ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΚΑΙ ΔΙΚΤΥΩΝ

"Διερεύνηση Ύπαρξης Κρυμμένου Περιεχομένου Βασισμένη στον Νόμο του Benford"

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΑΡΤΙΝΗ ΑΔΑΜΑΝΤΙΝΗ
ΖΑΧΑΡΗΣ ΑΛΕΞΑΝΔΡΟΣ

Εκπονήθηκε υπό την επίβλεψη των Καθηγητών:

Δρ. Ηλία Χούστη
Δρ. Γεωργίου Πάγκαλου
Δρ. Θεόδωρου Τρύφωνα

ΒΟΛΟΣ, ΙΑΝΟΥΑΡΙΟΣ 2010

Ευχαριστίες

Η μεταπτυχιακή εργασία αυτή αφιερώνεται σε αυτούς που μας στήριξαν και μας στηρίζουν τόσα χρόνια.
Θα θέλαμε να πούμε ένα ευχαριστώ στους ανθρώπους που ήταν πάντα εκεί για εμάς και στα δύσκολα και στα εύκολα.
Στους καθηγητές μας και ιδιαιτέρως στον κ.Χρήστο Ηλιούδη και τον Θ.Τρύφωνα για την ανιδιοτελή και αμέριστη βοήθεια που μας έχουν προσφέρει όλον αυτόν τον καιρό και για τις σημαντικές ακαδημαϊκές συμβουλές τους,καθώς επίσης τον κ.Χούστη και τον κ.Πάγκαλο.
Ιδιαίτερα θα θέλαμε να ευχαριστήσουμε τους φίλους μας Γιώργο και Μιχάλη που μας ανέχονται επιτυχώς ακόμα και σήμερα καθώς και την φίλη μας την Ιουλία.
Τέλος την αφιερώνουμε στις οικογένειες μας που μας στήριξαν ηθικά και υλικά τόσα χρόνια. Χωρίς την συνεισφορά και στήριξη των παραπάνω ανθρώπων,σήμερα τίποτα απο όλα αυτά δεν θα ήταν δυνατό.

ΜΕΡΟΣ ΙΙΙ - ΕΦΑΡΜΟΓΗ

Κεφάλαιο 7 – Εφαρμογή Προτεινόμενης Μεθοδολογίας Στεγανάλυσης

7.1 Αναδόμηση αρχείων.
7.2 Εφαρμογή της προτεινόμενης μεθόδου Στεγανάλυσης σε επιλεγμένα format αρχείων.

Κεφάλαιο 8 – Υλοποίηση Λογισμικού Στεγανάλυσης

8.1 Υλοποίηση Λογισμικού που χρησιμοποιεί την προτεινόμενη μέθοδο Στεγανάλυσης για JPEG format.
8.2 Επιμέρους τμήματα λογισμικού.
8.3 Παραδείγματα χρήσης λογισμικού.
8.4 Αποτελέσματα / Στατιστικά Στοιχεία Επιτυχημένου εντοπισμού.
8.5 Συγκριτικά αποτελέσματα σε σχέση με άλλα εργαλεία Στεγανάλυσης.

Κεφάλαιο 9 - Επίλογος

9.1 Συμπεράσματα.
9.2 Στόχοι που επιτεύχθηκαν.
9.3 Μελλοντικές Εφαρμογές.
9.4 Επίλογος.

Κεφάλαιο 10 - Παράρτημα

Κεφάλαιο 11 - Βιβλιογραφία

1.1 Αντικείμενο της πτυχιακής εργασίας

Η Σήμανση Υπολογιστών (Computer Forensics) αποτελεί έναν εξειδικευμένο και ταχέως αναπτυσσόμενο κλάδο της επιστήμης των Ηλεκτρονικών Υπολογιστών. Τα τελευταία χρόνια υπήρξε στροφή από την συμβατική διερεύνηση έντυπων στοιχείων, στην εντατική διερεύνηση της πληθώρας αρχείων που αποθηκεύονται σε ηλεκτρονική μορφή, τόσο σε υπολογιστικά συστήματα όσο και σε άλλες περιφερειακές ηλεκτρονικές συσκευές.
Αυτή η μεταστροφή στην διερεύνηση στοιχείων, οφείλεται κυρίως στην αύξηση της χρήσης των ηλεκτρονικών υπολογιστών, που διαθέτουν μεγάλη χωρητικότητα αποθήκευσης δεδομένων, σε συνδυασμό με την πεποίθηση των ερευνητών πως πολλά στοιχεία που δεν βρίσκονται σε έντυπη \ φυσική μορφή μπορεί να βρίσκονται αποθηκευμένα σε ηλεκτρονική μορφή.

Με τον όρο Computer Forensics [112] πρωταρχικά αναφερόμαστε στην επιστήμη εκείνη που έχει ως κύριο μέλημα την εντατική μελέτη και γνώση των υπολογιστών και των μεθόδων αποθήκευσης σε αυτούς, με βασικό σκοπό την άντληση δεδομένων / στοιχείων με τρόπο κατά τον οποίο να μην επηρεάζεται η ακεραιότητα αυτών. Ο όρος Computer Forensics βέβαια έχει συσχετιστεί σε μεγάλο βαθμό πλέον με την εξιχνίαση ηλεκτρονικών εγκλημάτων, που έκαναν την επιστήμη αυτή πιο δημοφιλή.
Κλειδί λοιπόν στην διαδικασία που ακολουθείται για την ανάκτηση αυτών των στοιχείων είναι η μεθοδική τήρηση κανόνων με σκοπό τα στοιχεία που θα ανακτηθούν να είναι αποδεδειγμένα ακέραια ως προς το περιεχόμενό τους ώστε να υπάρχει η δυνατότητα χρήσης τους σε νομικές διαδικασίες.

Η Σήμανση Υπολογιστών (Computer Forensics)[112] περιλαμβάνει συνοπτικά:

1. Ανάκτηση δεδομένων που έχουν διαγραφεί εσκεμμένα ή μη, είναι κρυφά, είναι προστατευμένα με κωδικούς, είναι κρυπτογραφημένα ή στεγανογραφημένα.
2. Ανάκτηση δεδομένων από υπολογιστικά συστήματα τα οποία έχουν υποστεί βλάβες εσκεμμένες ή μη, έχουν προσβληθεί από κάποιον ιό ή ακόμη χειρότερα όταν υπάρχει ολική απώλειά τους.

Το ερώτημα λοιπόν «ποιά ανάγκη καλύπτει» ή «γιατί χρειαζόμαστε αυτή την επιστήμη» απαντάται αν εξεταστούν οι στατιστικές που δείχνουν μεγάλη αύξηση σε μη εξουσιοδοτημένη χρήση πληροφοριακών συστημάτων σε όλο τον κόσμο [117].

Πολλοί φορείς δηλώνουν άγνοια για το αν τα πληροφοριακά τους συστήματα έχουν γίνει αντικείμενο κατάχρησης από παράνομη ή μη εξουσιοδοτημένη χρήση, κάτι που δείχνει την ανάγκη για απόδειξη του τρόπου χρήσης των συστημάτων αυτών. Αυτό είναι κάτι που πρέπει να γίνει με τέτοιο τρόπο ούτως ώστε να μπορούν τα στοιχεία να συλλεχθούν, να αναλυθούν και να μπορούν να στοιχειοθετήσουν κατηγορία σε πιθανές δικαστικές διαδικασίες. Καταλήγουμε δηλαδή στο συμπέρασμα ότι το Computer Forensics πρέπει να εξασφαλίζει κάποιες παραμέτρους προκειμένου να διασφαλίζει κριτήρια που

αποδεικνύουν την μη-παραποίηση των στοιχείων αφενός και αφετέρου την αναγνωσιμότητα/μεταφερσιμότητα των συμπερασμάτων σε οργανισμούς, υπηρεσίες, εταιρείες ακόμα και απλούς πολίτες.

Αντικειμενικός σκοπός αυτής της εργασίας είναι, η κατανόηση όλων των παραπάνω διαδικασιών και αξιών της επιστήμης του Computer Forensics με σκοπό την ανάπτυξη νέων εξελιγμένων μεθόδων εντοπισμού κρυμμένης πληροφορίας και υλοποίησης των μεθόδων αυτών σε ευχρηστα εργαλεία.

Μέσα από την βαθύτερη κατανόηση της φύσης αλλά και των νόμων που διέπουν την επιστήμη αυτή, επαναχρησιμοποιούνται παλίες και δοκιμασμένες μέθοδοι με διαφορετική προσέγγιση για να επιτευχθούν καλύτερα αποτελέσματα εντοπισμού.

1.2 Στόχοι

Στόχοι της πτυχιακής αυτής εργασίας είναι :

- Η παρουσίαση των βασικών Κατηγοριών / Τεχνικών Απόκρυψης Πληροφορίας (Data Hiding)[1][36] υπό την οπτική της επιστήμης του Computer Forensics.
- Ο ορισμός, οι τεχνικές και η εφαρμογή της Στεγανογραφίας[1][33][113] υπό την σκοπιά της επιστήμης του Computer Forensics.
- Ο ορισμός, οι τεχνικές και η εφαρμογή της Στεγανάλυσης[1][33][114] υπό την σκοπιά της επιστήμης του Computer Forensics.
- Σύντομη παρουσίαση των μέχρι τώρα εργαλείων[33] Στεγανογραφίας και Στεγανάλυσης καθώς και εξοικίωση με τον τρόπο χρήσης αλλά και τις δυνατότητες τους.
- Η παρουσίαση του νόμου του Benford [3-6] και των χρήσεων αυτού σε διάφορες επιστήμες με έμφαση στην επιστήμη της Στεγανάλυσης.
- Η παρουσίαση μιας πρωτότυπης θεωρητικής χρήσης γενίκευσης του νόμου του Benford για Εντοπισμό Κρυμμένου Περιεχομένου ανεξαρτήτως τύπου αρχείου.
- Η παρουσίαση μιας πρωτότυπης χρήσης της γενίκευσης του νόμου του Benford για Εντοπισμό Κρυμμένου Περιεχομένου σε αρχεία εικόνας τύπου JPEG.
- Η ανάπτυξη και παρουσίαση λογισμικού που χρησιμοποιεί την προτεινόμενη μέθοδο Στεγανάλυσης με συνδυαστική χρήση γενίκευσης του νόμου του Benford και άλλων τεχνικών .
- Η παρουσίαση των συγκριτικών αποτελεσμάτων του προτεινόμενου λογισμικού σε σχέση με προηγούμενα.

1.3 Διάρθρωση της πτυχιακής εργασίας

Η διάρθρωση της πτυχιακής εργασίας θα ακολουθήσει πυραμιδοειδή ανάπτυξη, καθώς αρχικά θα παρουσιαστούν πιο γενικές και θεωρητικές έννοιες, για να καταλήξουμε τελικά σε πρακτική εφαρμογή αυτών.
Οι τρείς θεματικές ενότητες στις οποίες χωρίζεται η πτυχιακή αυτή εργασία είναι:

- Το Θεωρητικό Υπόβαθρο, η ιστορική διαδρομή αλλά και γενικές γνώσεις και αρχές γύρω από την επιστήμη της απόκρυψης πληροφορίας, της Στεγανογραφίας αλλά και της Στεγανάλυσης.
- Οι Νόμοι, δηλαδή αλλά και η Προτεινόμενη Θεωρητική Προσέγγιση που θα ακολουθηθεί προκειμένου να εντοπιστεί η κρυμμένη πληροφορία.
- Η Εφαρμογή, δηλαδή η υλοποίηση των θεωρητικών προτάσεων σε ένα πραγματικό εργαλείο με μετρήσιμη απόδοση αλλά και η σύγκριση του με άλλες παραπλήσιες εφαρμογές.

