, θα γίνει εξαγωγή του αρχείου *dmca.pdf* από την τοποθεσία *http://www.linuxjournal.com/* και στη συνέχεια, θα εξαγάγουμε τα υπάρχοντα μεταδεδομένα του. Όλες οι ενέργειες θα λάβουν χώρα σε περιβάλλον *Linux*.



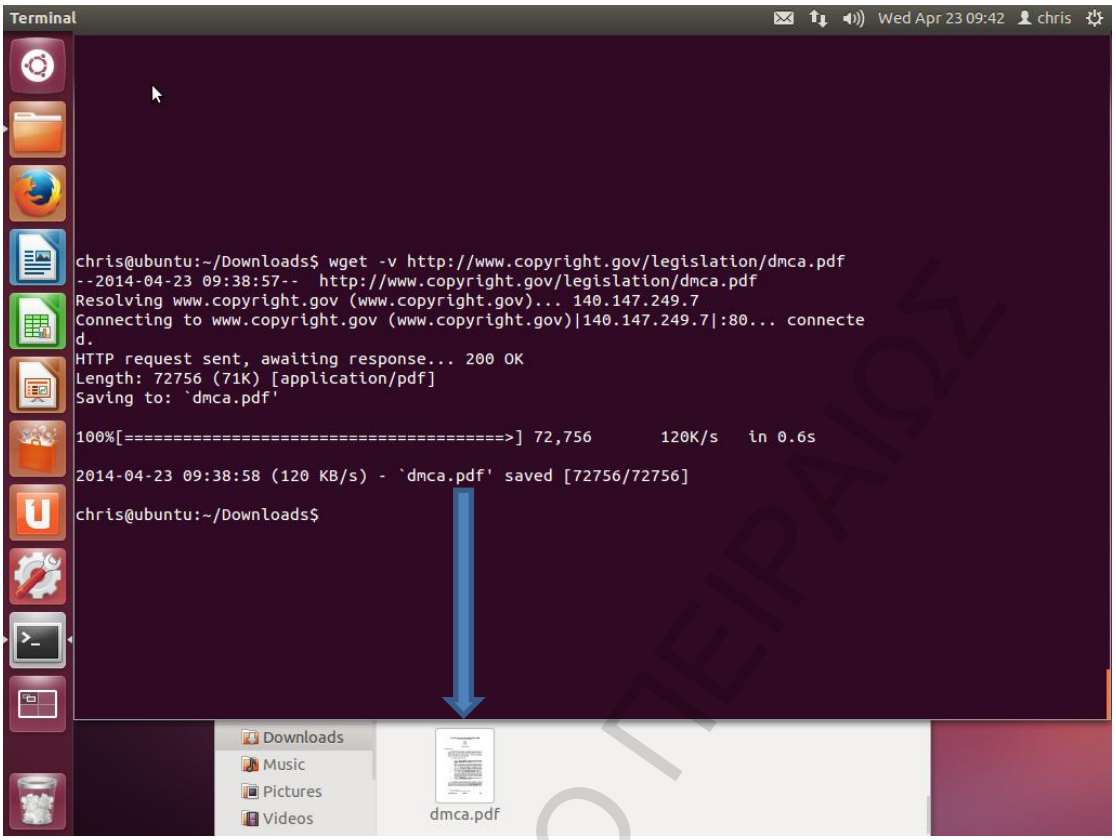
Εικόνα 11η: Εξαγωγή αρχείου *html*

Με την εντολή *wget -q http://www.linuxjournal.com/*, αποθηκεύεται στον κατάλογό μας το αρχείο *index.html*.

Με την ίδια εντολή, αυτή τη φορά όμως κάνοντας χρήση του *wget –v* (*v*=verbose), επιτυγχάνεται το ίδιο αποτέλεσμα και το αρχείο αποθηκεύεται ως *index.html.1*.

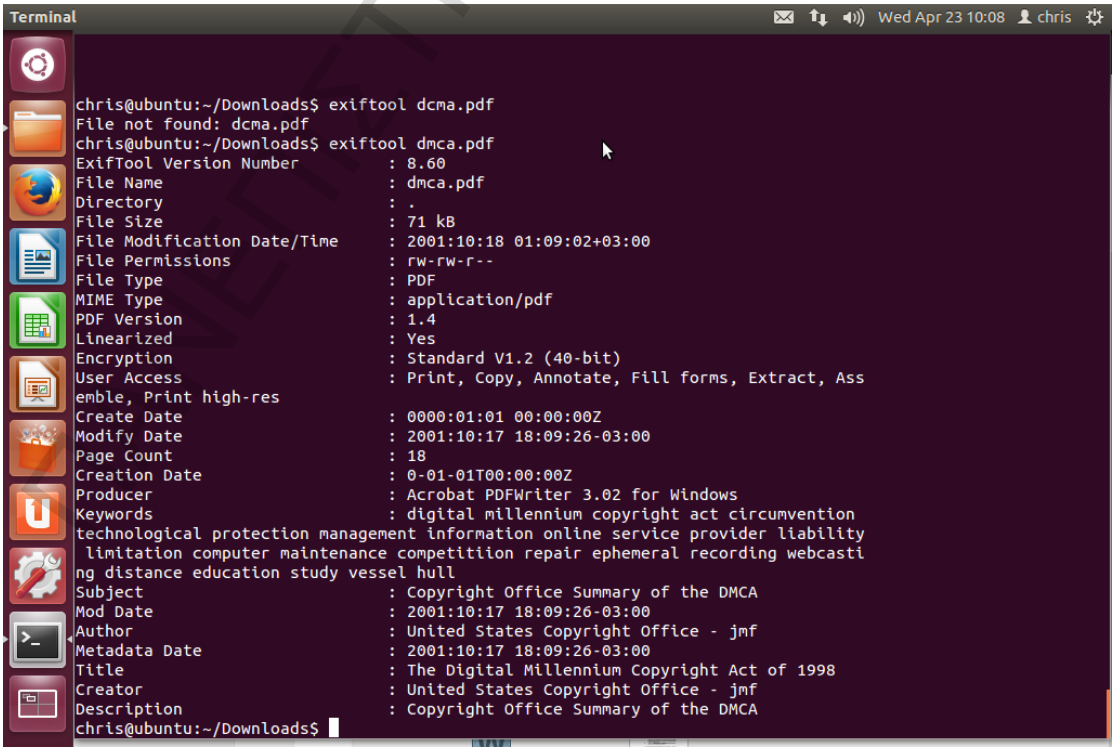
Στην επόμενη περίπτωση, εξάγεται το αρχείο *dmca.pdf*, από έναν ιστότοπο και μελετώνται τα μεταδεδομένα του.

Γράφοντας τις κατάλληλες εντολές, έχουμε τα εξής:



Εικόνα 12η: Εξαγωγή αρχείου *pdf* (wget σε *verbose* μορφή)

Χρησιμοποιώντας το εργαλείο *exiftool* παίρνουμε τα εξής μεταδεδομένα:



Εικόνα 13η: Μεταδεδομένα αρχείου *pdf* από τοποθεσία ιστού

4.5 AAC και MP3

Το Advanced Audio Coding (AAC), το Moving Picture Goup – 1 (MPEG - 1) και το επίπεδο ήχου 3 (MP3) αναπαριστούν συμπιεσμένα και με απώλειες, κωδικοποιημένες μορφές προτύπων, τα οποία αποτελούν μέρος του συνόλου προτύπων MPEG για κωδικοποίηση μουσικής. Το AAC βελτιώθηκε για το πρότυπο MPEG-4 ως MPEG - 4 AAC.

Το ακουστικό MPEG - 4 συντελεί με μια συλλογή από εφαρμογές που κυμαίνονται, από φωνή σε υψηλής ποιότητας πολυκάναλο ήχο και, από φυσικό σε σύνθετους (*synthesized*) ήχους. Το πρωταρχικό όφελος του MPEG - 4 AAC αντί του ήχου MP3 είναι ότι η καθαρότητα του MPEG -4 είναι περίπου διπλάσια από αυτή του MP3 για παρόμοιους ρυθμούς bit, και αρχειοθετεί το μισό μέγεθος για την ίδια αντιληπτή ποιότητα.

Τόσο το AAC όσο και το MP3 περιλαμβάνουν προδιαγραφές μεταδεδομένων. Ωστόσο, τα αυτόνομα αρχεία MP3 δεν περιλαμβάνουν ένα τυποποιημένο τρόπο για την αποθήκευση των μεταδεδομένων. Το 1996 ο Eric Kemp ανέπτυξε μια απλή προσέγγιση για την προσθήκη μιας μικρής ποσότητας δεδομένων στο τέλος ενός αρχείου ήχου. Το πρότυπο για την αποθήκευση αυτών των μεταδεδομένων ονομάσθηκε ID3v1, που σημαίνει "Identify an MP3". Η ετικέτα του ID3v1 καταλαμβάνει τα τελευταία 128 bytes ενός αρχείου MP3 και αρχίζει με τη συμβολοσειρά "TAG". Η ετικέτα δεσμεύει 30 bytes για τον τίτλο, τον καλλιτέχνη, το άλμπουμ και ένα «σχόλιο», 4 bytes για το έτος και 1 byte ως ένα προκαθορισμένο αναγνωριστικό ύφους, τοποθετημένο στο τέλος του αρχείου.

Οι ετικέτες ID3v1 θεωρήθηκαν πολύ μικρές για να περιέχουν αρκετά μεταδεδομένα με νόημα, έτσι, το 1998, εισήχθη το πρότυπο ID3v2. Κάθε ετικέτα ID3v2 περιέχει ένα ή περισσότερα πλαίσια (*frames*), τα οποία μπορεί να είναι έως και 16MB σε μέγεθος, για ένα σύνολο 256 MB για κάθε ετικέτα. Κάθε πλαίσιο μπορεί να περιέχει οποιοδήποτε τύπο πληροφορίας, όπως: το άλμπουμ, τον τίτλο, την τοποθεσία του Ιστού, τους στίχους, πληροφορίες πνευματικών δικαιωμάτων, εικόνες, κ.λπ. Ουσιαστικά, το ID3v1 είναι ένας ακόμη αποδέκτης προδιαγραφής. Το ID3v2 παρέχει κάποια ευελιξία για το γεγονός ότι περιέχει μεταδεδομένα που προστέθηκαν από το χρήστη.

Αν και τα αρχεία AAC περιλαμβάνουν μεταδεδομένα, οι πληροφορίες δεν περιέχονται στην ετικέτα ID3. Η Apple αντιθέτως, χρησιμοποιεί μια αποκλειστική μορφή αρχείου ήχου, που αναφέρεται ως Apple «*Core Audio Format*» για δεδομένα ήχου. Τα αρχεία κάτω από το Core Audio Format βασίζονται στον όγκο τους και περιέχουν AAC πρότυπα δεδομένων. Συγκεκριμένα, η Apple χρησιμοποιεί το «P*rotected AAC»* για να κωδικοποιήσει μουσικούς τίτλους προστατευμένους από αντιγραφή, αγορασμένα από το *iTunes Music Store*.