Αρχικά, ξεκινώντας από την βάση της πυραμίδας θα αναλυθεί η γενικευμένη έννοια της απόκρυψης πληροφορίας (Data Hiding), αλλά και η διερεύνηση αυτής μέσα από την οπτική της επιστήμης του Computer Forensics.
Έπειτα συγκεκριμενοποιώντας την έννοια του Data Hiding στην απόκρυψη πληροφορίας μέσα από την χρήση της Στεγανογραφίας θα πραγματοποιηθεί εκτενής ανάλυση και περιγραφή των πιο διαδεδομένων τεχνικών της Στεγανογραφίας όσο και των δημοφιλέστερων εργαλείων που χρησιμοποιούνται τα τελευταία χρόνια.

Θα ακολουθήσει παρουσίαση και περιγραφή των πιο διαδεδομένων τεχνικών της Στεγανάλυσης όσο και των δημοφιλέστερων εργαλείων για την πρακτική αυτή.
Ακολουθώντας όλο και πιο στοχευμένη ανάπτυξη, με βάση τους στόχους που τέθηκαν στην προηγούμενη ενότητα, εισάγουμε τον αναγνώστη στον Νόμο του Benford, μέσα από παραδείγματα χρήσης του σε διάφορους τομείς, για να καταλήξουμε στην χρήση που έχει ο νόμος αυτός στην Στεγανάλυση σήμερα.
Διακρίνοντας μια νέα προσέγγιση στην χρήση του νόμου του Benford ύστερα από παράθεση στατιστικών αποτελεσμάτων που συνηγορούν με την υπόθεσή μας, οδηγούμαστε σε μια γενίκευση του Νόμου που θα μπορούσε να χρησιμεύσει στον εντοπισμό κρυμμένου περιεχομένου.
Η υπόθεση μας μετασχηματίζεται σταδιακά, αρχικά σε γενική μεθοδολογία στεγανάλυσης με την χρήση του Νόμου του Benford[30] για εντοπισμό Κρυμμένου Περιεχομένου σε Forensics Investigation ανεξαρτήτως τύπου αρχείου (file type) (παράδειγμα στεγανογραφίας σε αρχεία κειμένου .txt), ώστε να αποκτήσει τελικά αποδοτική μορφή στοχεύοντας στον εντοπισμό Κρυμμένου Περιεχομένου σε αρχεία εικόνων τύπου .JPEG.

Μετά από την παρουσίαση της μεθοδολογίας, ακολουθεί η υλοποίηση της θεωρητικής προτεινόμενης προσέγγισης, σε ένα εργαλείο το οποίο θα μπορούσε να χρησιμοποιηθεί στα πλαίσια μιας Forensics Διερεύνησης (Investigation). Μέσα από παραδείγματα χρήσης, στατιστικά αποτελέσματα επιτυχίας αλλά και συγκριτικά αποτελέσματα σε σχέση με ανάλογο λογισμικό

στεγανάλυσης, αναδεικνύεται η αποτελεσματικότητα τόσο της προτεινόμενης μεθοδολογίας όσο και της υλοποίησης.

Τέλος παρουσιάζονται συνολικά τα συμπεράσματα αλλά και τα αποτελέσματα της εργασίας, η οποία επιτυγχάνει τους στόχους που τέθηκαν δίνοντας στην κοινότητα του Computer Forensics ένα αποδοτικό εργαλείο το οποίο χρησιμοποιεί ένα γνωστό μαθηματικό νόμο με πρωτότυπο τρόπο.

1.4 Κατανομή Εργασιών

Κεφάλαιο 1 - Εισαγωγή	
1.1 Αντικείμενο της πτυχιακής εργασίας.	Α. Μαρτίνη
1.2 Στόχοι	Α. Μαρτίνη
1.3 Διάρθρωση της πτυχιακής εργασίας.	Α.Ζαχαρής
1.4 Κατανομή Εργασιών	Α.Ζαχαρής
Κεφάλαιο 2 – Απόκρυψη Πληροφορίας	
2.1 Απόκρυψη πληροφορίας (Data Hiding) υπό την οπτική της επιστήμης του Computer Forensics.	Α. Μαρτίνη
2.2 Κατηγορίες / Είδη Data Hiding.	Α. Μαρτίνη
Κεφάλαιο 3 – Στεγανογραφία & Στεγανάλυση	
3.1 Ιστορική αναδρομή	Α. Μαρτίνη
3.2 Τι είναι Στεγανογραφία.	Α. Μαρτίνη
3.3 Κατηγορίες / Είδη Στεγανογραφίας	Α. Μαρτίνη
3.4 Η αξία της Στεγανάλυσης σε Forensics Investigations.	Α. Μαρτίνη
3.5 Κατηγορίες / Είδη Στεγανάλυσης.	Α. Μαρτίνη
3.6 Εργαλεία Στεγανογραφίας & Στεγανάλυσης.	Α. Μαρτίνη
Κεφάλαιο 4 – Ο Νόμος του Benford	
4.1 Ο νόμος του Benford και η επαλήθευση αυτού.	Α. Μαρτίνη
4.2 Χρήσεις του Νόμου του Benford μέχρι σήμερα σε διάφορους τομείς.	Α. Μαρτίνη
4.3 Παλαιότερες Χρήσεις του Νόμου του Benford για Στεγανάλυση.	Α. Μαρτίνη
Κεφάλαιο 5 – Στεγανογραφία και File Structure	
5.1 Στεγανογραφία και αλλοίωση του file structure.	Α.Ζαχαρής
5.2 Στατιστικά αλλοιώσεων για διαφορετικούς τύπους αρχείων.	Α.Ζαχαρής
5.3 Συμπεράσματα που οδηγούν στην Γενίκευση του Νόμου του Benford.	Α.Ζαχαρής
Κεφάλαιο 6 – Προτεινόμενη Μεθοδολογία Στεγανάλυσης	
6.1 Γενίκευση του Νόμου του Benford για εντοπισμό Κρυμμένου Περιεχομένου σε Forensics Investigation	Α.Ζαχαρής
6.2 Γενικευμένη Αναδόμηση αρχείων.	Α. Μαρτίνη
6.3 Προτεινόμενη μέθοδος Στεγανάλυσης με χρήση της Γενίκευσης του Νόμου του Benford.	Α.Ζαχαρής
Κεφάλαιο 7 – Εφαρμογή Προτεινόμενης Μεθοδολογίας Στεγανάλυσης	
7.1 Αναδόμηση αρχείων.	Α.Ζαχαρής
7.2 Εφαρμογή της προτεινόμενης μεθόδου Στεγανάλυσης σε επιλεγμένα format αρχείων.	Α.Ζαχαρής
Κεφάλαιο 8 – Υλοποίηση Λογισμικού Στεγανάλυσης	
8.1 Υλοτρείηση Λογισμικού που χρησιμοποιεί την προτεινόμενη μέθοδο Στεγανάλυσης για JPEG format.	Α.Ζαχαρής
8.2 Επιμέρους τμήματα λογισμικού.	Α.Ζαχαρής
8.3 Παραδείγματα χρήσης λογισμικού.	Α.Ζαχαρής
8.4 Αποτελέσματα / Στατιστικά Στοιχεία Επιτυχημένου εντοπισμού.	Α.Ζαχαρής
8.5 Συγκριτικά αποτελέσματα σε σχέση με άλλα εργαλεία Στεγανάλυσης	Α.Ζαχαρής
Κεφάλαιο 9 - Επίλογος	
9.1 Συμπεράσματα.	Α.Ζαχαρής
9.2 Στόχοι που επιτεύχθηκαν.	Α. Μαρτίνη
9.3 Μελλοντικές Εφαρμογές.	Α.Ζαχαρής
9.4 Επίλογος.	Α. Μαρτίνη
ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΕΡΓΑΛΕΙΟΥ ΣΤΕΓΑΝΟΓΡΑΦΙΑΣ ΣΕ JAVA	Α. Μαρτίνη, Α.Ζαχαρής

ΜΕΡΟΣ Ι - ΘΕΩΡΙΑ

2.1 Απόκρυψη πληροφορίας υπό την οπτική της επιστήμης του Computer Forensics.

Η ψηφιοποίηση των δεδομένων διευκολύνει την πρόσβαση, την μεταφερσιμότητα και την ακρίβεια των δεδομένων που παρουσιάζονται. Παράλληλα με τα θετικά στοιχεία της ψηφιοποίησης των δεδομένων υπάρχουν και κάποια αρνητικά όπως η ευκολία με την οποία αντιγράφονται, καταπατούνται τα πνευματικά τους δικαιώματα ή χρησιμοποιούνται σαν φορείς κρυφών δεδομένων.

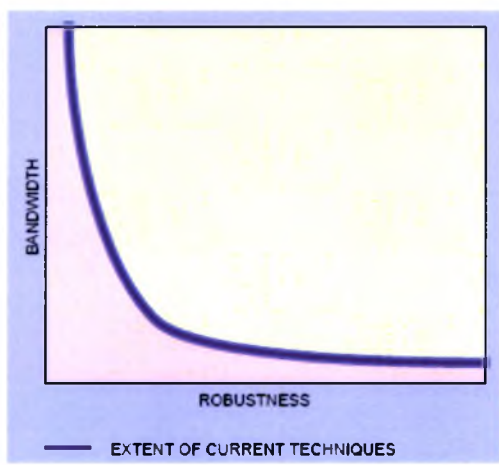
Η ηλεκτρονική απόκρυψη πληροφορίας ("Data Hiding")[36][37], εισάγει δεδομένα σε υπαρκτά ψηφιακά μέσα με σκοπό:

1. Την ταυτοποίηση
2. Τον σχολιασμό
3. Τα πνευματικά δικαιώματα
4. Την μεταφορά κρυφής πληροφορίας

Κατά την διαδικασία αυτή υπάρχουν πολλοί περιορισμοί οι οποίοι πρέπει να ληφθούν υπόψη όπως :

1. Ο αριθμός των αρχείων που θα κρυφθούν.
2. Το μέγεθος (file size) των κρυφών δεδομένων.
3. Ο βαθμός σταθερότητας των δεδομένων προς απόκρυψη, υπό προϋποθέσεις (π.χ. lossy or lossless compression).
4. Ο βαθμός ανοχής σε αλλοιώσεις, τροποποιήσεις ή και αφαίρεση του κρυφού περιεχομένου από τρίτους (third party).

Η ηλεκτρονική απόκρυψη πληροφορίας αποτελεί ένα εργαλείο με πολλές πτυχές, ανάλογες του σκοπού χρήσης της. Χρησιμοποιείται σε διάφορα είδη αρχείων όπως εικόνες, βίντεο, κείμενο με απώτερο σκοπό την ελάχιστη δυνατή αντιληπτή αλλαγή / τροποποίηση / αλλοίωση του αρχικού φέροντος σήματος (host signal) κατά την εισαγωγή κρυφής πληροφορίας (hidden signal).



Εικόνα 2.1: Αναλογία κρυφής πληροφορίας - στιβαρότητας*[2]*.

Όσο αυξάνεται ο όγκος της κρυφής πληροφορίας (bandwidth), τόσο μειώνεται η στιβαρότητα (robustness) του φέροντος σήματος (host signal) με αποτέλεσμα η κρυφή πληροφορία (hidden signal) να γίνεται πιο εύκολα αντιληπτή όπως παρουσιάζεται και στην *Εικόνα 2.1[2].*

Πρέπει να σημειωθεί ότι η ηλεκτρονική απόκρυψη πληροφορίας ("Data Hiding"), παρ' όλο που μοιάζει με την συμπίεση πληροφορίας (compression), διαφέρει εξ' ολοκλήρου από την κρυπτογράφηση (encryption).
Σκοπός της απόκρυψη πληροφορίας δεν είναι να περιορίζει ή να ελέγχει το φέρον σήμα (host signal), αλλά να εξασφαλίζει ότι τα ενσωματωμένα αρχεία παραμένουν αναλλοίωτα και εύκολα ανακτήσιμα.