Τα αρχεία που αγοράζονται μέσω του *iTunes Music Store* περιλαμβάνουν τα ακόλουθα μεταδεδομένα:

- Όνομα
- Διεύθυνση *e-mail* του αγοραστή
- Έτος

- Άλμπουμ
- Ομαδοποίηση
- Σχόλια
- Ύφος
- Στίχοι
- Καλλιτέχνημα
-

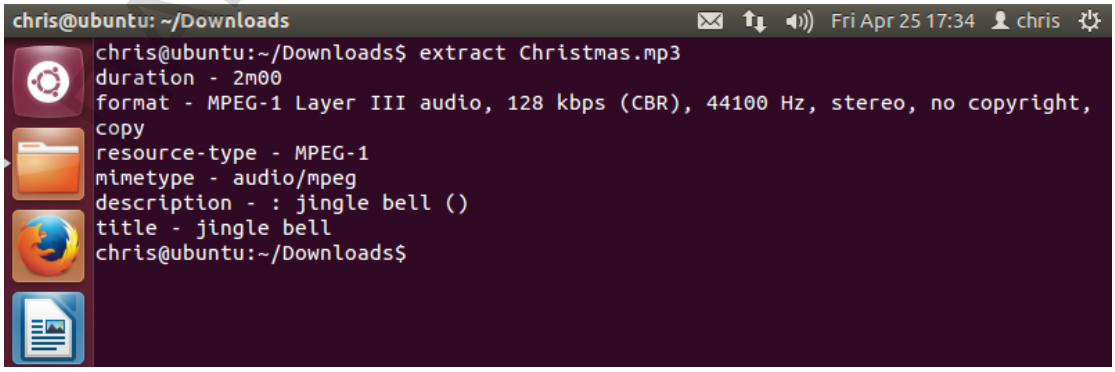
Τα μεταδεδομένα είναι χρήσιμα για πολλούς λόγους. Πρώτον, η διεπαφή χρήστη στο iTunes και το iPod είναι «χτισμένα» από τα μεταδεδομένα. Στη συνέχεια, τα μεταδεδομένα βελτιώνουν την αποτελεσματικότητα της περιήγησης και της αναζήτησης αρχείων. Τέλος, τα μεταδεδομένα χρησιμεύουν στην υποστήριξη και ενίσχυση του περιεχομένου.

Από πλευράς επιστήμης της εγκληματολογίας, τα μεταδεδομένα μπορεί να είναι χρήσιμα στον εντοπισμό του ιδιοκτήτη ή στη συσχέτιση των αρχείων με το δυνητικό ιδιοκτήτη. Η ενσωμάτωση της διεύθυνσης ηλεκτρονικού ταχυδρομείου παρέχει μία σύνδεση (*link*), και τα σχόλια που πρόσθεσε ένας χρήστης ή τις φωτογραφίες θα μπορούσαν να μας δώσουν μια άλλη σύνδεση.

Υπάρχουν μερικά εργαλεία που είναι διαθέσιμα για την εξαγωγή μεταδεδομένων από το ID3 MP3 αρχεία και τα αρχεία AAC. Ένα εργαλείο είναι το *MP3 ::Tag module*, το οποίο είναι ένας συντακτικός αναλυτής ετικετών (*tag parser*) της ID3 Perl, το οποίο είναι διαθέσιμο στο CPAN (*Comprehensive Perl Archive Network*). Το *MP3 ::Perl* μπορεί να διαβάσει τόσο το ID3v1 όσο και το ID3v2. Επιπλέον, το MP3 :: Tag επιτρέπει στο χρήστη να επεξεργασθεί το περιεχόμενο μιας ετικέτας αρχείων MP3 ή να δημιουργήσει μια νέα ετικέτα.

Ένα άλλο εργαλείο το οποίο είναι ικανό να εξαγάγει μεταδεδομένα από τα αρχεία MP3 είναι το (γνωστό από πριν) *libextractor*. Στην παρακάτω εικόνα παρατηρούμε τα μεταδεδομένα που εξάγονται από το αρχείο *Christmas.mp3* με το εργαλείο *extract* (με βιβλιοθήκη *Libextractor*).

Για τα αρχεία AAC, η εφαρμογή MPEG4IP παρέχει πολλά εργαλεία για την εργασία με αρχεία ήχου και εικόνας, συμπεριλαμβανομένων και των αρχείων σε μορφή AAC. Συγκεκριμένα, το *mp4dump* είναι ένα βοηθητικό πρόγραμμα, το οποίο μπορεί να εξαγάγει mp4 μεταδεδομένα αρχείου σε μορφή κειμένου. Η εφαρμογή MPEG4IP περιλαμβάνει, επίσης, το βοηθητικό πρόγραμμα *mp4info*, το οποίο παρέχει συνοπτικές πληροφορίες για αρχεία mp4.



Εικόνα 14η : Libextractor – Μεταδεδομένα στο αρχείο *Christmas.mp3*

4.5.1 Τροποποίηση μεταδεδομένων mp3 και aac - Easytag

Τα μεταδεδομένα που εξάγονται από τα αρχεία *mp3* μπορεί να τροποποιηθούν με τα αντίστοιχα εργαλεία και να προστεθούν καινούργιες πληροφορίες κατά το δοκούν. Ένα τέτοιο εργαλείο, το οποίο χρησιμοποιείται στην πλατφόρμα του *Linux* είναι το *Easytag*.

Το EasyTAG είναι ένα βοηθητικό πρόγραμμα για την προβολή, επεξεργασία και εγγραφή των ετικετών ID3 διαφόρων αρχείων ήχου.

Το EasyTAG, προς το παρόν, υποστηρίζει τα εξής :

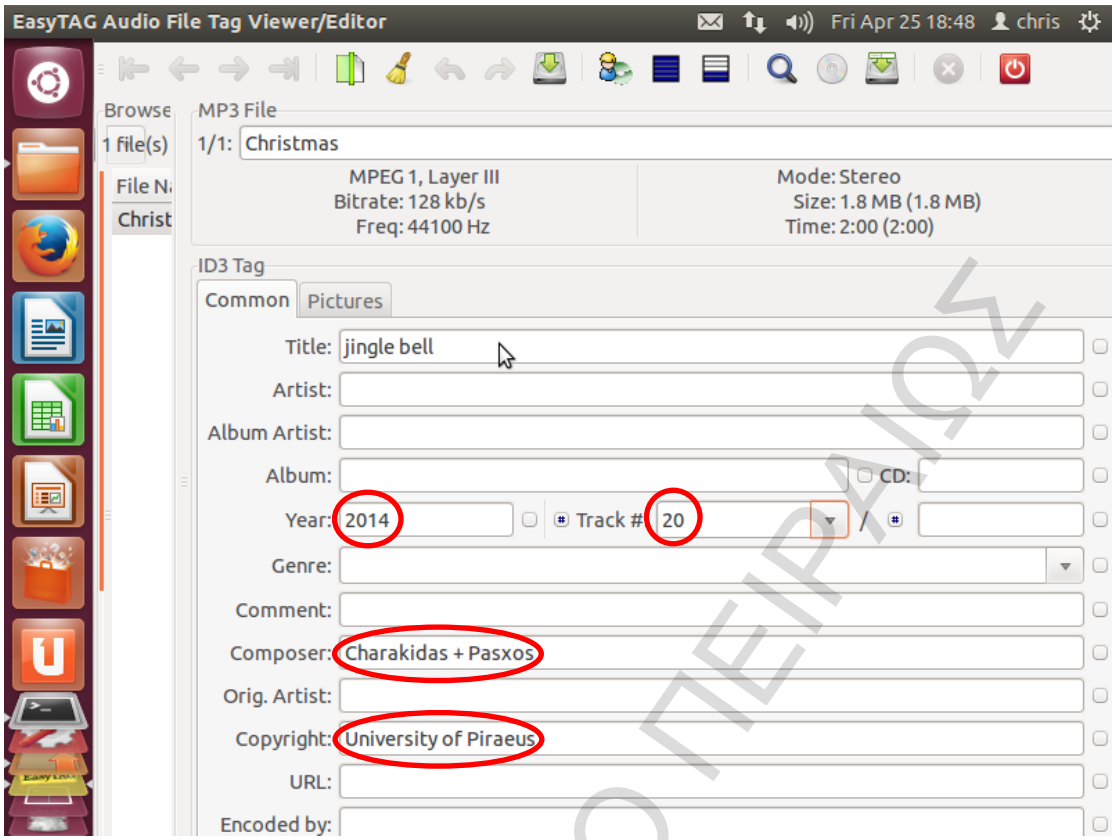
- Προβολή, επεξεργασία, εγγραφή ετικετών σε αρχεία MP3, MP2 (ID3 tag), αρχεία FLAC (FLAC Vorbis tag), αρχεία Ogg Vorbis (Ogg Vorbis tag) και Musepack
- Ανάλυση ονόματος αρχείου και καταλόγου για την αυτόματη ολοκλήρωση των πεδίων
- Δυνατότητα μετονομασίας των αρχείων από την ετικέτα (χρησιμοποιώντας μάσκες) ή με τη φόρτωση ενός αρχείου κειμένου
- Επεξεργασία επιλεγμένων αρχείων του επιλεγμένου καταλόγου
- Δυνατότητα περιήγησης στους υποκαταλόγους
- Αναδρομή για ετικετοποίηση, διαγραφή, μετονομασία, αποθήκευση
- Ανάγνωση πληροφοριών κεφαλίδας του αρχείου (ρυθμός μετάδοσης, χρόνος) και εμφάνισή τους
- Αναίρεση και επανάληψη των τελευταίων αλλαγών
- Δυνατότητα να ανοίξει έναν κατάλογο ή ένα αρχείο με ένα εξωτερικό πρόγραμμα
- Υποστήριξη CDDB (από το πρωτόκολλο http)

Στην παρακάτω εικόνα (Εικόνα 14) παρατηρούμε τις τροποποιήσεις που μπορεί να λάβουν χώρα, με τη συμβολή του εργαλείου EasyTAG. Δίνοντας την εντολή:

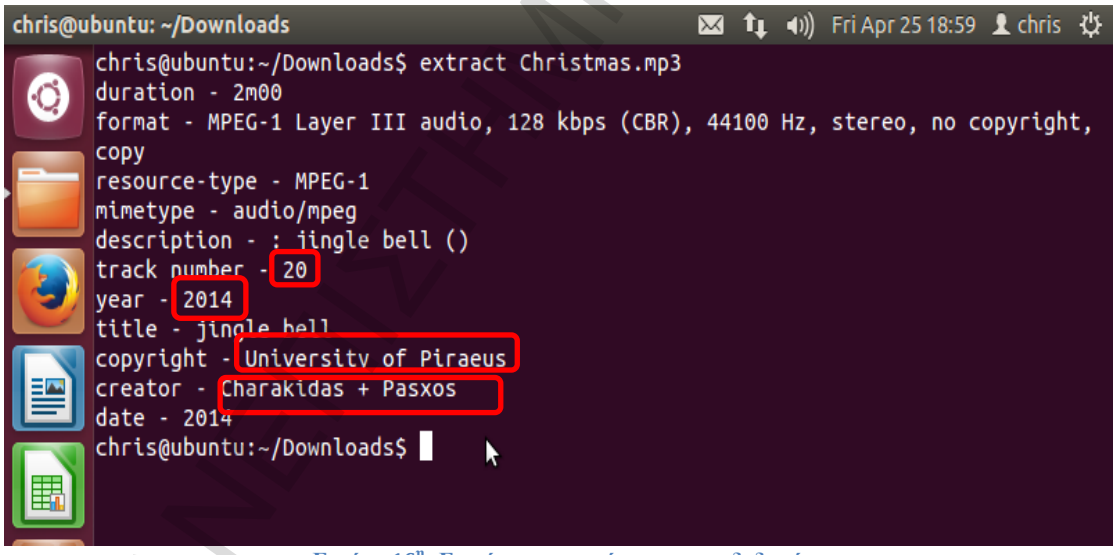
easytag Christmas.mp3

εμφανίζεται μία φόρμα, στην οποία μπορούμε να προβούμε σε αρκετές τροποποιήσεις μεταδεδομένων του αρχείου που ερευνούμε. Έτσι, παρατηρούνται οι αλλαγές που έλαβαν χώρα, πλαισιωμένες σε κόκκινα σχήματα.

Στην εικόνα 15, παρατηρείται ότι, μετά την τροποποίηση των μεταδεδομένων στο αρχείο *Christmas.mp3*, εμφαίνονται τα μεταδεδομένα που προσθέσαμε.



Εικόνα 15η: EasyTAG – Τροποποίηση μεταδεδομένων αρχείου *Christmas.mp3*



Εικόνα 16η: Εμφάνιση καινούργιων μεταδεδομένων

4.6 Tagged Image File Format (TIFF)

Τα αρχεία αυτού του μορφότυπου είναι αρχεία αποδέκτες, που μπορούν να φέρουν δεδομένα και μεταδεδομένα σε μια ποικιλία μορφών. Πληροφορίες που μπορούν να συμπεριληφθούν σε ένα αρχείο *TIFF*, έκδοσης 6.0, περιλαμβάνει τα παρακάτω:

- Ημερομηνία και ώρα που δημιουργήθηκε η εικόνα
- Μία περιγραφή της εικόνας

- Μάρκα και μοντέλο του εξοπλισμού που παρήγαγε την εικόνα
- Έκδοση λογισμικού που παρήγαγε την εικόνα
- Δημιουργός της εικόνας (καλλιτέχνης)
- Κάτοχος πνευματικών δικαιωμάτων

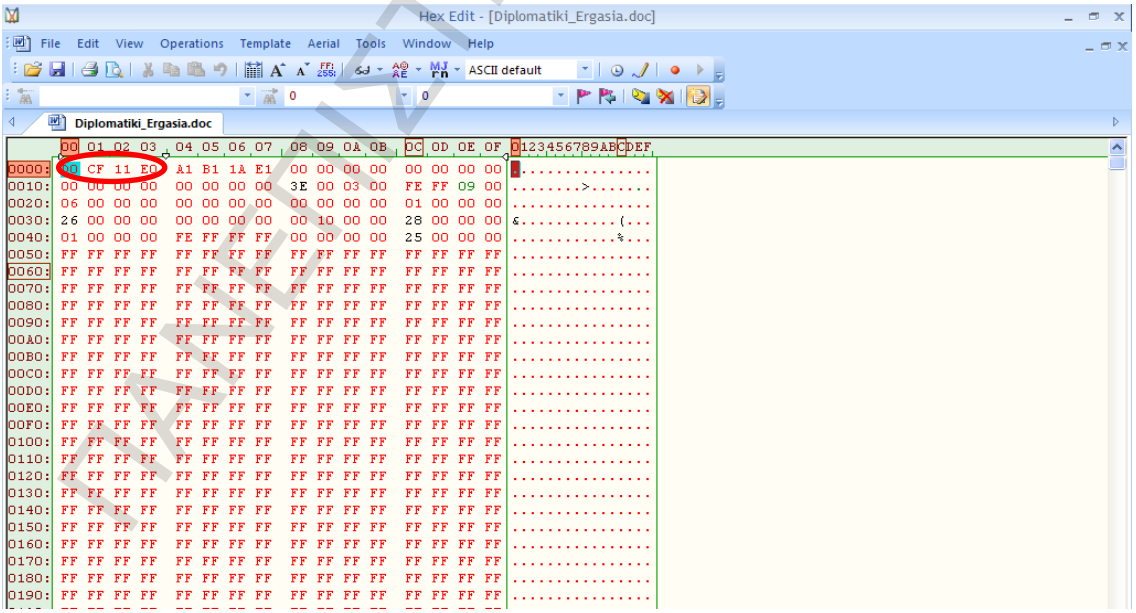
4.7 Κεφαλίδες αρχείων

Για να γίνει αντιληπτό αν η επέκταση ενός αρχείου έχει αλλοιωθεί, θα πρέπει να κατανοήσουμε τον τρόπο, με τον οποίο τα αρχεία αναγνωρίζονται από τα λειτουργικά συστήματα και την εφαρμογή λογισμικού. Ένα πρόγραμμα εφαρμογής, γενικότερα, αναγνωρίζει ένα αρχείο είτε από την κεφαλίδα του αρχείου ή την επέκτασή του, ενώ τα λειτουργικά συστήματα έχουν την τάση να βασίζονται κυρίως στην επέκταση του αρχείου για να καθορίσουν τον τύπο του αρχείου .

Μια κεφαλίδα αρχείου είναι συνήθως μια ακολουθία χαρακτήρων στην αρχή ενός αρχείου που υποδηλώνει το είδος του αρχείου που πραγματικά είναι. Κυριολεκτικά υπάρχουν χιλιάδες διαφορετικοί τύποι αρχείων. Έτσι, η εύρεση κεφαλίδας αρχείου μπορεί να είναι μια πρόκληση, αν το αρχείο έχει δημιουργηθεί από ένα ασαφές πρόγραμμα.

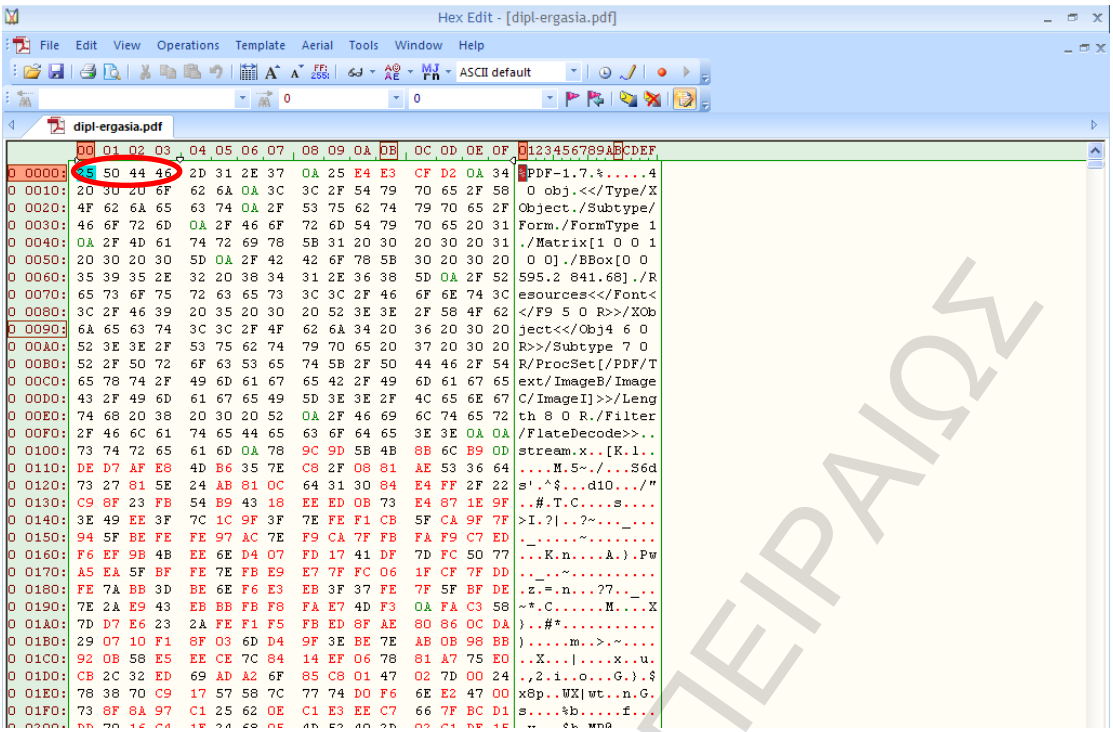
Ευτυχώς, τα περισσότερα αρχεία ανοίγουν με τα γνωστά πακέτα λογισμικού, όπως της Microsoft, της Novell, της Adobe , της Sun, κ.λπ.

Στο παρακάτω σχήμα φαίνεται μια κεφαλίδα ενός αρχείου Microsoft Word. Η ακολουθία *DO CF 11 EO* για το αρχείο αυτό είναι πάντα η ίδια, ακόμη και αν αλλάξει η επέκταση αρχείου. Η κεφαλίδα του συγκεκριμένου αρχείου ανήκει σε αρχείο με επέκταση *.doc*.

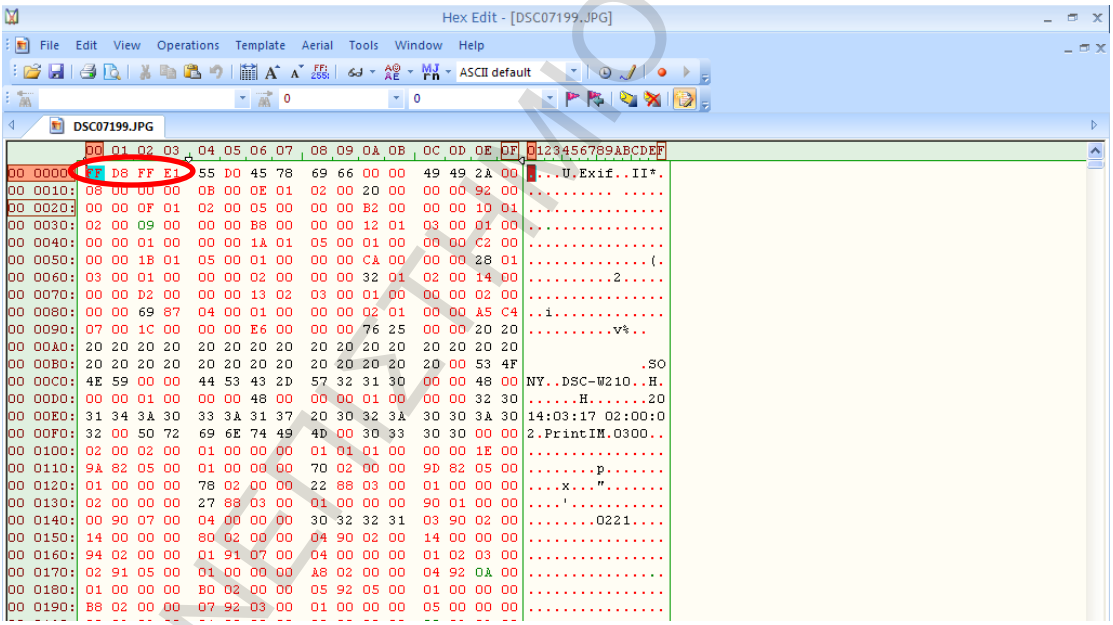


Εικόνα 17[η]: Κεφαλίδα αρχείου *doc σε* hex editor

Στο αμέσως επόμενο παράδειγμα, φαίνεται η κεφαλίδα ενός αρχείου PDF και μιας εικόνας JPG, σύμφωνα με τον αναλυτή *Hex Editor*.



Εικόνα 18η: Κεφαλίδα αρχείου *pdf* σε hex editor



Εικόνα 19η: Κεφαλίδα αρχείου JPG σε hex editor

Παρακάτω φαίνονται οι κεφαλίδες γνωστών τύπων αρχείων.