Δύο είναι οι βασικές αρχές που διέπουν την ηλεκτρονική απόκρυψη πληροφορίας είναι :

1. Η απόδειξη των πνευματικών δικαιωμάτων του φέροντος σήματος.
2. Η πιστοποίηση της ακεραιότητας του κρυφού περιεχομένου.

Για του δύο παραπάνω λόγους τα δεδομένα πρέπει να παραμένουν κρυμμένα παρ' όλες τις εξωτερικές αλλοιώσεις που μπορεί να υποστεί το φέρον σήμα.
Οι τεχνικές που έχουν αναπτυχθεί μέχρι σήμερα, με σκοπό την ηλεκτρονική απόκρυψη πληροφορίας διαφέρουν ανάλογα με :

- Την ποσότητα / μέγεθος της πληροφορίας που πρόκειται να κρυφτεί.
- Την στιβαρότητα των δεδομένων σε εξωτερικές αλλαγές.

Εφόσον δεν υπάρχει μία μοναδική μέθοδος η οποία να επιτυγχάνει όλους τους παραπάνω στόχους, οι υλοποιήσεις της ηλεκτρονικής απόκρυψης πληροφορίας διαφέρουν ανάλογα με την εφαρμογή.

Οι τεχνικές προκλήσεις βρίσκονται όχι απλά στον εντοπισμό «κενών» τα οποία μπορούν να γεμίσουν με τμήματα της κρυφής πληροφορίας, αλλά στον εντοπισμό κενών τα οποία δεν πρόκειται να αλλοιωθούν κατά την μεταφορά, συμπίεση ή και τροποποίηση του φέροντος σήματος.

2.1.1 Χαρακτηριστικά

Μερικά από τα χαρακτηριστικά αλλά και περιορισμοί που περιγράφουν μία απόπειρα απόκρυψης πληροφορία είναι[1-2] :

1. Το φέρον σήμα ανεπαίσθητα υποβιβάζεται σε ποιότητα και τα κρυφά δεδομένα γίνονται αντιληπτά σε ελάχιστο βαθμό.
2. Η κρυμμένη πληροφορία πρέπει να έχει ενσωματωθεί στο ίδιο το μέσο (encoded) και όχι να έχει κρυφτεί σε κεφαλίδες (headers) του αρχείου, έτσι ώστε τα κρυφά δεδομένα να παραμένουν αναλλοίωτα αλλάζοντας μεταξύ διαφορετικών formats αρχείων.
3. Τα κρυφά δεδομένα πρέπει να παραμένουν αναλλοίωτα από εσκεμμένες προσπάθειες τροποποίησης ή απομάκρυνσης της κρυφής πληροφορίας.

4. Η ανάκτηση της κρυφής πληροφορίας πρέπει να γίνεται εύκολα, στόχος που επιτυγχάνεται μέσα από ασύμμετρη κωδικοποίηση της κρυφής πληροφορίας.
5. Error Correction κωδικοποίηση θα πρέπει να μπορεί να επαναφέρει την κρυφή πληροφορία, που θα έχει υποστεί αλλοιώσεις, στην αρχική της κατάσταση.

2.1.2 Εφαρμογές

Όπως αναφέρθηκε και σε προηγούμενη ενότητα, κατά την ηλεκτρονική απόκρυψη πληροφορίας υπάρχει πάντα ένας βαθμός ανταλλαγής μεταξύ όγκου κρυφής πληροφορίας και βαθμού που γίνεται αντιληπτή η κρυφή πληροφορία από τρίτους.

Μειώνοντας τον όγκο κρυφής πληροφορίας μπορούμε να αυξήσουμε την στιβαρότητα του φέροντος σήματος και αντίστροφα. Η παραπάνω διαπίστωση μπορεί να αποδειχθεί μαθηματικά για κάποιες μεθόδους απόκρυψης πληροφορίας όπως η spread spectrum και τείνει να επαληθεύεται και όλες τις υπόλοιπες τεχνικές απόκρυψη πληροφορίας.

Λαμβάνοντας υπ' όψη τα παραπάνω γίνεται αντιληπτό ότι ανάλογα με το αποτέλεσμα που θέλουμε να πετύχουμε με την εφαρμογή μας αλλάζει και η τεχνική απόκρυψης πληροφορίας.

Για παράδειγμα:

1. Η τοποθέτηση ψηφιακού υδατογραφήματος (digital watermark) απαιτεί ελάχιστη κρυφή πληροφορία που λειτουργεί σαν ψηφιακή υπογραφή. Η τεχνική που θα χρησιμοποιηθεί σε μία τέτοια εφαρμογή πρέπει να λαμβάνει υπ' όψη την πιθανότητα αλλοίωσης ή απαλοιφής του υδατογραφήματος και πρέπει να προστατεύει αυτό από πιθανές εσκεμμένες ή μη προσπάθειες.

2. Η τοποθέτηση κρυφής πληροφορίας για εξασφάλιση μη τροποποίησης του περιεχομένου ενός αρχείου από μη εξουσιοδοτημένο χρήστη. Σε περίπτωση που αλλοιωθεί το περιεχόμενο του αρχείου, αλλοιώνεται και η κρυφή πληροφορία. Η τεχνική που θα χρησιμοποιηθεί σε μία τέτοια εφαρμογή πρέπει να λαμβάνει υπ' όψη την πιθανότητα αλλοίωσης ή απαλοιφής και πρέπει να προστατεύει αυτό από πιθανές εσκεμμένες ή μη προσπάθειες.

3. Μεταφορά κρυφού περιεχομένου χωρίς να γίνει αντιληπτό από τρίτους. Σκοπός η μεταφορά δεδομένων μεταβλητού μεγέθους, πολλές φορές κρυπτογραφημένου. Η τεχνική που θα χρησιμοποιηθεί σε μία τέτοια εφαρμογή πρέπει να λαμβάνει υπ' όψη την πιθανότητα η κρυφή πληροφορία να γίνει αντιληπτή λόγω μεγάλου μεγέθους κρυφής πληροφορίας που αλλοιώνει το φέρων σήμα.

Ακολουθεί εκτενής κατηγοριοποίηση των τεχνικών απόκρυψης.

2.2 Κατηγορίες / Είδη Data Hiding

Οι τεχνικές απόκρυψης αρχείων (Data Hiding) είναι πάρα πολλές και ένας Forensics Investigator πρέπει να είναι ενημερωμένος για όλες προκειμένου να μην παραβλέψει σημεία του υπολογιστικού συστήματος στο οποίο θα μπορούσαν να βρεθούν στοιχεία. Σύμφωνα με μια γενική κατηγοριοποίηση το Data Hiding χωρίζεται σε :

- Media Management Layer
- File System Layer
- Application Layer

2.2.1 Media Management Layer

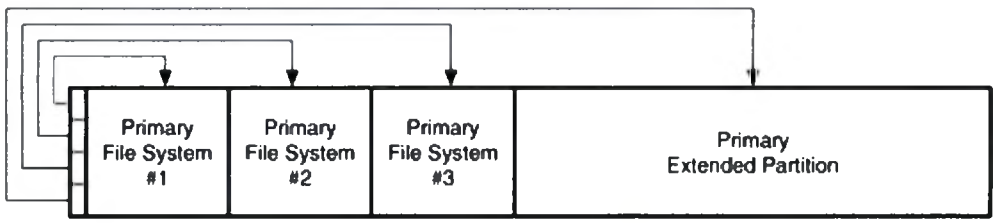
Το πρώτο επίπεδο είναι εκείνο στο οποίο ένα αποθηκευτικό μέσο (π.χ. σκληρός δίσκος), χωρίζεται από την φυσική μορφή του σε μικρότερα κομμάτια.

Ένας σκληρός δίσκος μπορεί να χωρίζεται σε λογικές ενότητες που ονομάζονται **partitions** τα οποία μορφοποιούνται με κάποιο **file system** και κατηγοριοποιούνται σε δυο είδη, τα **primary** και τα **extended**.
Σε ένα σκληρό δίσκο μπορεί να συνυπάρξουν παραπάνω από ένα partitions αλλά μόνο ένα κάθε φορά είναι το ενεργό, από το οποίο δηλαδή ενεργοποιείται η λειτουργία της εκκίνησης του υπολογιστή.
Γι' αυτή την λειτουργία είναι υπεύθυνο το τμήμα του σκληρού δίσκου που ονομάζεται Master Boot Record (MBR), το οποίο βρίσκεται στην πρώτη λογική μονάδα αποθήκευσης κάθε υπολογιστή και περιέχει:

- Τον κώδικα εκκίνησης, ο οποίος καθορίζει στον υπολογιστή ποια είναι τα προσπελάσιμα partitions και που μπορεί να βρει το λειτουργικό σύστημα από το οποίο θα γίνει η εκκίνηση.
- Τον πινάκα των partitions.
- Μια αναγνωριστική τιμή. (signature value)

Ο MBR μπορεί να περιγράψει / υποστηρίξει μέχρι τέσσερα partitions, αλλά επειδή πολλά λειτουργικά χρειάζονται πολύ περισσότερα, χρησιμοποιούνται έγγραφες από τον πινάκα των partitions και δημιουργείται ένα extended partition το οποίο καταλαμβάνει όλο τον υπόλοιπο χώρο. Σε αυτό το χώρο, η διάταξη διαφοροποιείται και από εδώ και περά κάθε partition έχει στην αρχή του πληροφορίες όπως:

- Ποιο είναι το μέγεθος του
- Ποιο σύστημα διαχείρισης αρχείων (file system) χρησιμοποιεί
- Που βρίσκεται το επόμενο partition



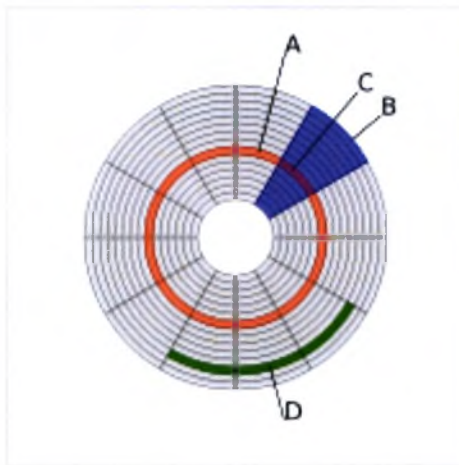
Εικόνα 2.2

Οι σκληροί δίσκοι από φυσικής πλευράς χωρίζονται σε:

- Tracks
- Cylinders
- Sectors

Οι **sectors** διαμορφώνουν / ομαδοποιούνται σε **clusters** διαφόρων μεγεθών τα όποια αποτελούν το μικρότερο λογικό μέγεθος του δίσκου στο οποίο μπορεί να ανατεθεί ένα αρχείο και είναι η βασική μονάδα αποθήκευσης της πληροφορίας πάνω σε ένα δίσκο, ενώ το μέγεθος τους καθορίζεται από το **file system** ή αντιστοίχως από το λειτουργικό σύστημα, το οποίο καθορίζει και το **file system**.

Sector: Ονομάζεται το μικρότερο διευθυνσιοδοτήσιμο κομμάτι πάνω στο δίσκο και μπορεί να περιέχει 512 bytes δεδομένων.



Disk structure:
(A) track
(B) geometrical sector
(C) track sector
(D) ***cluster***

Εικόνα 2.3

Σε αυτό το επίπεδο οι τεχνικές συγκάλυψης αρχείων είναι εκείνες οι οποίες ασχολούνται με την χρήση χώρων που το λειτουργικό σύστημα είτε δεν γνωρίζει, είτε θεωρεί μη δεσμευμένους. Σε αυτή την κατηγορία εμπίπτουν περιοχές όπως «unallocated space» και «Host Protected Area». Στην ίδια

κατηγορία ανήκουν και τα δεδομένα που μπορεί να βρίσκονται σε «Partition Gap», το μη χρησιμοποιούμενο κομμάτι που βρίσκεται αμέσως μετά τον MBR, το «volume slack» κ.α.