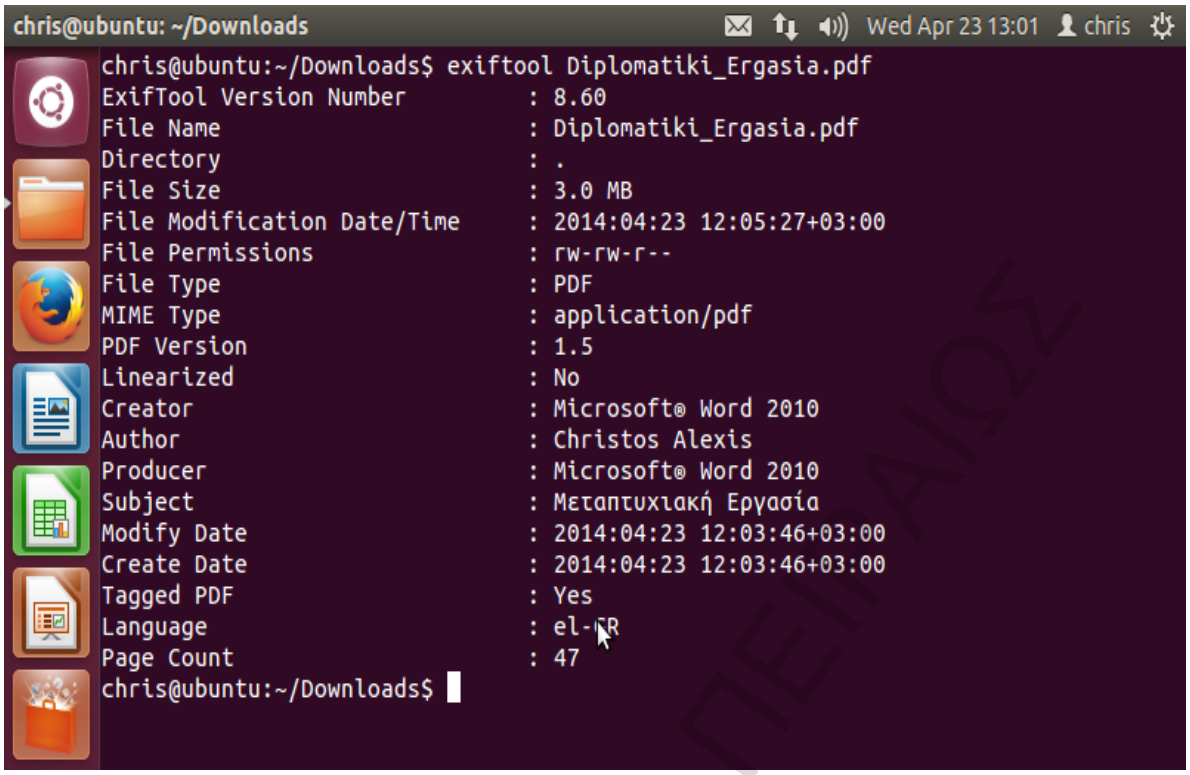
Κεφαλίδα	ISO 8859-1	Offset	Επέκταση Αρχείου	Περιγραφή
30 26 B2 75 8E 66 CF 11 A6 D9 00 AA 00 62 CE 6C	0&²u.fÏ. ¦Ù.ª.bÎl	0	asf, wma, wmv	Advanced Systems Format[8]
D0 CF 11 E0			doc	Microsoft Office
66 4C 61 43	fLaC	0	flac	Free Lossless Audio Codec[
46 4F 52 4D nn nn nn nn 46 54 58 54	FORM....FTXT	0, any	ftxt, txt, iff	IFF Formatted Text
47 49 46 38 37 61 47 49 46 38 39 61	GIF87a GIF89a	0	gif	Graphics Interchange Format
00 00 01 00	....	0	ico	ICO
43 44 30 30 31	CD001	0x8001, 0x8801 or 0x9001	iso	ISO9660
FF D8 FF	ÿØÿà	0	jpg, jpeg	JPEG
4D 54 68 64	MThd	0	mid, midi	MIDI sound file[12]
FF FB	˙ű	0	mp3	MPEG-1 Layer 3 ID3v1
49 44 33	ID3	0	mp3	MP3 file with an ID3v2 container
00 00 00 nn 66 74 79 70 33 67 70 35	....ftyp 3gp5	0	MP4	MPEG-4 αρχεία video
25 50 44 46	%PDF	0	pdf	PDF;
89 50 4E 47 0D 0A 1A 0A	.PNG....	0	png	Portable Network Graphics
25 21 50 53	%!PS	0	ps	PostScript document

Κεφαλίδα	ISO 8859-1	Offset	Επέκταση Αρχείου	Περιγραφή
38 42 50 53	8BPS	0	psd	Adobe Photoshop's
52 61 72 21 1A 07 01 00	Rar!....	0	rar	RAR
52 61 72 21 1A 07 00	Rar!...	0	rar	RAR
52 49 46 46 nn nn nn nn 57 41 56 45	RIFF....WAVE	0	wav	Waveform Audio File Format
1F A0	..	0	z, tar.z	Συμπιεσμένο αρχείο (συνήθως tar zip) με χρήση Αλγορίθμου LZH
50 4B 03 04, 50 4B 05 06 (empty	PK..	0	zip, jar, odt, ods, odp, docx, xlsx, pptx, apk	zip file format και formats όπως των: JAR, ODF, OOXML

5 Τροποποίηση Μεταδεδομένων

Υπάρχουν αρκετοί τρόποι να τροποποιήσεις τα μεταδεδομένα ενός αρχείου, τα οποία εξήχθησαν με κάποιο εργαλείο. Στο παρακάτω παράδειγμα θα αναφερθούμε στον τρόπο με τον οποίο αλλάζουμε τις πληροφορίες που παρέχονται από τα μεταδεδομένα, κατά το δοκούν.

Αν από τον παρακάτω πίνακα ανακτήσουμε τα μεταδεδομένα του αρχείο *FINAL.pdf*, τότε παίρνουμε τις εξής πληροφορίες:



Εικόνα 20η: Μεταδεδομένα αρχείου FINAL.pdf

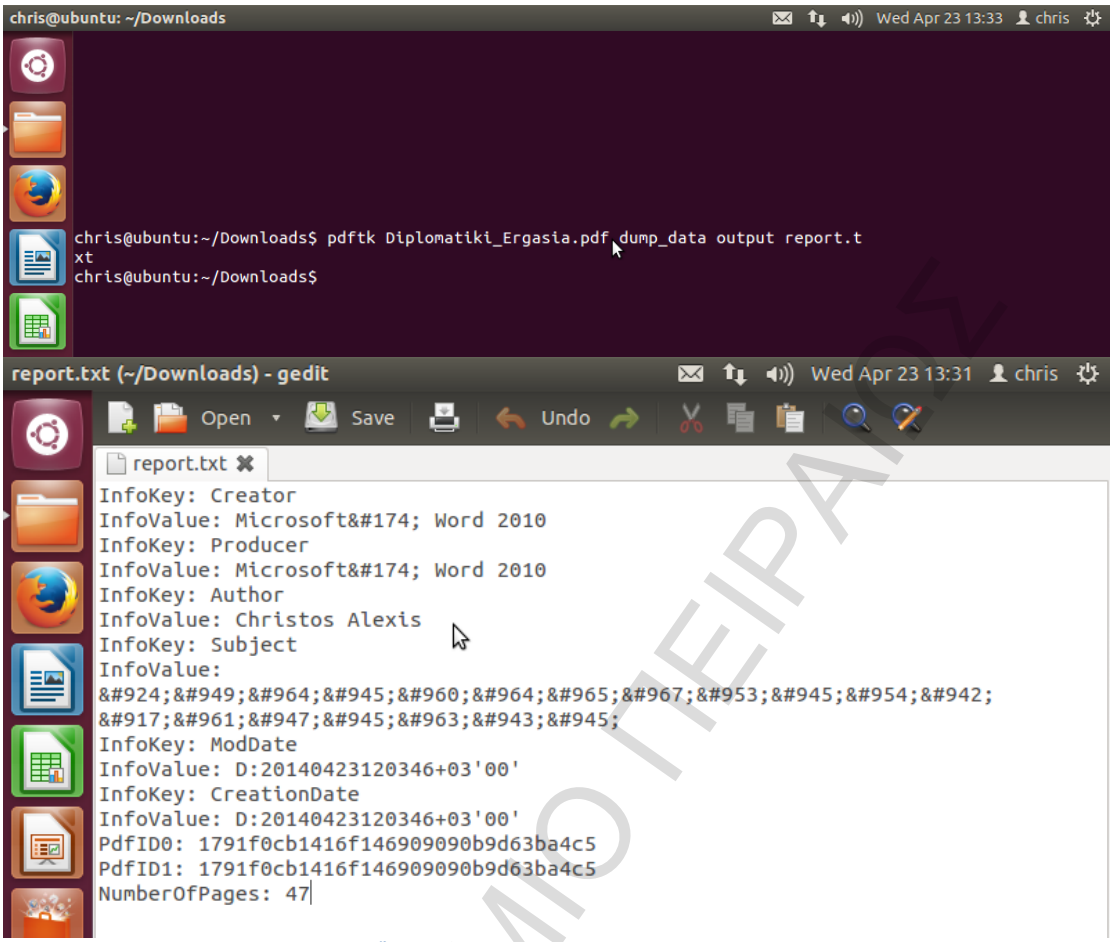
Αυτό που επιχειρείται στη συνέχεια είναι η εξαγωγή μεταδεδομένων σε ένα αρχείο, π.χ., στο *report.txt*. Κάτι τέτοιο γίνεται, προκειμένου να τροποποιήσουμε τα δεδομένα που εμπεριέχονται στο κείμενο.

Η εντολή που χρησιμοποιούμε σε περιβάλλον Linux είναι η εξής:

pdftk Diplomatiki_Ergasia.pdf dump_data output report.txt

Στο σημείο αυτό πρέπει να ειπωθεί ότι το εργαλείο που χρησιμοποιείται στην περίπτωση τροποποίησης μεταδεδομένων (σε αρχεία PDF) είναι το *pdftk*, το οποίο είναι μια διασυστημική πλατφόρμα ανοιχτού κώδικα, με δυνατότητα διαχωρισμού, ένωσης, κρυπτογράφησης, κ.λπ., των αρχείων pdf.

Με την ανωτέρω εντολή, αποτυπώνονται τα μεταδεδομένα του αρχείου *Diplomatiki_Ergasia.pdf* στο αρχείο *report.txt*.



Εικόνα 21η: Αποτύπωση μεταδεδομένων σε αρχείο *txt*

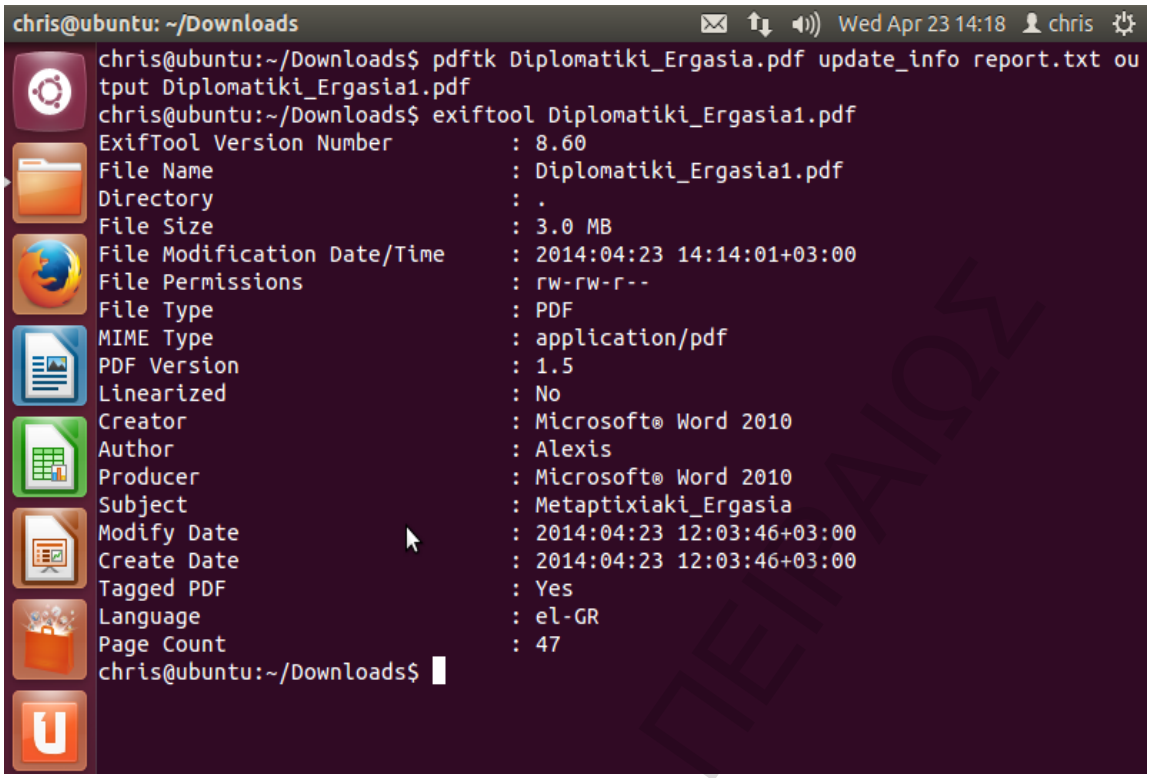
Στο παραπάνω αρχείο *report.txt*, μπορούμε να τροποποιήσουμε όλα σχεδόν τα πεδία του, αντικαθιστώντας τα με οποιαδήποτε επιθυμητή τιμή.