2.2.2 File System Layer

Στο δεύτερο επίπεδο οι τεχνικές συγκάλυψης χρησιμοποιούν τις δομές των file systems εκμεταλλευόμενες τις αδυναμίες και τα χαρακτηριστικά της εκάστοτε έκδοσης.

Τα **file systems** που χρησιμοποιούνται περισσότερο είναι τα:

- Τύπου FAT
- Τύπου NTFS
- Τύπου EXT2/EXT3

Σε αυτή την κατηγορία εμπίπτουν οι τεχνικές που χρησιμοποιούν οποιουδήποτε είδους «Slack Space», τα Alternate Data Streams (ADS του NTFS), τα Reserved i-nodes των Unix file systems όπως το EXT2/3 (τα οποία αναφέρονται ονομαστικά μονό) και η χρήση «ειδικών» file names τα οποία προσποιούνται αρχεία του συστήματος.

Παρακάτω παρουσιάζονται αναλυτικά διάφορες τεχνικές που χρησιμοποιούνται μεμονωμένα ή σε συνδυασμό για την επιτυχή απόκρυψη πληροφορίας και περιλαμβάνονται στις κατηγορίες 2.2.1 και 2.2.2

Ορισμός: Κατάλοιπα Δεδομένων είναι τα δεδομένα τα όποια δεν είναι ενεργά σε ένα πληροφοριακό σύστημα και περιλαμβάνουν:

1. Τα δεδομένα τα όποια βρίσκονται σε ελεύθερο χώρο πάνω στο μέσο
2. Τα δεδομένα που βρίσκονται στο “slack space”
3. Τα δεδομένα μέσα σε αρχεία τα όποια έχουν διαγραφεί στο παρελθόν και δεν είναι εμφανή με χρήση των εφαρμογών που τα έχουν δημιουργήσει χωρίς την χρήση ειδικών εφαρμογών τύπου “undelete” και ειδικές τεχνικές recovery.

Οι κατηγορίες αυτού του είδους δεδομένων περιγράφονται ως εξής :

File Slack: Είναι ο χώρος που δεν χρησιμοποιείται μεταξύ του λογικού τέλους του αρχείου σε σχέση με το φυσικό. Όταν τα δεδομένα αποθηκεύονται σε clusters υπάρχει περίπτωση να μην καλύπτουν ακριβώς αυτές τις δομικές αποθηκευτικές μονάδες και να περισσεύει κάποιος χώρος. Το file system που χρησιμοποιείται είναι καθοριστικός παράγοντας για το πόσος τέτοιος χώρος παραμένει κενός.
Στο FAT, που χρησιμοποιούνται clusters μεγέθους 64Kb, αν αποθηκευτεί ένα αρχείο με μέγεθος 1Kb, χάνονται 63Kb.
Στο NTFS το μέγεθος των clusters είναι πολύ μικρότερο, συνήθως 4Kb .Τα υπολειπόμενα bytes σε ένα cluster περιέχουν πιθανώς υπολείμματα από προηγούμενα αρχεία που είχαν αποθηκευτεί εκεί. Αυτά τα bytes μπορούν να

ανακτηθούν με forensics τεχνικές και να αποκαλύψουν πρόσθετα στοιχεία (evidence) .

Εικόνα 2.4 : Cluster Μεγέθους 4 Kb

Volume Slack: Είναι ο μη χρησιμοποιούμενος χώρος μεταξύ του τέλους ενός file system και το τέλος του partition το οποίο το περιέχει. Το μέγεθος των δεδομένων που κρύβονται στο Volume Slack είναι απεριόριστο επειδή μπορεί να ρυθμιστεί από το χρήστη.

File System Slack: Είναι ο χώρος στο τέλος ενός file system στον οποίο δεν έχει ανατεθεί κανένα cluster. Αυτό συμβαίνει επειδή το μέγεθος του partition μπορεί να μην είναι πολλαπλάσιο του μεγέθους του cluster που χρησιμοποιεί το file system.Για παράδειγμα, μπορεί να υπάρχουν 10001 sectors σε ένα partition και οι 1000 πρώτοι ανατίθενται σε 2500 clusters με μέγεθος 4 sectors το καθένα, τότε ο τελευταίος sector γίνεται file system slack. Το μέγεθος των δεδομένων που μπορεί να κρυφτούν εδώ είναι σχετικά μικρό και εξαρτάται από το file system και το μέγεθος των clusters που χρησιμοποιεί.

Partition Gap: Όταν κάποιος σκληρός δίσκος μορφοποιείται με παραπάνω από ένα partitions, τότε είναι πιθανόν να υπάρχουν κενά μεταξύ των partitions τα οποία μπορούν να χρησιμοποιηθούν για να κρυφτούν δεδομένα. Τα κενά αυτά μπορεί να περιέχουν επίσης δεδομένα που υπήρχαν εκεί πριν γίνει η μορφοποίηση στην περίπτωση που ο δίσκος επαναμορφοποιείται.

MBR-area: Κάθε DOS partition δεσμεύει χώρο στην αρχή του, για τον MBR ο οποίος έχει περιγραφεί παραπάνω. Επειδή ο MBR χρησιμοποιεί έναν μοναδικό sector και τα partitions πρέπει να αρχίζουν και να τελειώνουν σε cylinder boundary, μένουν 62 κενοί sectors που ανήκουν στην περιοχή του MBR.

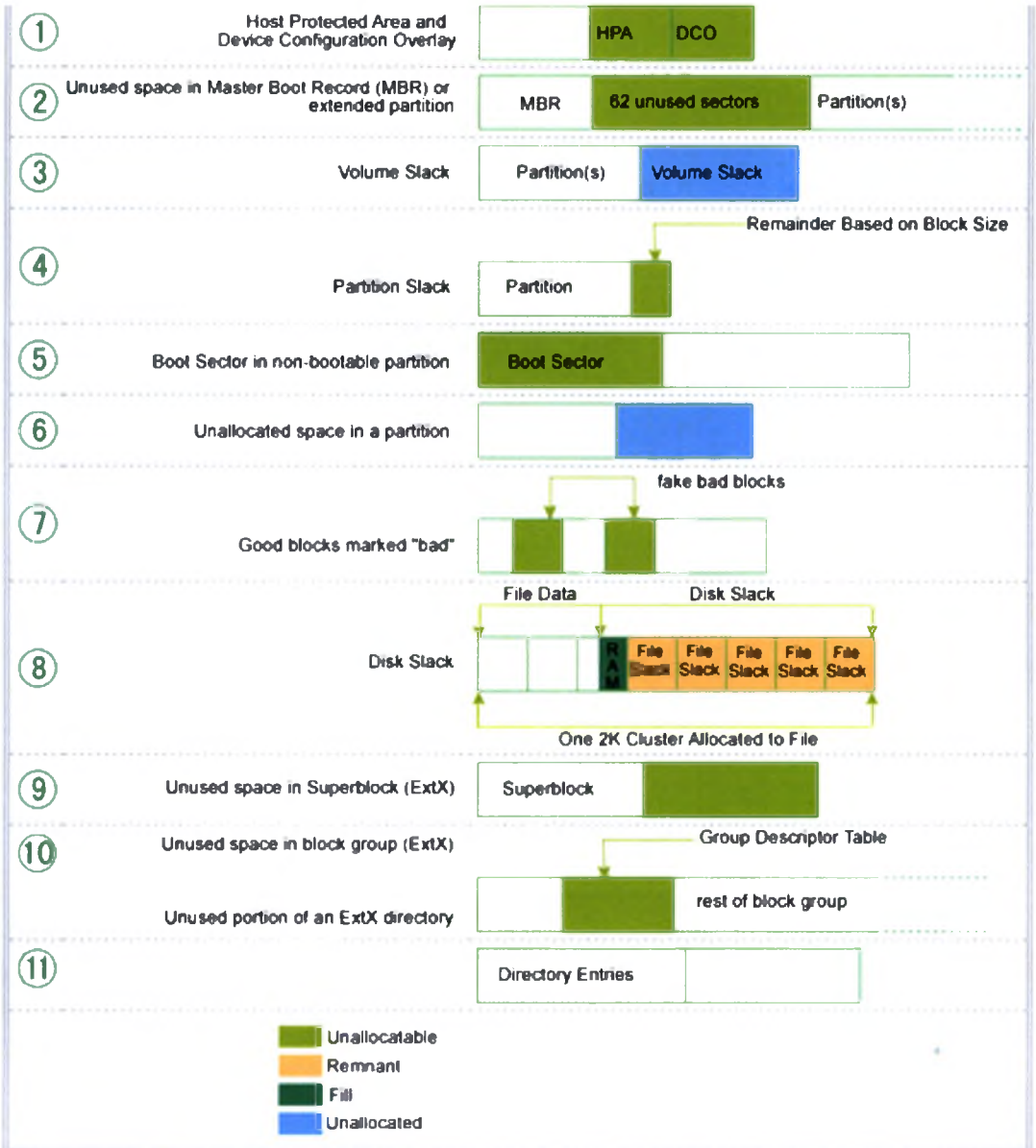
RAM slack: Είναι ένα φαινόμενο που παρατηρούνταν σε παλιότερα λειτουργικά κατά το οποίο όταν το λειτουργικό σύστημα έπρεπε να γράψει κάτι στο δίσκο, γνωρίζοντας το μέγεθος του, συμπλήρωνε το sector που μπορεί να είχε περισσευούμενο χώρο με δεδομένα από την μνήμη. Είναι κάτι πολύ επικίνδυνο, καθώς στα δεδομένα της RAM μπορεί να βρίσκονται passwords, αριθμοί πιστωτικών καρτών και πολλά ακόμα δεδομένα τα όποια δεν υπήρχε πρόθεση να αποθηκευτούν στο δίσκο. Σήμερα το φαινόμενο έχει εξαλειφθεί καθώς τα καινούρια λειτουργικά συστήματα γεμίζουν τον περισσευούμενο χώρο με μηδενικά.

Swap File: Είναι η virtual RAM που χρησιμοποιείται από το σύστημα όταν χρειάζεται (όταν δεν υπάρχει αρκετή RAM). Εδώ μπορούν να βρεθούν δεδομένα της RAM που κανονικά θα είχαν χαθεί σε ένα σύστημα το οποίο θα βρισκόταν κλειστό. Αλλά επειδή τα συγκεκριμένα αποθηκεύονται στο δίσκο, μπορούν να ανακτηθούν και να δώσουν δεδομένα όπως πληροφορίες για αρχεία, ουρές αρχείων προς εκτύπωση (printer queues) και δεδομένα που

αποθηκεύτηκαν \ σώθηκαν από τον χρηστή μαζί με τις χρονοσφραγίδες τους γραμμένες από το λειτουργικό σύστημα.

Unallocated space: Είναι το ακόμα μη χρησιμοποιημένο μέρος του δίσκου. Αν για παράδειγμα φτιάξουμε ένα partition μεγέθους 80 Gb σε ένα σκληρό μεγέθους 100 Gb, τα υπόλοιπα 20Gb που μένουν είναι το Unallocated space. Επειδή είναι έξω από την δικαιοδοσία του λειτουργικού συστήματος, δεν μπορεί να προστατεύει από αυτό, ούτε να καταγράφει η δραστηριότητα που παίρνει μέρος σε αυτό το κομμάτι. Συνεπώς εάν κάποιος με δικαιώματα administrator χρησιμοποιήσει το Unallocated space για να γράψει ένα αρχείο το οποίο θα ήθελε να δείξει σε κάποιον έκτος δικαιοδοσίας, ο δεύτερος δεν χρειάζεται να έχει δικαιώματα πάνω στο λειτουργικό για να το δει. Κατά αυτόν τον τρόπο μπορεί να λάβει μέρος κάποιο είδος επικοινωνίας μεταξύ χρηστών πάνω σε πληροφορίες που δεν θα έπρεπε εξ' αρχής να ανταλλάσσουν. Ένας πολύ απλός τρόπος με τον οποίο θα μπορούσε κάποιος να χρησιμοποιήσει unallocated space για να κρύψει δεδομένα είναι και ο ακόλουθος:
Μορφοποιούνται δύο ξεχωριστά partitions εκ των οποίων στο ένα τοποθετούνται τα δεδομένα που πρέπει να συγκαλυφθούν .Έπειτα σβήνεται το partition το οποίο τα περιέχει, κάτι που όμως δεν συνεπάγεται την διαγραφή των δεδομένων που αρχικά τοποθετήθηκαν εκεί .Οπότε αυτά τα δεδομένα παραμένουν σε μη δεσμευμένο χώρο "αόρατο" από το υπόλοιπο σύστημα.

Host Protected Area: Μερικοί σκληροί δίσκοι έχουν δεσμεύσει μια ειδική περιοχή που λέγεται Host Protected Area.Είχε αρχικά σχεδιαστεί έτσι ώστε να υπάρχει μια περιοχή στο δίσκο τέτοια ώστε οι πωλητές να μπορούν να αποθηκεύουν κάποια δεδομένα πάνω στο δίσκο, που να μην μπορούν να προσπελαστούν από τον κοινό χρηστή, το λειτουργικό (format, delete) ή το file system Σε αυτό το χώρο αποθηκεύονται διαγνωστικά και booting εργαλεία, προαποθηκευμένο λειτουργικό σύστημα για εγκατάσταση και για επαναφορά σε περίπτωση βλάβης κλπ. Συχνά χρησιμοποιείται και από εταιρίες ασφάλισης υπολογιστών από κλοπή, εκμεταλλευόμενες το γεγονός ότι είναι μια περιοχή που δεν ανιχνεύεται και δεν επηρεάζεται από format όλου του συστήματος. Έτσι χρησιμοποιώντας λογισμικό που «επικοινωνεί» με τους servers της εταιρείας Ασφάλισης όταν βρεθεί σε δίκτυο, εντοπίζεται το κλεμμένο σύστημα.
Επίσης χρησιμοποιείται από *rootkits* για να μην μπορούν να ανιχνευτούν από αντί-ιικα προγράμματα.
Από την άλλη είναι και ένας ιδανικός χώρος στον οποίο μπορούν να κρυφτούν δεδομένα, τα όποια θα περιείχαν στοιχειά σημαντικά για μια ερευνά forensics.Χρηστές ή εφαρμογές θα μπορούσαν να γραφούν και να διαβάζουν δεδομένα από αυτό το χώρο και το λειτουργικό να μην μπορεί να το ελέγξει.



Εικόνα 2.5 : Αναλυτική απεικόνιση τεχνικών απόκρυψης.
Source: Hal Berghel, David Hoelzer and Michael Sthultz, "Data Hiding Tactics for Windows and Unix File Systems"

2.2.3 Application Layer

Στο τρίτο επίπεδο, τέλος ανήκουν πιο γνωστές τεχνικές, όπως αυτές της κρυπτογραφίας, της υδατογραφίας και της στεγανογραφίας με την οποία θα ασχοληθούμε εκτενώς σε επόμενη ενότητα. Σήμερα, μη φυσικές μέθοδοι απόκρυψης πληροφορίας μπορούν να έχουν μία ή και περισσότερες από τις παρακάτω γενικές ψηφιακές μορφές:

- Κρυπτογραφία (cryptography)
- Στεγανογραφία (steganography)
- Υδατογράφημα (watermarking)[118]

Αν και οι παραπάνω μορφές σχετίζονται απόλυτα μεταξύ τους, με την έννοια ότι όλες προσπαθούν να προσφέρουν ασφαλή επικοινωνία, δεν παύουν να

έχουν σε μεγάλο βαθμό και διαφορές οι οποίες τις διαχωρίζουν. Για παράδειγμα:

- Ο κρυπτογράφος έχει σαν κύριο σκοπό του την προστασία / απόκρυψη του περιεχομένου ενός μηνύματος χωρίς να τον ενδιαφέρει, να αποκρύψει τον τρόπο / μέσο με τον οποίο θα σταλεί το μήνυμα.
- Το υποκείμενο που εκτελεί Στεγανογραφία από την άλλη πλευρά, έχει ως κύριο σκοπό να κρύψει κυρίως το μέσο με το οποίο θα εκτελέσει την επικοινωνία. Να δημιουργήσει δηλαδή ένα κρυφό κανάλι επικοινωνίας.
- Τέλος ο κύριος σκοπός της χρήσης υδατογραφημάτων είναι η προσπάθεια να προσθέσουμε στο επίμαχο αρχείο που υδατογραφούμε ικανή ποσότητα μεταδεδομένων (metadata) ώστε να αποδείξουμε την αυθεντικότητα ή ιδιοκτησία αυτού.

Η Κρυπτογραφία και Στεγανογραφία μοιράζονται την ιδιότητα του ότι το αντικείμενο ενδιαφέροντος είναι αυτό που πρέπει να παραμείνει κρυφό, να ενσωματωθεί σε άλλο αρχείο ή να κρυπτογραφηθεί. Αντίθετα, το υδατογράφημα κρύβεται / ενσωματώνεται μέσα στο μέσο του οποίου θέλουμε να εξακριβώσουμε την αυθεντικότητα. Επίσης, Υδατογραφία και Στεγανογραφία μπορούν να χρησιμοποιηθούν με ή χωρίς την χρήση Κρυπτογραφίας.

Η Στεγανογραφία η "Stego" όπως αναφέρεται στην κοινότητα της Πληροφορικής, στην κυριολεξία σημαίνει "καλυμμένη γραφή" και προέρχεται από την Ελληνική γλώσσα. Η Στεγανογραφία ορίζεται από τον Markus Kahn [1] ως ακολούθως: "Στεγανογραφία είναι η τέχνη και η επιστήμη της επικοινωνίας με έναν τρόπο ο όποιος κρύβει την ύπαρξη της επικοινωνίας". Σε αντίθεση με την Κρυπτογραφία, όπου στον εχθρό επιτρέπεται να ανιχνεύει, να παρεμβαίνει και να τροποποιεί τα μηνύματα χωρίς να μπορεί να παραβιάζει συγκεκριμένες προϋποθέσεις ασφαλείας που ορίζονται από ένα Κρυπτογραφικό σύστημα, ο στόχος της Στεγανογραφίας είναι να κρύψει μηνύματα μέσα σε άλλα "ακίνδυνα" μηνύματα με έναν τρόπο που δεν επιτρέπει σε κανένα εχθρό ούτε καν να ανιχνεύσει ότι υπάρχει ένα δεύτερο μήνυμα.

Σε έναν ψηφιακό κόσμο, Στεγανογραφία και Κρυπτογραφία θέλουν να προστατεύσουν την πληροφόρηση από ανεπιθύμητες ομάδες. Η Στεγανογραφία αλλά και η Κρυπτογραφία είναι άριστα μέσα για να το επιτελέσουν, αλλά καμία από τις δυο τεχνολογίες μεμονωμένα δεν είναι τέλεια, καθώς και οι δυο μπορούν να παραβιαστούν. Γι' αυτό το λόγο οι περισσότερο ειδικοί προτείνουν τη χρησιμοποίηση και των δυο τεχνολογιών για την πρόσθεση πολλαπλών στρωμάτων προστασίας. Η Στεγανογραφία μπορεί να χρησιμοποιηθεί για έναν μεγάλο αριθμό τύπων δεδομένων του σήμερα. Οι πιο δημοφιλείς τύποι δεδομένων που χρησιμοποιούνται είναι οι : bmp, .doc, .gif, .jpeg, .mp3, .txt, και .wav. κυρίως λόγω της δημοτικότητας τους στο Ίντερνετ και της ευκολίας χρήσης των Στενογραφικών εργαλείων που χρησιμοποιούν αυτούς τους τύπους δεδομένων.

Αυτοί οι τύποι είναι επίσης δημοφιλείς λόγω της σχετικής ευκολίας με την οποία μπορούν τα περιττά δεδομένα, ή θόρυβος, να αφαιρεθούν από αυτούς και να αντικατασταθούν με ένα κρυφό μήνυμα.

Οι Στεγανογραφικές τεχνολογίες είναι ένα πολύ σημαντικό κομμάτι του μέλλοντος της προστασίας και της εμπιστευτικότητας στα Ανοικτά Συστήματα όπως είναι το Ίντερνετ. Η Στεγανογραφική έρευνα καθοδηγείται κυρίως από την έλλειψη δύναμης στα Κρυπτογραφικά συστήματα και την επιθυμία να υπάρξει ολοκληρωτική μυστικότητα στο περιβάλλον Ανοικτών Συστημάτων.

Πολλές κυβερνήσεις έχουν δημιουργήσει νόμους που είτε οριοθετούν την δύναμη των Κρυπτογραφικών συστημάτων, είτε τα εμποδίζουν εξ' ολοκλήρου. Αυτό δυστυχώς αφήνει την πλειονότητα της κοινότητας του Ίντερνετ με σχετικά ανίσχυρους και τις περισσότερες φορές εύθραυστους αλγορίθμους απόκρυψης ή ακόμα και καθόλου προστασία.

Οι υπέρμαχοι των αστικών ελευθεριών αντιτίθενται σε αυτό με το επιχείρημα ότι "αυτοί οι περιορισμοί είναι προσβολή της ιδιοτικότητας". Εδώ "εισέρχεται" η Στεγανογραφία μιας και μπορεί να χρησιμοποιηθεί για να κρύψει σημαντικά δεδομένα μέσα σε έναν άλλο αρχείο έτσι ώστε μόνο οι νόμιμες ομάδες χρηστών να ξέρουν ότι υπάρχει κρυφό μήνυμα.

Για να προστεθούν πολλαπλά στρώματα προστασίας και για να μειωθούν τα σχετικά προβλήματα με θέμα "Κρυπτογραφία και Νόμος" που αναφέρθηκαν νωρίτερα, μια καλή πρακτική είναι να χρησιμοποιούνται Κρυπτογραφία και Στεγανογραφία μαζί. Όπως αναφέρθηκε νωρίτερα, ούτε η Κρυπτογραφία ούτε η Στεγανογραφία θεωρούνται "σίγουρες" λύσεις για να ανοιχθεί η εμπιστευτικότητα των συστημάτων, αλλά χρησιμοποιώντας και τις δυο τεχνολογίες μπορεί να διασφαλιστεί ένα αποδεκτό επίπεδο εμπιστευτικότητας σε οποιονδήποτε συνδέεται και επικοινωνεί μέσω αυτών των συστημάτων.

3.1 Ιστορική Αναδρομή

Οι πρώτες καταγραφές Στεγανογραφίας είχαν γίνει από τον Έλληνα ιστορικό Ηρόδοτο στα χρονικά του γνωστά ως "Ιστορίες" και χρονολογούνται γύρω στο 440 π.χ. Ο Ηρόδοτος κατέγραψε δυο ιστορίες με Στεγανογραφικές τεχνικές κατά τη διάρκεια εκείνης της περιόδου στην Ελλάδα. Στην πρώτη καταγράφηκε ότι ο Βασιλιάς Δαρείος της Σούσας, ξύρισε το κεφάλι ενός κρατουμένου του και έγραψε ένα κρυφό μήνυμα στο δέρμα του κρανίου του. Όταν τα μαλλιά του κρατουμένου ξαναμεγάλωσαν, αυτός στάλθηκε στον γαμπρό του Δαρείου, τον Αρισταγόρα της Μιλήτου χωρίς να ανακαλυφθεί.