Αφού τροποποιήσουμε, για παράδειγμα, το πεδίο του *Author*, από *Christos Alexis* σε *Alexis* και το *Subject* από *Diplomatiki_Ergasia* σε *Metaptixiaki_Ergasia*, ενημερώνουμε το αρχείο μας με τα καινούργια πλέον μεταδεδομένα.

Η εντολή που εκτελείται είναι η εξής:

pdftk Diplomatiki_Ergasia.pdf update_info report.txt output Diplomatiki_Ergasia1.pdf

Τα αποτελέσματα της τροποποίησης των μεταδεδομένων του αρχείου μάς εμφανίζουν τα παρακάτω μεταδεδομένα:



Εικόνα 22η: Αρχείο με τροποποιημένα μεταδεδομένα

6 Extract και Libextractor - Ανάγνωση μεταδεδομένων

Οι σύγχρονες μορφές αρχείων περιλαμβάνουν όρους σχετικά με το σχολιασμό του περιεχόμενου του αρχείου με περιγραφικές πληροφορίες. Η εξέλιξη αυτή οδηγείται από την ανάγκη να βρεθεί ένας καλύτερος τρόπος για να οργανωθούν τα δεδομένα, από το να γίνεται απλώς χρήση του ονόματος αρχείου. Το πρόβλημα με αυτά τα μεταδεδομένα είναι ότι δεν αποθηκεύονται με έναν τυποποιημένο τρόπο σε διαφορετικές μορφές αρχείων. Αυτό το καθιστά δύσκολο για τα εργαλεία του τύπου *format-agnostic*, όπως διαχειριστές αρχείων ή οι *file-sharing* εφαρμογές, να κάνουν χρήση των πληροφοριών. Συντελεί, επίσης, σε μια πληθώρα εργαλείων συγκεκριμένης μορφής, που χρησιμοποιούνται για την εξαγωγή των μεταδεδομένων, όπως τα: AVInfo, id3edit, jpeginfo και Vocoditor.

Στη συνέχεια, αναλύεται η βιβλιοθήκη *Libextractor* και το εργαλείο *Ectract.* Ο στόχος του σχεδίου *Libextractor* είναι να παρέχει μια ενιαία διεπαφή για τη λήψη μεταδεδομένων από διαφορετικές μορφές αρχείων.

Η libextractor επιτυγχάνει αυτές τις πληροφορίες χρησιμοποιώντας ειδικό κωδικό parser για πολλά δημοφιλή πρότυπα. Ο κατάλογος, προς το παρόν, περιλαμβάνει αρχεία MP3, Ogg, Real Media, MPEG, RIFF (avi), GIF, JPEG, PNG, TIFF, HTML, PDF, PostScript, Zip, OpenOffice.org, StarOffice, Microsoft Office, tar, DVI, man, Deb , elf, RPM, ASF, καθώς επίσης και γενικές μεθόδους, όπως η ανίχνευση τύπου MIME. Υπάρχουν πολλά άλλα πρότυπα, και από τα πιο δημοφιλή, μόνο λίγα αποκλειστικά (*proprietary*) πρότυπα δεν υποστηρίζονται.

Το να ενσωματώσει κανείς υποστήριξη για νέα πρότυπα, είναι εύκολο, γιατί η libextractor χρησιμοποιεί plug-ins για τη συλλογή δεδομένων. Τα plugins της libextractor είναι κοινόχρηστες βιβλιοθήκες που παρέχουν συνήθως κώδικα για να αναλύσει ένα συγκεκριμένο πρότυπο.

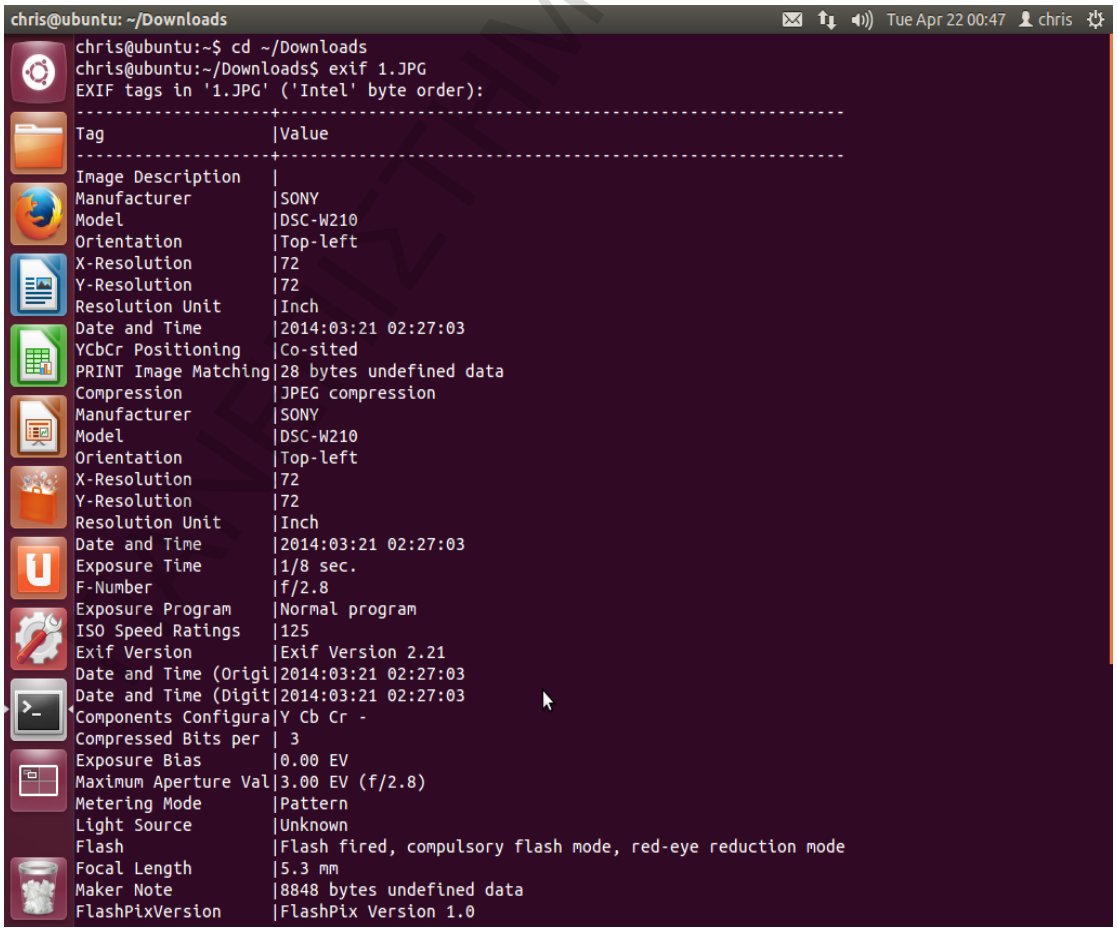
Η libextractor συγκεντρώνει τα μεταδεδομένα που λαμβάνονται από διάφορα plugins και παρέχει στους πελάτες μια λίστα από ζεύγη, που αποτελείται από μία ακολουθία ταξινόμησης και χαρακτήρων. Η ταξινόμηση χρησιμοποιείται για να οργανώσει τα μεταδεδομένα σε κατηγορίες, όπως ο τίτλος, ο δημιουργός, το θέμα και την περιγραφή.

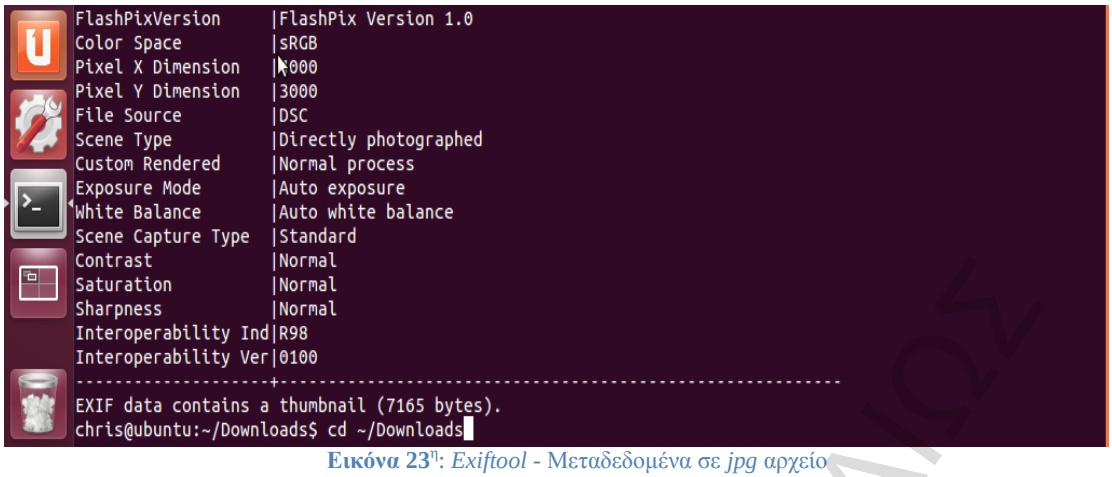
6.1 Πρόσθετα προγράμματα (Plug-ins) εξαγωγής μεταδεδομένων.

6.1.1 JPEG

Όπως επισημάνθηκε και προηγουμένως, το jpeg είναι το πιο δημοφιλές είδος αρχείου σήμερα, όσον αφορά την αναπαράσταση ψηφιακών εικόνων. Η εξαγωγή μεταδεδομένων από τις εικόνες είναι σημαντικό για τους ερευνητές. Ένα από τα πιο γνωστά προγράμματα εξαγωγής μεταδεδομένων σε αρχεία JPEG είναι το πρόγραμμα *exif*.

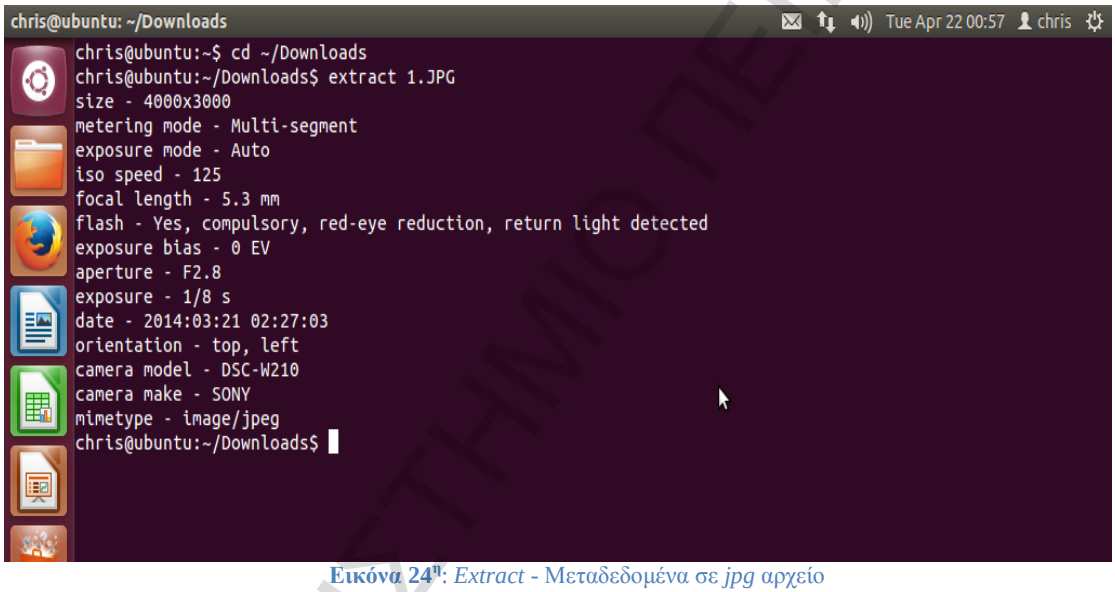
Στο παρακάτω σχήμα αποτυπώνονται τα μεταδεδομένα που εξάγονται με τη βοήθεια του προγράμματος *exif*.