Η δεύτερη ιστορία καταγράφηκε επίσης από τον Ηρόδοτο, η οποία υποστηρίζει ότι ένας στρατιώτης με το όνομα Δεμέρατος χρειαζόταν να στείλει ένα μήνυμα στην Σπάρτη για να ενημερώσει ότι ο Ξέρξης ετοιμαζόταν να εισβάλλει στην Ελλάδα. Εκείνη την εποχή, το μέσο του γραπτού λόγου ήταν κείμενο γραμμένο σε πλακέτες καλυμμένες με κερί. Ο Δεμέρατος αφαίρεσε το κερί από την πλακέτα, έγραψε το μήνυμα στο ξύλο που βρισκόταν από κάτω, πρόσθεσε ξανά το κερί στη πλακέτα ούτως ώστε να φαίνεται κενή και εν τέλει έστειλε το έγγραφο χωρίς να ανακαλυφθεί από κανέναν.

Οι Ρωμαίοι χρησιμοποιούσαν αόρατα μελάνια, τα οποία βασίζονταν σε φυσικές ουσίες όπως χυμούς φρούτων και γάλα. Προκειμένου να αποκαλυφθεί το κρυφό περιεχόμενο, απαιτούνταν η θέρμανση του κρυφού κειμένου. Τα αόρατα μελάνια έχουν γίνει πολύ πιο προηγμένα και χρησιμοποιούνται περιορισμένα ακόμα και σήμερα. Κατά τη διάρκεια του 15^{ου} και 16^{ου} αιώνα, πολλοί συγγραφείς συμπεριλαμβανόμενων των Johannes Trithemius (συγγραφέα του Steganographia) και Gaspari Schotti (συγγραφέα του Steganographica) έγραψαν για τις Στεγανογραφικές τεχνικές όπως η κωδικοποίηση τεχνικών για κείμενα, αόρατα μελάνια, και ενσωμάτωση κρυφών μηνυμάτων σε μουσική.

Μεταξύ 1883 και 1907, περαιτέρω ανάπτυξη μπορεί να αποδοθεί στις εκδόσεις των Auguste Kerckhoff (συγγραφέα του Cryptographic Militaire) και Charles Briquet (συγγραφέα του Les Filigranes). Αυτά τα βιβλία είχαν κυρίως για θέμα τους την Κρυπτογραφία, άλλα και στα δυο μπορεί να αποδοθεί η θεμελίωση κάποιων Στεγανογραφικών συστημάτων και πιο συγκεκριμένα η τεχνική Watermarking.

Κατά τη διάρκεια του 1^{ου} και 2^{ου} Παγκοσμίου Πολέμου, έγιναν σημαντικές πρόοδοι στον τομέα αυτό. Έννοιες όπως η "null ciphers" (απόκρυψη του τρίτου γράμματος κάθε λέξης σε ένα "ακίνδυνο" μήνυμα για την δημιουργία

ενός κρυφού μηνύματος, κτλ), η αλλαγή εικόνων και η “microdot” (συμπίεση δεδομένων όπως εικόνες στο μέγεθος μιας τελείας σε ένα φύλλο χαρτί) εισήχθησαν και έγιναν αποδεκτές ως μεγάλες Στεγανογραφικές τεχνικές.

Στον ψηφιακό κόσμο του σήμερα, δηλαδή από το 1992 μέχρι και σήμερα, η Στεγανογραφία χρησιμοποιείται σε συστήματα υπολογιστών σε όλο τον κόσμο. Πολλά εργαλεία και τεχνολογίες έχουν δημιουργηθεί για να εκμεταλλεύονται παλιές Στεγανογραφικές τεχνικές όπως η “ null ciphers”, η κωδικοποίηση σε εικόνες, βίντεο, ήχο και η “microdot”. Με την έρευνα που γίνεται πάνω σε αυτόν τον τομέα θα δούμε πολλές εξαιρετικές εφαρμογές Στεγανογραφίας στο κοντινό μέλλον.

3.2 Τι είναι Στεγανογραφία

Σε αυτό το κεφάλαιο θα αναλύσουμε την Στεγανογραφία διεξοδικά. Σκοπός μας είναι να ορίσουμε όσο καλύτερα γίνεται το τι θεωρείται τέλεια Στεγανογραφημένη Επικοινωνία.

Αρχικά, ας κοιτάξουμε από τι αποτελείται μια θεωρητικά τέλεια κρυφή επικοινωνία (με χρήση Στεγανογραφίας). Για την επεξήγηση αυτής της έννοιας, θα χρησιμοποιήσουμε τρεις φανταστικούς χαρακτήρες που θα ονομάσουμε Amy, Bret και Crystal [33].

Η Amy θέλει να στείλει ένα κρυφό μήνυμα (M) στον Bret χρησιμοποιώντας ένα τυχαίο (R) “ακίνδυνο” μήνυμα για να δημιουργήσει κάλυψη (C), το οποίο θα σταλεί στον Bret χωρίς να δημιουργήσει υποψίες. Έπειτα η Amy αλλάζει το μήνυμα κάλυψης (C) σε ένα αντικείμενο Στεγανογραφίας (S) ενσωματώνοντας το κρυφό μήνυμα (M) μέσα στο μήνυμα κάλυψης (C) χρησιμοποιώντας ένα κλειδί Στεγανογραφίας (K). Η Amy θα μπορεί τότε να στείλει το αντικείμενο Στεγανογραφίας (S) στον Bret χωρίς να γίνει αντιληπτό από την Crystal. Ο Bret θα μπορεί έπειτα να ξαναπροσθέσει το κρυφό μήνυμα (M) αφού γνωρίζει το κλειδί Στεγανογραφίας (K) που χρησιμοποιήθηκε στο το μήνυμα κάλυψης (C).

Όπως επισημαίνει ο A.P. Petitcolas:

«Σε ένα τέλειο σύστημα, μια συνηθισμένη κάλυψη δεν ξεχωρίζει από ένα αντικείμενο Στεγανογραφίας, ούτε από άνθρωπο αλλά ούτε και από υπολογιστή που ψάχνει για στατιστικά δείγματα».

Στην πράξη, ωστόσο, αυτό δεν είναι το μόνο που πρέπει να ληφθεί υπ’ όψιν. Για να ενσωματωθούν κρυφά δεδομένα σε ένα μήνυμα κάλυψης, η κάλυψη πρέπει να περιέχει ένα επαρκές μέγεθος περιττών δεδομένων η θόρυβο. Αυτό ισχύει διότι η διαδικασία ενσωμάτωσης που χρησιμοποιεί η Στεγανογραφία ουσιαστικά αντικαθιστά αυτά τα περιττά δεδομένα με το κρυφό μήνυμα. Αυτό περιορίζει τους τύπους δεδομένων που μπορούμε να χρησιμοποιήσουμε για Στεγανογραφία. Στην πράξη, χρησιμοποιούνται κατά βάση τρεις τύποι Στεγανογραφικών Πρωτοκόλλων :

1. “Καθαρή Στεγανογραφία”(Pure Steganography)
2. “Στεγανογραφία Κρυφού Κλειδιού”(Secret Key Steganography)
3. “Στεγανογραφία Δημοσίου Κλειδιού”(Public Key Steganography)

Η “Καθαρή Στεγανογραφία” ορίζεται ως ένα Στεγανογραφικό Σύστημα που δεν απαιτεί την ανταλλαγή κρυπτογραφήματος όπως ένα Στεγανογραφικό κλειδί. Αυτή η μέθοδος Στεγανογραφίας είναι η λιγότερο ασφαλής λόγω του ότι η κρυφή επικοινωνία μεταξύ του αποστολέα και του παραλήπτη βασίζεται στην υπόθεση ότι καμία άλλη ομάδα χρηστών δεν γνωρίζει για το κρυφό αυτό μήνυμα. Με την χρησιμοποίηση Ανοικτών Συστημάτων όπως το Ίντερνετ γνωρίζουμε ότι κάτι τέτοιο δεν μπορεί να υφίσταται.

Η “Στεγανογραφία Κρυφού Κλειδιού” ορίζεται ως ένα Στεγανογραφικό σύστημα που απαιτεί να έχει προηγηθεί η ανταλλαγή ενός Κρυφού Κλειδιού (stego-key) για επικοινωνία. Η “Στεγανογραφία Κρυφού Κλειδιού” παίρνει ένα μήνυμα κάλυψης και ενσωματώνει το κρυφό μήνυμα μέσα του χρησιμοποιώντας ένα κρυφό κλειδί (stego-key). Μονό οι ομάδες χρηστών που γνωρίζουν το κρυφό κλειδί μπορούν να αντιστρέψουν τη διαδικασία και να διαβάσουν το κρυφό μήνυμα. Αντίθετα με την “Καθαρή Στεγανογραφία” όπου υπάρχει ένα αντιληπτό αόρατο επικοινωνιακό κανάλι, η “Στεγανογραφία Κρυφού Κλειδιού” ανταλλάσσει κρυφό κλειδί (stego-key), κάτι που την κάνει πιο επιρρεπή σε υποκλοπή. Το πλεονέκτημα της “Στεγανογραφίας Κρυφού Κλειδιού” είναι το ότι ακόμα και αν υποκλαπεί το μήνυμα, μόνο οι ομάδες χρηστών που γνωρίζουν το κρυφό κλειδί μπορούν να αποσπάσουν το κρυφό μήνυμα.

Η “Στεγανογραφία Δημοσίου Κλειδιού” ορίζεται ως ένα Στεγανογραφικό Σύστημα το οποίο χρησιμοποιεί ένα δημόσιο και ένα ιδιωτικό κλειδί για να διασφαλίσει την επικοινωνία μεταξύ ομάδων χρηστών που θέλουν να επικοινωνούν κρυφά. Ο αποστολέας θα χρησιμοποιήσει το δημόσιο κλειδί κατά την διαδικασία κωδικοποίησης και μόνο το ιδιωτικό κλειδί, το οποίο έχει άμεση μαθηματική σχέση με το δημόσιο κλειδί, μπορεί να αποκρυπτογραφήσει το κρυφό μήνυμα. Η “Στεγανογραφία Δημοσίου Κλειδιού” παρέχει έναν πιο σθεναρό τρόπο εφαρμογής ενός Στεγανογραφικού συστήματος λόγω του ότι μπορεί να αξιοποιεί μια πολύ πιο στιβαρή και διερευνημένη τεχνολογία. Επίσης έχει πολλαπλά επίπεδα προστασίας μέσω των οποίων οι ανεπιθύμητες ομάδες χρηστών θα έπρεπε πρώτα να υποψιαστούν την χρήση της Στεγανογραφίας και μετά θα έπρεπε να βρουν ένα τρόπο να σπάσουν τον αλγόριθμο που χρησιμοποιείται από το σύστημα δημοσίου κλειδιού πριν τελικά καταφέρουν να υποκλέψουν το κρυφό μήνυμα.

3.3 Κατηγορίες / Είδη στεγανογραφίας

3.3.1 Κωδικοποίηση Κρυφών Μηνυμάτων σε κείμενο

Η κωδικοποίηση κρυφών μηνυμάτων σε κείμενο μπορεί να γίνει πολύ δύσκολη αποστολή. Αυτό ισχύει λόγω του ότι τα αρχεία κειμένου έχουν έναν πολύ μικρό αριθμό περιττών δεδομένων για να αντικατασταθούν με ένα κρυφό μήνυμα.

Ένα άλλο μειονέκτημα είναι η ευκολία με την οποία η Στεγανογραφία κειμένου μπορεί να μεταβληθεί από ανεπιθύμητες ομάδες χρηστών απλά με την αλλαγή του ίδιου του κειμένου ή την επαναδόμηση του κειμένου σε άλλο τύπο αρχείου (π.χ.TXT σε .PDF).