Εικόνα 23η: *Exiftool* - Μεταδεδομένα σε *jpg* αρχείο

Κάνοντας τις ίδιες ενέργειες, με το πρόγραμμα *extract* αυτή τη φορά, παίρνουμε τα παρακάτω αποτελέσματα:



Εικόνα 24η: *Extract* - Μεταδεδομένα σε *jpg* αρχείο

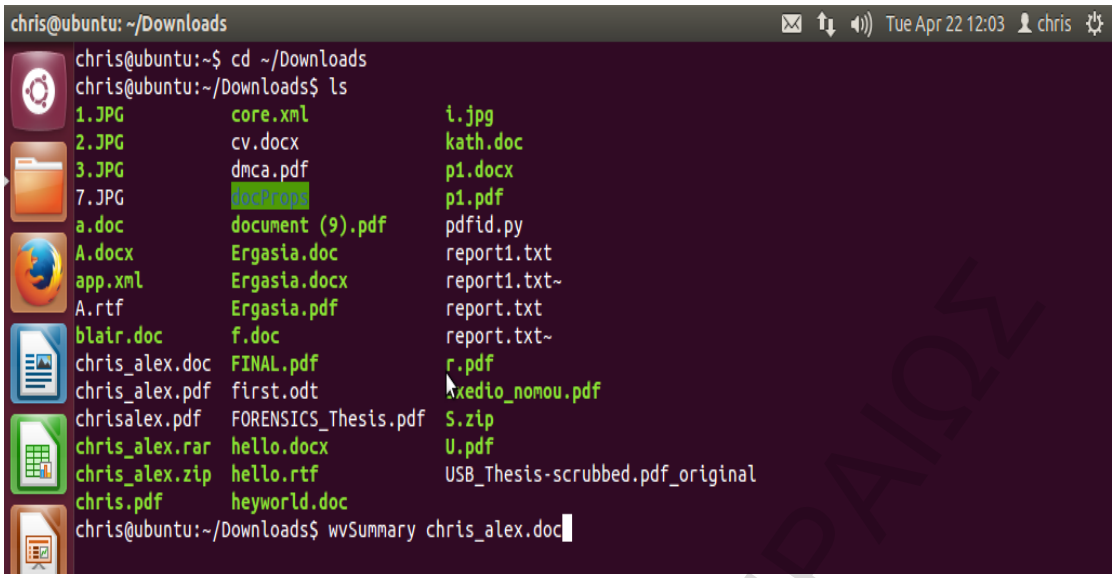
Είναι εμφανής η διαφορά ότι στην πρώτη περίπτωση, το πρόγραμμα *exif* διεισδύει σε βάθος και προσφέρει μία πληθώρα από μεταδεδομένα, σε αντίθεση με το πρόγραμμα *extract* το οποίο, κυρίως, αναφέρεται σε άλλου τύπου έγγραφα.

6.1.2 MS Office '97-2003

Προκειμένου να εξαχθούν μεταδεδομένα από έγγραφα του MS Office '97-2003, γίνεται χρήση – κυρίως – ενός πρόσθετου προγράμματος (*plug-in*), του *wvSummary*, το οποίο χαρακτηρίζεται ως βοηθητικό πρόγραμμα του *wvWare*. Το *wvWare*, με τη σειρά του, είναι ένας συντακτικός αναλυτής (*parser*), γραμμένο σε γλώσσα Python.

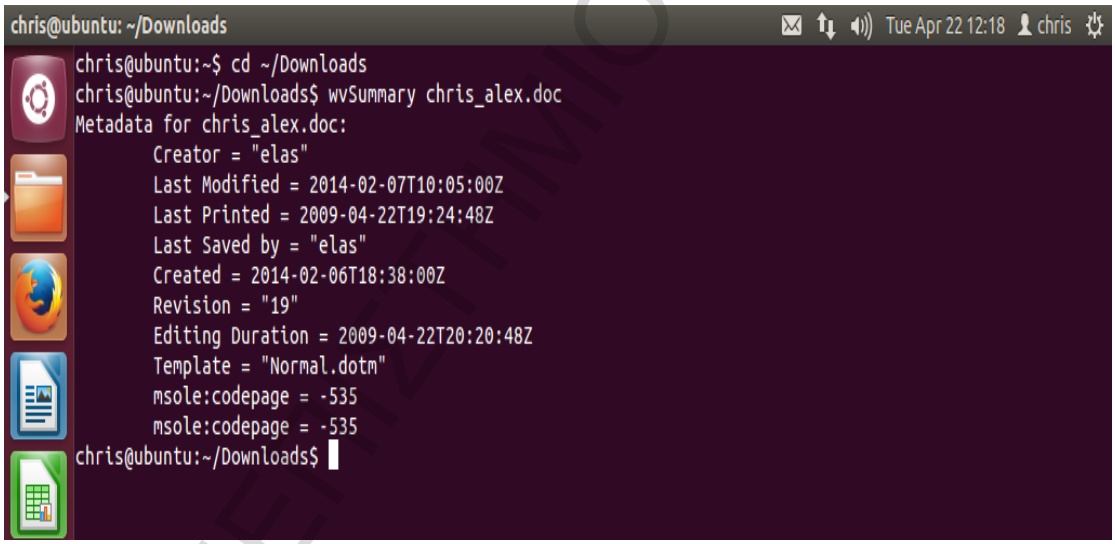
Το πρόσθετο αυτό πρόγραμμα, χρησιμοποιείται για την εξαγωγή μεταδεδομένων από αρχεία Office Open XML, το εσωτερικό πρότυπο του Microsoft Office 2007 (Windows) και 2008 (Macintosh).

Από τα αρχεία που διαθέτουμε στον κατάλογό μας (Εικόνα 25η),



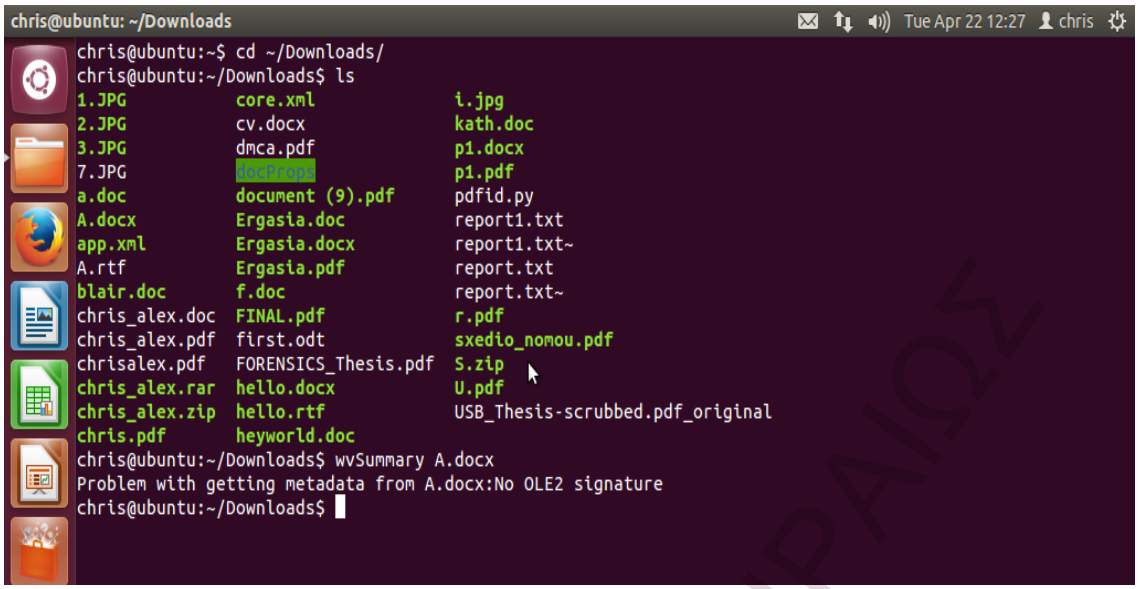
Εικόνα 25η: Αρχεία καταλόγου

κάνουμε αναζήτηση μεταδεδομένων, με τη χρήση του βοηθητικού προγράμματος *wvSummary*. Έτσι, τα μεταδεδομένα που λαμβάνουμε από το έγγραφο *chris_alex.doc* είναι τα εξής:



Εικόνα 26η: *wvSummary* - Μεταδεδομένα αρχείου *doc*

Κάνοντας τις ίδιες ενέργειες σε αρχεία της σουίτας του Micrsoft Office 2007 και μεταγενέστερα, διαπιστώνουμε ότι η εξαγωγή μεταδεδομένων σε τέτοια αρχεία είναι αδύνατη, αφού τα αρχεία αυτά δεν ανήκουν στην κατηγορία των αρχείων σύνδεσης και ενσωμάτωσης αντικειμένων (*Object Linking and Embedding Files*).



Εικόνα 27η: *wvSummary* - Μεταδεδομένα μόνο σε αρχεία τύπου OLE

7 EnCase

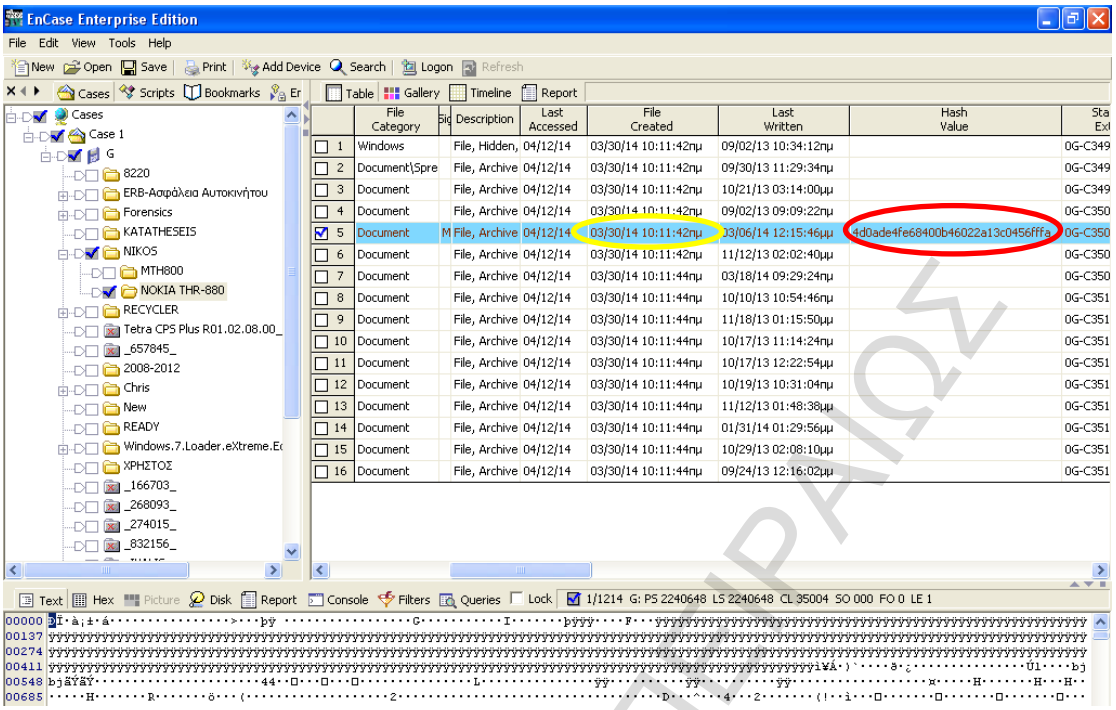
Το EnCase είναι ένα δημοφιλές και πολύτιμο εργαλείο στον τομέα της ανάλυσης ψηφιακών πειστηρίων. Υπάρχουν τέσσερις κλάσεις αρχείων ενδείξεων (*evidence files*), που υποστηρίζονται από τις εφαρμογές του EnCase.

- EnCase αρχεία ενδείξεων
- Λογικά αρχεία ενδείξεων
- Ακατέργαστες εικόνες (*Raw Images*)
- Ατομικά αρχεία (συμπεριλαμβανομένων και των καταλόγων)

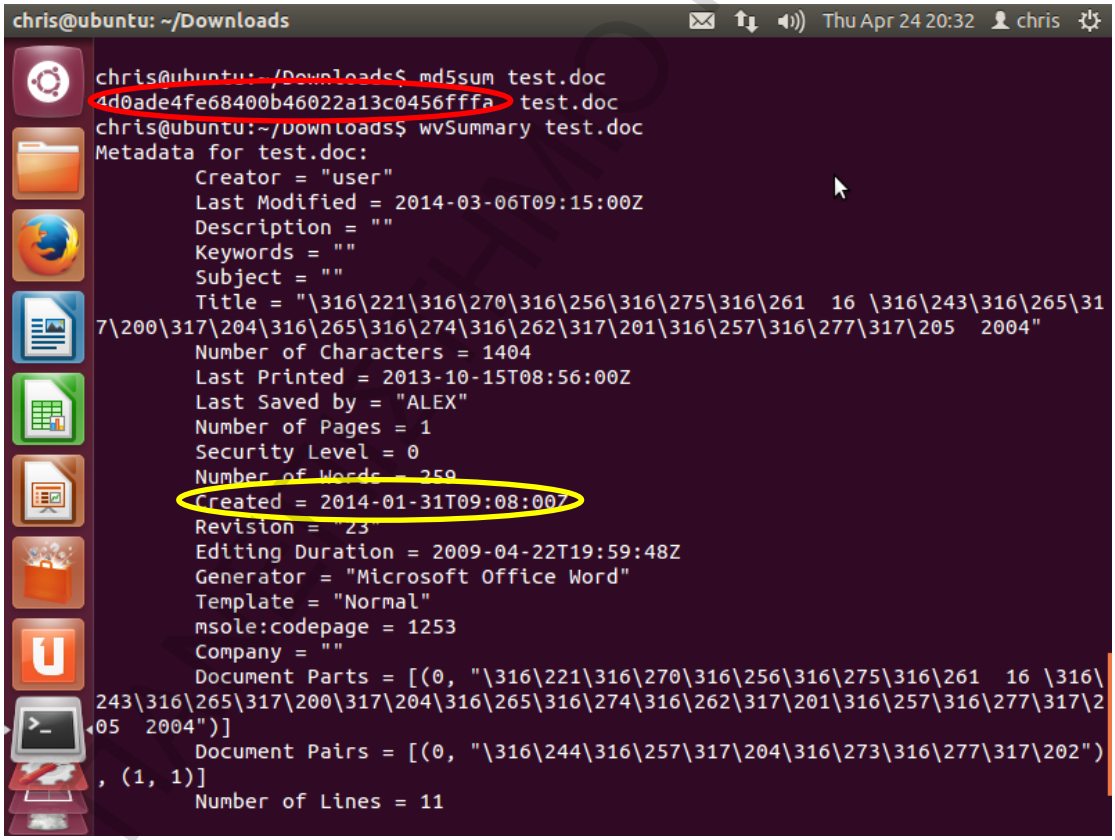
Η πρώτη και η τρίτη κλάση είναι οι πλέον ενδιαφέρουσες από πλευράς μεταδεδομένων. Τα αρχεία ενδείξεων περιλαμβάνουν τα περιεχόμενα μιας αποκτηθείσας συσκευής και παρέχουν τη βάση για μελλοντική ανάλυση. Αυτά τα αρχεία ενσωματώνουν μεταδεδομένα προς έρευνα, την τιμή επιπέδου-*hash* της συσκευής και το περιεχόμενο από τη συσκευή. Αντιθέτως, τα αρχεία ακατέργαστων εικόνων δεν περιλαμβάνουν μεταδεδομένα ή τιμές *hash*. Τα ατομικά στοιχεία (ή κατάλογοι) περιλαμβάνουν τα υπάρχοντα μεταδεδομένα και μπορούν να προστεθούν στην ανοιχτή υπόθεση του ερευνητή.

Στο εργαλείο *EnCase* προστέθηκε υποστήριξη για τα αρχεία του Microsoft Office 2007 (*Word, Excel, PowerPoint*). Το *EnCase* έχει τη δυνατότητα να εξαγάγει τιμές σε φύλλα εργασίας *Excel*, που έχουν δημιουργηθεί σε αρχείο προτύπου Office Open XML.

Δοκιμάζοντας το *EnCase*, διαπιστώθηκε ότι, εκτός του γεγονότος ότι μπορεί να εξαγάγει κείμενο από έγγραφα του Microsoft Office 2007, είναι επίσης δυνατή η εξαγωγή κειμένου από ενσωματωμένα (*embedded*) αρχεία μέσα σε αρχεία του Microsoft Office 2007. (20)



Εικόνα 28η: EnCase – Μεταδεδομένα εγγράφου *test.doc*



Εικόνα 29η: Linux – Μεταδεδομένα εγγράφου *test.doc*

Αυτό που παρατηρείται και συμπεραίνεται από τις δύο εικόνες είναι ότι κατά τον έλεγχο του ιδίου αρχείου, την πρώτη φορά με το εργαλείο *EnCase* και τη δεύτερη σε πλατφόρμα Linux, τα μεταδεδομένα που εξήχθησαν δε συμφωνούν σε όλες τις πληροφορίες. Έτσι, παρατηρούμε ότι όντως πρόκειται για το ίδιο αρχείο, μια και η

συνάρτηση κατακερματισμού *MD5* βγάζει το ίδιο αποτέλεσμα, τόσο με το *EnCase* όσο και με Linux

Ένα σημαντικό εργαλείο που περιέχεται στο *EnCase* είναι ο Επεξεργαστής Υπόθεσης (*Case Processor*). Ο επεξεργαστής αυτός επιτρέπει στους χρήστες να «τρέξουν» μία ή περισσότερες *EnScript* [εσωτερική γλώσσα δέσμης ενεργειών (*scripting language*)] αυτόνομες λογισμικές μονάδες (*modules*), σε μία ανοιχτή υπόθεση. Από τη σκοπιά των μεταδεδομένων, οι πιο εφαρμόσιμες αυτόνομες λογισμικές μονάδες περιλαμβάνουν τα εξής:

- Εφαρμογή προβολής EXIF (*EXIF Viewer*)
- Συντακτικός αναλυτής πληροφοριών ενεργού καταλόγου (*Active Directory Information Parser*)
- Ανιχνευτής αρχείων (*File Finder*)
- Αναφορά αρχείων (*File Report*)
- Ανιχνευτής προστατευμένων αρχείων (*Find Protected Files*)
- Συντακτικός αναλυτής HTML (*HTML Carver*)

Τα χαρακτηριστικά του *EnCase*, όπως οι αυτόνομες λογισμικές μονάδες (*modules*) που προσδιορίστηκαν πιο πάνω, η βοήθεια στη συλλογή των μεταδεδομένων, όπως:

- το όνομα
- το φίλτρο
- ο τύπος αρχείου
- το αρχείο κατηγορίας
- η υπογραφή
- η περιγραφή
- το διαγραμμένο αρχείο
- ο χρόνος τελευταίας προσπέλασης
- η ημερομηνία/ώρα δημιουργίας
- η τελευταία χρονοσφραγίδα
- η χρονοσφραγίδα τροποποίησης εισόδου.

Αρκετοί συνιστούν τον υπολογισμό των *MD5 hashes* των αρχείων, προτού πριν την εξαγωγή των μεταδεδομένων, για την παροχή ενός μέσου επαλήθευσης της ακεραιότητας των αρχείων. Αφού αρχικά υπολογισθούν οι *hashes*, ο χρήστης μπορεί να επιλέξει από μια λίστα επιλογών, αναφορικά με το ποια μεταδεδομένα να εξαγάγει και πού να εξαγάγει τα αποτελέσματα.

8 Sleuthkit (TSK)

Το πιο δημοφιλές εργαλείο ανοιχτού κώδικα στη δικανική πληροφορική είναι το *Sleuthkit* και το *Autopsy*. Το πρόγραμμα περιήγησης του Sleuthkit (TSK) και του *Autopsy* είναι εργαλεία που βασίζονται σε Unix. (21)

Το TSK είναι πολυ-σύνθετο εργαλείο, με περισσότερα από 20 εργαλεία γραμμής εντολών να χρησιμοποιούνται για την ανάλυση του δίσκου και το εικόνων συστήματος αρχείων, προκειμένου να συλλεγούν αποδείξεις . Το *Autopsy* είναι ένα front-end πρόγραμμα περιήγησης για το TSK και δημιουργήθηκε για να καταστήσει τη διαδικασία ανάλυσης ευκολότερη για τον χρήστη.

Εκτός από την ικανότητα του TSK για ανάλυση εικόνων, που βρίσκονται σε ένα δίσκο, ένα από τα πρωταρχικά του δυνατά σημεία είναι οι βιβλιοθήκες, οι οποίες διατίθενται προς χρήση, από προγραμματιστές εργαλείων .

Το TSK αποτελείται από τέσσερις λογικές βιβλιοθήκες. Η πρώτη βιβλιοθήκη αναφέρεται στο πρότυπο αρχείου ειδώλου του δίσκου. Αυτή η βιβλιοθήκη παρέχει μια αφαίρεση στις διάφορες μορφές αρχείων, όπως είναι το πρότυπο του Expert Witness, το πρότυπο του Advanced Forensic (AFF) και του ακατέργαστου προτύπου. Το πρότυπο αρχείου ειδώλου του δίσκου παρουσιάζει μια διεπαφή μόνο προς ανάγνωση σε αρχεία ειδώλου του δίσκου και επιτρέπει στα προγράμματα να διαβάζουν δεδομένα από αυθαίρετες θέσεις στο αρχείο, χωρίς να χρειάζεται να γνωρίζουμε τι πρότυπο χρησιμοποιείται.

Η επόμενη βιβλιοθήκη είναι η βιβλιοθήκη συστήματος (διαχείριση των μέσων), η οποία αναλύει τα διάφορα είδη των διαμερισμάτων (*partitions*) που υποστηρίζει το TSK. Η βιβλιοθήκη συστήματος τόμου έχει δύο σημαντικές διεπαφές. Η πρώτη είναι μια λειτουργία για να ανοίξει ένα αρχείο ειδώλου δίσκου και να εντοπίσει τον τύπο του συστήματος τόμου. Η δεύτερη είναι μια συνάρτηση «*walk*» που προσδιορίζει τον τόμο σε ένα δίσκο και επεξεργάζεται μια λειτουργία επανάκλησης για κάθε τόμο.

Η τρίτη βιβλιοθήκη είναι για τα εργαλεία του συστήματος αρχείων και είναι η μεγαλύτερη. Ο σχεδιασμός είναι παρόμοιος με τη βιβλιοθήκη συστήματος τόμου, στο ότι ένα πρόγραμμα ανοίγει το αρχείο ειδώλου δίσκου ή διαμερίσματος και μετά προσπελάζει συναρτήσεις«*walk*» επιπέδου δεδομένων, μεταδεδομένων, ονόματος αρχείου, περιεχόμενου αρχείου και ημερολογίου.

Η τελική βιβλιοθήκη είναι η βιβλιοθήκη του συστήματος αρχείων που ερμηνεύει δομές του συστήματος αρχείων και επιτρέπει την ανάκτηση πληροφοριών καταλόγου και συστήματος αρχείων.