Υπάρχουν πολυάριθμες μέθοδοι μέσω των οποίων μπορεί να επιτευχθεί Στεγανογραφία κειμένου. Παρακάτω θα παρουσιαστούν κάποιες από τις πιο δημοφιλείς μεθόδους:

- Η κωδικοποίηση μέσω αλλαγής γραμμής περιλαμβάνει την αλλαγή κάθε γραμμής κειμένου καθέτως πάνω ή κάτω με όριο 3 εκατοστών. Με βάση το αν η γραμμή ήταν πάνω η κάτω από την σταθερή γραμμή θα έχουμε μια εξίσωση με μια τιμή που θα μπορούσε να κωδικοποιηθεί σε ένα κρυφό μήνυμα.
- Η κωδικοποίηση μέσω αλλαγής λέξεων δουλεύει σχεδόν με τον ίδιο τρόπο που δουλεύει η κωδικοποίηση μέσω αλλαγής γραμμής, μόνο που χρησιμοποιούμε τα οριζόντια κενά ανάμεσα στις λέξεις για να έχουμε μια εξίσωση με μια τιμή για το κρυφό μήνυμα. Αυτή η μέθοδος κωδικοποίησης είναι λιγότερο ορατή από την κωδικοποίηση μέσω αλλαγής γραμμής αλλά απαιτεί ο τύπος κείμενου να υποστηρίζει μεταβλητή διάταξη.
- Η πιο διαδεδομένη κωδικοποίηση κρυφών μηνυμάτων μέσα σε διαμορφωμένο κείμενο αλλάζοντας συγκεκριμένες ιδιότητες του κειμένου όπως το κάθετο / οριζόντιο μήκος των γραμμάτων (b, d, T, κτλ). Αυτή είναι με διαφορά η μέθοδος κωδικοποίησης κειμένου με την μεγαλύτερη δυσκολία υποκλοπής καθώς κάθε τύπος διαμορφωμένου κειμένου έχει ένα μεγάλο ποσοστό χαρακτηριστικών που μπορούν να χρησιμοποιηθούν για την κωδικοποίηση του κρυφού μηνύματος.

Και οι τρεις αυτές μέθοδοι κωδικοποίησης κείμενου απαιτούν είτε το αρχικό αρχείο είτε τη γνώση των αρχικών αρχείων που διαμορφώθηκαν για να μπορέσει κανείς να αποκρυπτογραφήσει το κρυφό μήνυμα.

3.3.2 Κωδικοποίηση Κρυφών Μηνυμάτων σε Εικόνες

Η κωδικοποίηση κρυφών μηνυμάτων σε ψηφιακές εικόνες είναι με διαφορά η πιο ευρέως διαδεδομένη μέθοδος στον ψηφιακό κόσμο. Αυτό ισχύει λόγω του ότι μπορεί να εκμεταλλευθεί την περιορισμένη ικανότητα του ανθρώπινου συστήματος όρασης (HVS). Σχεδόν κάθε απλό κείμενο, κρυπτογραφημένο κείμενο, εικόνα και κάθε άλλο μέσο που μπορεί να κωδικοποιηθεί σε ένα bit stream μπορεί να κρυφθεί σε μια ψηφιακή εικόνα. Με την συνεχή ανάπτυξη

δυνατών γραφικών στους υπολογιστές και την έρευνα που γίνεται στην Στεγανογραφία μέσω εικόνων, αυτός ο τομέας θα συνεχίσει να αναπτύσσεται με πολύ γρήγορο ρυθμό.

Πριν διεισδύσουμε στις τεχνικές κωδικοποίησης ψηφιακών εικόνων, πρέπει να επεξηγηθεί με συντομία η αρχιτεκτονική των ψηφιακών εικόνων και οι τεχνικές συμπίεσης ψηφιακών εικόνων.

Όπως εξηγεί ο Duncan Sellars:

"Για έναν υπολογιστή, μια εικόνα είναι μια παράταξη αριθμών που αναπαριστάνουν ελαφρές εντάσεις σε διάφορα σημεία, ή pixels. Αυτά τα pixels αντισταθμίζουν τα raster data των εικόνων".

Όταν έχουμε να κάνουμε με Στεγανογραφία ψηφιακών εικόνων, τα εικονικά αρχεία 8-bit και 24-bit ανά pixel είναι ενδεικτικά. Και τα δυο έχουν πλεονεκτήματα και μειονεκτήματα, όπως θα εξηγήσουμε παρακάτω.

Οι εικόνες 8-bit είναι ένας πολύ καλός τύπος για να χρησιμοποιήσουμε λόγω του ότι είναι σχετικά μικρές σε μέγεθος. Το μειονέκτημα είναι ότι μπορούν χρησιμοποιηθούν μόνο 256 πιθανά χρώματα, το οποίο μπορεί να αποτελέσει πρόβλημα κατά την κωδικοποίηση. Συνήθως χρησιμοποιείται γκρίζα παλέτα χρωμάτων όταν έχουμε να κάνουμε με εικόνες 8-bit (όπως GIF.) διότι η σταδιακή αλλαγή χρώματος θα είναι δυσκολότερο να ανιχνευθεί μετά την κωδικοποίηση της εικόνας με το κρυφό μήνυμα.

Οι εικόνες 24-bit προσφέρουν περισσότερη ευελιξία όταν χρησιμοποιούνται για Στεγανογραφία. Ο μεγάλος αριθμός χρωμάτων (πάνω από 16 εκατομμύρια) που μπορούν να χρησιμοποιηθούν είναι κατά πολύ πάνω από το ανθρωπινό σύστημα όρασης (HVS), κάτι το οποίο κάνει αυτομάτως πολύ δύσκολη την ανίχνευση ενός κρυφού μηνύματος. Ένα άλλο πλεονέκτημα είναι ότι σε μια ψηφιακή εικόνα 24-bit μπορεί να κωδικοποιηθεί ένας πολύ μεγαλύτερος αριθμός κρυφών δεδομένων, σε αντίθεση με μια ψηφιακή εικόνα 8-bit.Το μοναδικό σημαντικό μειονέκτημα στις ψηφιακές εικόνες 24-bit είναι ότι το μεγάλο μέγεθος τους (συνήθως σε MB) τις κάνει πιο ύποπτες σε σχέση με τις πολύ μικρότερες ψηφιακές εικόνες 8-bit (συνήθως σε KB) όταν στέλνονται σε ένα ανοικτό σύστημα όπως είναι το Ίντερνετ.

Η συμπίεση των ψηφιακών εικόνων είναι μια καλή λύση αναφορικά με τις μεγάλες ψηφιακές εικόνες όπως οι 24-bit που αναφέρθηκαν προηγουμένως. Υπάρχουν 2 είδη συμπίεσης για τις ψηφιακές εικόνες, η "lossy" και η "lossless".

Η συμπίεση "lossy" (JPEG) μειώνει σημαντικά το μέγεθος μιας ψηφιακής εικόνας μέσω της απομάκρυνσης περιττών δεδομένων της εικόνας και του υπολογισμού σε προσέγγιση του αρχικού αρχείου. Η συμπίεση "lossy" χρησιμοποιείται συνήθως για ψηφιακές εικόνες 24-bit και μείωση του μεγέθους τους, αλλά έχει ένα σημαντικό μειονέκτημα. Οι τεχνικές συμπίεσης "lossy" αυξάνουν την πιθανότητα το ασυμπίεστο κρυφό μήνυμα να χάσει

κομμάτια από τα περιεχόμενα του λόγω του ότι η συμπίεση “lossy” απομακρύνει ότι “βλέπει” ως περιττά δεδομένα της εικόνας.

Η τεχνική συμπίεσης “lossless”, όπως συνιστά το όνομα, διατηρεί το αρχείο ανέπαφο χωρίς την πιθανότητα απώλειας δεδομένων. Γι αυτό το λόγο αυτή η τεχνική συμπίεσης επιλέγεται για Στεγανογραφική χρήση. Παραδείγματα συμπίεσης “lossless” είναι τα αρχεία τύπου .GIF και .BMP. Το μοναδικό μειονέκτημα της συμπίεσης εικόνων “lossless” είναι ότι δεν είναι πολύ αποτελεσματικό στην συμπίεση του μεγέθους της εικόνας.

Κάποιες από τις πιο διαδεδομένες τεχνικές κωδικοποίησης ψηφιακών εικόνων του σήμερα είναι:

- Η τεχνική “least significant bit (LSB) encoding”
- Η τεχνική “masking and filtering encoding”.

Η τεχνική “least significant bit (LSB) encoding” είναι με διαφορά η πιο διαδεδομένη τεχνική κωδικοποίησης που χρησιμοποιείται για ψηφιακές εικόνες. Χρησιμοποιώντας την LSB για κάθε byte (8 bits) σε μια εικόνα για ένα κρυφό μήνυμα, μπορεί κανείς να αποθηκεύσει 3 bits δεδομένων σε κάθε pixel εικόνων 24-bit και 1 bit σε κάθε pixel εικόνων 8-bit. Όπως είναι προφανές, σε εικονικό αρχείο 24-bit μπορούν να αποθηκευθούν πολύ περισσότερες πληροφορίες.

Με βάση το χρώμα παλέτας που χρησιμοποιείται για την εικόνα κάλυψης (π.χ. γκρίζα παλέτα), είναι πιθανόν να πάρουμε 2 LSB από 1 byte χωρίς το ανθρωπινό σύστημα όρασης (HVS) να μπορεί να καταλάβει τη διαφορά.

Το μοναδικό πρόβλημα με αυτή την τεχνική είναι ότι είναι πολύ ευπαθές σε επιθέσεις όπως αλλαγές και διαμόρφωση των εικόνων (π.χ. αλλαγή από.GIF σε .JPEG).

Οι τεχνικές κωδικοποίησης ψηφιακής εικόνας “masking and filtering” όπως η τεχνική “Watermarking” (πχ.- ταύτιση του σήματος μιας εταιρίας με το διαδικτυακό της περιεχόμενο) ταυτίζονται με τεχνικές συμπίεσης “lossy” (όπως η JPEG).Αυτή η τεχνική πρακτικά αυξάνει τα δεδομένα μιας εικόνας με το να καλύπτει τα κρυφά δεδομένα πάνω από το αρχικό αρχείο και όχι με το να κρύβει πληροφορίες μέσα στα δεδομένα. Κάποιοι ειδικοί ισχυρίζονται ότι αυτή είναι αναμφίβολα μια μορφή κάλυψης πληροφοριών, αλλά τεχνικά δεν είναι Στεγανογραφία. Η γοητεία των τεχνικών “masking and filtering” είναι ότι είναι απρόσβλητες από παραποίηση εικόνων, κάτι που κάνει τις πιθανές τους χρήσεις πολύ στιβαρές.

Σαν επιπρόσθετη σημείωση, υπάρχουν πολλές άλλες, λιγότερο διαδεδομένες, τεχνικές που δεν καλύπτονται σε αυτή την εργασία και μπορούν να διερευνηθούν από τον καθένα που ενδιαφέρεται για την χρησιμοποίηση ψηφιακών εικόνων για Στεγανογραφικούς σκοπούς. Τεχνικές που χρησιμοποιούν περίπλοκους αλγορίθμους, μετασχηματισμό εικόνων και απόκρυψη εικόνων είναι σχετικά νέες, αλλά έχουν προοπτικές να γίνουν πιο

ασφαλείς και στιβαροί τρόποι χρησιμοποίησης ψηφιακών εικόνων στην Στεγανογραφία.

3.3.3 Κωδικοποίηση κρυφών μηνυμάτων σε Ήχο.