8.1 Εξαγωγή μεταδεδομένων στο TSK

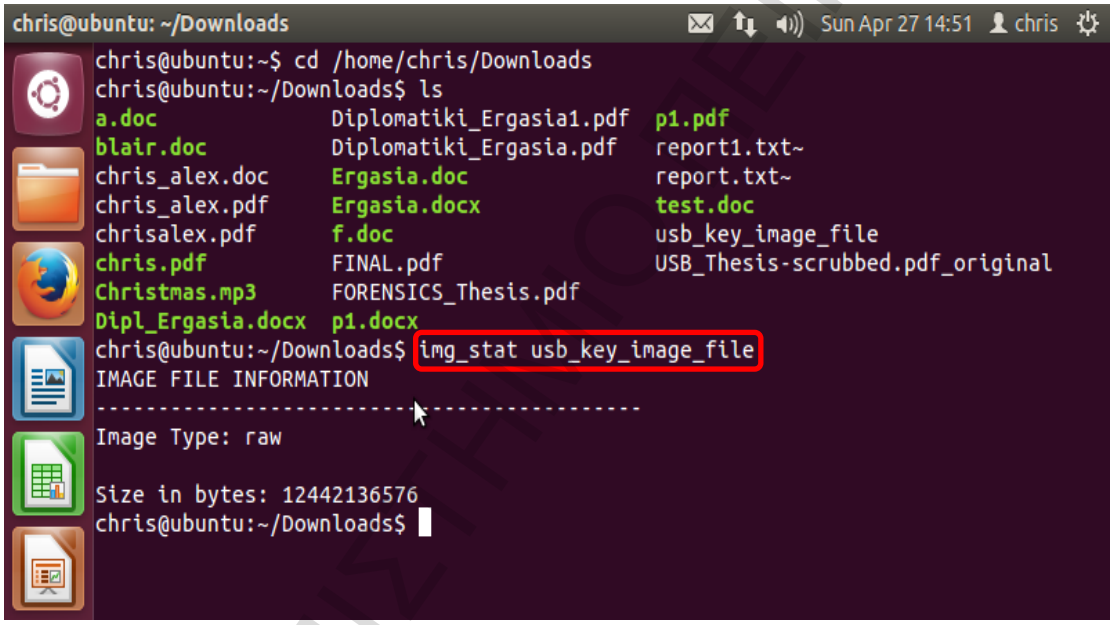
Το TSK δεν έχει κανένα εργαλείο που να επεξεργάζεται μεταδεδομένα, αλλά έχει μία πληθώρα εργαλείων που μπορούν να επεξεργασθούν μεταδεδομένα συστήματος αρχείων, όπως: το *istat, ils, ifind* και *icat.*

Το εργαλείο *istat* παρέχει πληροφορίες που αφορούν το μέγεθος του αρχείου, δεδομένα χρόνου, πεδία δικαιωμάτων και διευθύνσεις δεσμευμένων μονάδων δεδομένων.

Το εργαλείο *ils* αποτυπώνει λεπτομέρειες πολλών δομών μεταδεδομένων.

Το *ifind*, με τη σειρά του, χρησιμοποιείται όταν μια μονάδα δεδομένων περιέχει αποδεικτικά στοιχεία με δυνητική τιμή και παρέχει επιλογές για την αναζήτηση όλων των εγγραφών μεταδεδομένων και, επίσης, για την εύρεση της εγγραφής μεταδεδομένων που μας υποδεικνύει ένα συγκεκριμένο όνομα αρχείου.

Τέλος, το *icat* επιτρέπει τα περιεχόμενα κάθε αρχείου να προβληθούν, χρησιμοποιώντας τη διεύθυνση μεταδεδομένων, σε αντίθεση με το όνομα του αρχείου. Το κύριο όφελος του *icat* είναι ότι τα αρχεία που δεν έχουν ένα όνομα αρχείου να υποδεικνύει την εγγραφή των μεταδεδομένων τους (αδέσμευτα αρχεία), μπορούν να ανακτηθούν και να προβληθούν.



Εικόνα 30η: Εξαγωγή μεταδεδομένων με το εργαλείο *istat*

9 Σύγκριση αποτελεσμάτων

Πραγματοποιήθηκε δοκιμή σε έγγραφο του Word (*Microsoft Office 2007*), όπου το *EnCase* περιείχε ενσωματωμένα αρχεία επιπέδου τρία (*three layers deep*), και ενός εγγράφου Word (*Microsoft Office 2003*), που περιείχε ενσωματωμένα αρχεία τετάρτου επιπέδου (*four layers deep*). Και στις δύο περιπτώσεις, το *EnCase* ήταν σε θέση να εξαγάγει το κείμενο, σε αντίθεση με άλλα εργαλεία ανοιχτού κώδικα, όπως αυτά που δοκιμάσθηκαν παραπάνω (*wvSummary*, *wvText*).

Επιπρόσθετα από τα αρχεία του Microsoft Office 2007, το *EnCase* παρέχει λειτουργικότητα να δει κανείς ατομικά στοιχεία άλλων σύνθετων τύπων αρχείων,

όπως τα αρχεία μητρώου, τα αρχεία *OLE*, τα αρχεία *MS Outlook email, Windows Thumbs.db* και αρχεία *Macintosh PAX*.

10 Σύνοψη

Κατά την εκπόνηση της παρούσας εργασίας, εξήχθησαν – εκτός από μεταδεδομένα – και αρκετά σημαντικά συμπεράσματα.

Έτσι, τα αρχεία εγγράφων (προ - Microsoft Office 2007, Microsoft Office 2007, και Open Office) βρέθηκαν να περιέχουν μεταδεδομένα που θα ήταν ενδιαφέροντα από τη σκοπιά ενός ερευνητή ψηφιακών πειστηρίων. Παρόλο που τα αρχεία πολυμέσων περιελάμβαναν, επίσης, μεταδεδομένα, οι διαθέσιμες πληροφορίες θα είναι λιγότερο χρήσιμες για τον ερευνητή, σε σύγκριση με τα μεταδεδομένα που υπάρχουν διαθέσιμα σε αρχεία εγγράφων.

Για παράδειγμα, τα μεταδεδομένα που σχετίζονται με ένα αρχείο mp3 (καλλιτέχνης, τίτλος, έτος), συγκρινόμενα με τα μεταδεδομένα που σχετίζονται με ένα αρχείο εγγράφου (δημιουργός του αρχείου, αριθμός αναθεώρησης, αυτός τροποποίησε το έγγραφο, περιγραφή) δεν είναι τόσο σχετικά σε μία έρευνα, στην επιστήμη της ψηφιακής εγκληματολογίας. Ωστόσο, ορισμένα αρχεία πολυμέσων, όπως τα αρχεία JPEG, πιθανώς να περιέχουν μεταδεδομένα, όπως τον κατασκευαστή της κάμερας και το σειριακό αριθμό, τα οποία ένας ερευνητής θα ενδιαφερόταν να τα αποκτήσει.

Ερευνώντας τόσο τα αρχεία του Open Document Format (ODF - Open Office) όσο και του Office Open XML, υπάρχει σαφές συμπέρασμα ότι τα διαθέσιμα μεταδεδομένα αυτών των αρχείων είναι πολύτιμα, αναφορικά με έρευνες στην επιστήμη της δικανικής πληροφορικής.

Ύστερα από επεξεργασία αρκετών εγγράφων, βρέθηκαν χρονοσφραγίδες σε έγγραφα τόσο του Open Office, όσο και του Office Open XML. Ωστόσο, οι χρονοσφραγίδες αυτές δεν ήταν πάντα ακριβείς .

Ελαττώματα στα εργαλεία εξαγωγής μεταδεδομένων

Ένας από τους στόχους αυτής της εργασίας ήταν και η πιστοποίηση του κατά πόσο ακριβή θα ήταν τα μεταδεδομένα που θα εξάγονταν, με τη βοήθεια εργαλείων ανοιχτού κώδικα. Κατά τη διάρκεια της επεξεργασίας διαφόρων εγγράφων, διαπιστώθηκαν κάποια ελαττώματα, σχετικά με τα εργαλεία που εξετάσαμε πιο πάνω.

Κατανοώντας την ύπαρξη αυτών των ελαττωμάτων, είναι πολύ σημαντικό για έναν ερευνητή μιας υπόθεσης, καθότι – μερικά αποτελέσματα – ίσως να απαιτηθεί περισσότερη έρευνα ή ανάλυση.

Χαρακτηριστικά, το εργαλείο w*vSummary* αποδείχθηκε χρήσιμο στην εξαγωγή των περισσότερων μεταδεδομένων στα έγγραφα του προ – 2007 Microsoft Office. Ωστόσο, παρουσιάστηκαν ορισμένες αντιφάσεις και ανακρίβειες, όπως για παράδειγμα, ο αριθμός σελίδων σε κάποια έγγραφα δεν ήταν ακριβής.

Επίσης, παρόλο που προστέθηκαν σχόλια μέσα σε έγγραφα, το εργαλείο w*vSummary* δεν κατάφερε ορισμένες φορές να τα εντοπίσει. Ωστόσο, όμως, εξήγαγε μεταδεδομένα που είχαν να κάνουν με *Λέξεις Κλειδιά, Τροποποίηση, Έλεγχος*, κ.λπ.

Επίσης, παρουσιάστηκαν και κάποια θέματα σχετικά με το εργαλείο *Libextractor*. Στις πρώτες δοκιμές, η ανάκτηση μεταδεδομένων ήταν περιορισμένη. Τα αρχεία

δοκιμής περιελάμβαναν αρχεία *.jpg, .html, .pdf*, κ.λπ. Ενώ θα έπρεπε να υπάρχει πληθώρα από μεταδεδομένα κατά την εξαγωγή τους, σύμφωνα με αυτά που υποστηρίζει ότι παρέχει το παραπάνω εργαλείο, ωστόσο, σε ορισμένες περιπτώσεις, δεν ήταν δυνατή η εξαγωγή βασικών μεταδεδομένων ενός αρχείου, π.χ., σχόλια ή τίτλος.

Βιβλιογραφία

1. **Cheong Kai Wee, Edith Cowan University.** Analysis of hidden data in NTFS file system .

2. **Casey, Eoghan.** *Windows Forensics Analysis.*

3. **DARSHANAM, PRAVEEN.** FORENSICS - Analyzing an Unknown Image .

4. https://ext4.wiki.kernel.org/index.php/Ext4_Disk_Layout. [Online]

5. **Fairbanks, Kevin D.** An analysis of Ext4 for digital forensics.

6. **Kernan, D.** Hidden Data in Electronic Documents.

7. **Carrier, Brian.** File System Forensic Analysis.

8. **Linda Volonino, Reynaldo Anzaldua.** Computer Forensics for Dummies.

9. **FEI, BENNIE KAR LEUNG.** DATA VISUALISATION IN DIGTIAL FORENSICS.

10. **http://www.crimelibrary.com/serial_killers/unsolved/btk/index_1.html.**

11. **http://www.casi.org.uk/discuss/2003/msg00457.html.**

12. **http://office.microsoft.com/.**

13. **http://support.microsoft.com/.**

14. **http://msdn.microsoft.com/.**

15. **http://www.sfgate.com/news/article/Suspected-creator-of-Melissa-virus-arrested-3090073.php.**

16. **Aashish Kumar Purohit, Naveen Hemrajani, Ruchi Dave.** Role of metadata in cyber forensic and status of Indian cyber law.

17. **Microsystems, Sun.** OpenOffice.org XML File Format 1.0 Technical Reference Manual, Version 2. [Online]

18. **Article, Microsoft Knowledgebase.** "How and why unique identifiers are created in Office documents," Revision: 2.2 January 24, 2007, http://support.microsoft.com/kb/222180/. [Online]

19. **Incorporated, Adobe Systems.** "Extensible Metadata Platform (XMP) homepage", http://www.adobe.com/products/xmp/. [Online]

20. **Bunting, Steve.** *EnCase Computer Forensics - The Official EnCE: EnCase Certified Examiner Study Guide, Second Edition.*

21. **Migletz, James.** AUTOMATED METADATA EXTRACTION.