Η κωδικοποίηση κρυφών μηνυμάτων σε ήχο είναι η πιο απαιτητική τεχνική που μπορεί να χρησιμοποιήσει κανείς στην Στεγανογραφία. Αυτό ισχύει λόγω του ότι το ανθρώπινο ακουστικό σύστημα (HAS) έχει μεγάλο δραστικό βεληνεκές. Για να γίνει αυτό πιο κατανοητό, το HAS αντιλαμβάνεται μια κλίμακα με δυναμική μεγαλύτερη του 1.000.000 προς 1 και μια κλίμακα συχνοτήτων μεγαλύτερη των 1000 προς 1, κάνοντας το υπερβολικά δύσκολο να προσθέσει ή να αφαιρέσει κανείς δεδομένα από την αρχική δομή δεδομένων.

Το μοναδικό μειονέκτημα του HAS παρουσιάζεται στην προσπάθεια της διαφοροποίησης των ήχων και αυτό είναι που πρέπει να αξιοποιηθεί για την κωδικοποίηση κρυφών μηνυμάτων σε ήχο χωρίς να ανιχνεύεται. Υπάρχουν δυο έννοιες που πρέπει να μελετηθούν πριν την επιλογή τεχνικής κωδικοποίησης ήχου. Αυτές είναι η ψηφιακή μορφή του ήχου και το μέσο μετάδοσης του ήχου. Υπάρχουν τρεις (3) κύριες μορφές ψηφιακού ήχου που χρησιμοποιούνται στην πράξη. Αυτές είναι οι "Sample Quantization", "Temporal Sampling Rate" και "Perceptual Sampling".

- Η μορφή "Sample Quantization", που αποτελεί μια 16-bit γραμμική αρχιτεκτονική sampling και χρησιμοποιείται από διαδεδομένες μορφές ήχου όπως WAV ή AIFF.
- Η μορφή "Temporal Sampling Rate" χρησιμοποιεί επιλεγμένες συχνότητες (σε KHz) για να δειγματίσει τον ήχο. Γενικότερα , όσο πιο υψηλός είναι ο ρυθμός sampling, τόσο πιο πολύ μεγαλώνει ο χώρος των δεδομένων.
- Η τελευταία μορφή ήχου είναι η "Perceptual Sampling". Αυτή η μορφή τα μεταβάλει στατιστικά του ήχου δραστικά, κωδικοποιώντας Μονό τα κομμάτια που δέχεται ο ακροατής, έτσι διατηρείται ο ήχος αλλά αλλάζει το σήμα. Αυτή η μορφή χρησιμοποιείται από τους πιο διαδεδομένους ψηφιακούς ήχους που υπάρχουν σήμερα στο Ίντερνετ (ISO, MPEG, MP3). Το μέσο μετάδοσης (η διαδρομή που κάνει ο ήχος από τον αποστολέα στον παραλήπτη) πρέπει επίσης να ληφθεί υπ' όψιν κατά την κωδικοποίηση κρυφών μηνυμάτων σε ήχο. Ο W. Bender [8] παρουσιάζει τέσσερα (4) πιθανά μέσα μετάδοσης:

1) "Digital end to end" – από συσκευή σε συσκευή χωρίς διαφοροποίηση.

2) "Increased/decreased resampling"- ο ρυθμός sampling μεταβάλλεται αλλά παραμένει ψηφιακός.

3) "Analog and resampled"- το σήμα αλλάζει σε αναλογικό και επαναδειγματίζεται σε διαφορετικό ρυθμό.

4) "Over the air "- το σήμα μεταδίδεται σε ραδιοφωνικές συχνότητες και επαναδειγματίζεται από ένα μικρόφωνο.

Τώρα θα κοιτάξουμε τρεις από τις πιο διαδεδομένες μεθόδους κωδικοποίησης για απόκρυψη δεδομένων σε ήχο. Αυτές είναι οι “low-bit encoding”, “phase-coding” και “spread spectrum”.

- Η μέθοδος “low-bit encoding” ενσωματώνει κρυφά δεδομένα στο LSB του ηχητικού αρχείου. Η χωρητικότητα του καναλιού είναι 1 KB ανά δευτερόλεπτο ανά kilohertz (44 kbps για 44 KHz αλληλουχίας). Αυτή η μέθοδος είναι εύκολο να ενσωματωθεί αλλά είναι ταυτόχρονα πολύ επιρρεπής στην απώλεια δεδομένων λόγω του θορύβου στο κανάλι αλλά και του resampling.
- Η μέθοδος "Phase coding" αντικαθιστά την φάση ενός αρχικού τμήματος του ήχου με μια φάση “αναφοράς” (reference) που αναπαριστά τα κρυφά δεδομένα. Αυτό μπορεί να θεωρηθεί ως ένα είδος απόκρυψης του ηχητικού σήματος χρησιμοποιώντας κάτι γνωστό ως “Discrete Fourier Transform” (DFT), το οποίο δεν είναι τίποτε παραπάνω από μια μετατροπή του αλγόριθμου του ηχητικού σήματος.
- Η μέθοδος “spread spectrum” κωδικοποιεί τον ήχο σχεδόν σε όλο το φάσμα των συχνοτήτων. Μεταδίδει τον ήχο σε διαφορετικές συχνότητες οι οποίες ποικίλλουν ανάλογα με το ποια μέθοδος μετάδοσης φάσματος χρησιμοποιείται. Η “Direct Sequence Spread Spectrum” (DSSS) είναι μια τέτοια μέθοδος που μεταδίδει το σήμα πολλαπλασιάζοντας το σήμα της Πηγής μέσω κάποιας τυχαίας αλληλουχίας γνωστής ως “CHIP”. Κατόπιν χρησιμοποιείται το sampling rate ως chip rate για την επικοινωνία ηχητικού σήματος. Οι τεχνικές κωδικοποίησης “spread spectrum” είναι το πιο ασφαλές μέσο για αποστολή κρυφών μηνυμάτων σε ήχο, αλλά μπορεί να δημιουργήσει τυχαίο θόρυβο στο αρχείο και κατά συνέπεια μπορεί να δημιουργηθεί η πιθανότητα απώλειας δεδομένων.

Υπάρχουν πολλές εφαρμογές για την Στεγανογραφία, κάποιες για καλό και κάποιες για κακό σκοπό, που οδηγούν στην ανάγκη για αποτελεσματικό της χρήσης της μέσα από τεχνικές Στεγανάλυσης (Steganalysis) σε Forensic Investigation που θα περιγραφούν σε επόμενο κεφάλαιο.

3.3.4 Εφαρμογές Στεγανογραφίας σε ένα Περιβάλλον Ανοικτών Συστημάτων

Σε αυτό το κεφάλαιο θα εξετάσουμε κάποιες από τις πιθανές εφαρμογές Στεγανογραφίας.

Οι τρεις πιο διαδεδομένες και διερευνημένες χρήσεις της στεγανογραφίας σε περιβάλλον ανοικτών συστημάτων είναι οι: “covert channels”, “embedded data” και “digital watermarking”.

- Η χρήση “covert channels” σε TCP/IP προϋποθέτει την κάλυψη Πληροφοριών ταυτότητας για την απόκρυψη της αληθινής ταυτότητας ενός η περισσότερων συστημάτων. Αυτό μπορεί να είναι πολύ χρήσιμο για τις ανάγκες κάθε ασφαλούς επικοινωνίας σε ανοικτά συστήματα όπως είναι το Ίντερνετ όταν χρειάζεται απόλυτη

μυστικότητα για μια ολόκληρη διαδικασία επικοινωνίας και όχι ένα μεμονωμένο έγγραφο, κάτι που αναφέρεται στη συνέχεια.

- Η χρήση μηνυμάτων κάλυψης (containers) για ενσωμάτωση κρυφών μηνυμάτων είναι με διάφορα η πιο διαδεδομένη χρήση της Στεγανογραφίας σήμερα. Αυτή η μέθοδος Στεγανογραφίας είναι πολύ χρήσιμη όταν μια ομάδα χρηστών πρέπει να στείλει ένα απόρρητο, ιδιωτικό η "ευαίσθητο" έγγραφο σε ένα περιβάλλον ανοικτών συστημάτων όπως το Ίντερνετ. Με την ενσωμάτωση των κρυφών δεδομένων μέσα σε ένα κρυφό μήνυμα και την αποστολή του, ο αποστολέας έχει μια αίσθηση προστασίας λόγω του ότι κανείς δεν γνωρίζει ότι έχει στείλει κάτι περισσότερο από ένα "ακίνδυνο" μήνυμα εκτός από τους επιδιωκόμενους παραλήπτες. Αυτή είναι η χρήση της Στεγανογραφίας στην οποία θα εστιάσουμε την ερευνητική μας προσπάθεια στα πλαίσια αυτής της εργασίας.
- Αν και δεν είναι αυθεντική τεχνική Στεγανογραφίας, το "digital watermarking" είναι πολύ συνηθισμένο στον σημερινό κόσμο και χρησιμοποιεί Στεγανογραφικές τεχνικές για την ενσωμάτωση πληροφοριών σε έγγραφα.Το "digital watermarking" χρησιμοποιείται κυρίως για λόγους πνευματικής ιδιοκτησίας από εταιρίες ή πρόσωπα που επιθυμούν να προστατεύσουν την ιδιοκτησία τους είτε με την ενσωμάτωσή της εμπορικής τους επωνυμίας στην περιουσία τους είτε με την απόκρυψη σειριακών αριθμών / Πληροφοριών άδειας σε λογισμικό κτλ. Το "digital watermarking" παίζει πολύ σημαντικό ρόλο για την ανίχνευση και την δίωξη πειρατικών λογισμικών / ψηφιακών κλεφτών.

3.4 Η αξία της στεγανάλυσης σε Forensics Investigations.

Στεγανάλυση είναι τέχνη και η επιστήμη που ασχολείται με την διακοπή ή την ανίχνευση όλων των Στεγανογραφικών Τεχνικών.

Οι Στεγανογραφικές εφαρμογές είναι συνήθως ουδέτερες. Ωστόσο, αν χρησιμοποιηθούν με ανάρμοστο τρόπο μπορούν να δημιουργήσουν σοβαρούς κινδύνους, π.χ. αν τρομοκρατικές οργανώσεις χρησιμοποιούσαν Στεγανογραφικά εργαλεία για κρυφή επικοινωνία, αυτό θα έκανε την δουλειά των οργάνων ασφαλείας πολύ δύσκολη αναφορικά με τον εντοπισμό τέτοιας επικοινωνίας. Σε μια προσπάθεια να διευθετηθεί αυτό το θέμα μπορούν να χρησιμοποιηθούν τεχνικές Στεγανάλυσης.
Στεγανάλυση είναι η τέχνη και η επιστήμη εντοπισμού κρυφών μηνυμάτων που δημιουργήθηκαν με τεχνικές Στεγανογραφίας. Μόλις εντοπιστεί η παρουσία Στεγανογραφικού υλικού σε κάποια ηλεκτρονικά μέσα (εικόνες, ήχοι, βίντεο, δικτυακά πακέτα κτλ), τα «μολυσμένα» μέσα μπορούν να απομονωθούν για περαιτέρω ανάλυση ή το Στεγανογραφικό υλικό μπορεί να καταστραφεί σβήνοντας τα «μολυσμένα» μέσα. Τα εργαλεία Στεγανάλυσης είναι πολύ σημαντικά για έναν Εγκληματολογικό Ερευνητή γιατί του δίνουν τη δυνατότητα να αναλύει, εντοπίζει ή ακόμα και να καταστρέφει την κρυφή επικοινωνία.

Βέβαια βασικός σκοπός μίας έρευνας Computer Forensics είναι ο εντοπισμός κάποιου ατόμου που έχει διαπράξει εγκληματική ή άλλη παράνομη ενέργεια

με χρήση υπολογιστικών συστημάτων αλλά και η ανάκτηση της κρυφής επικοινωνίας σαν αποδεικτικό στοιχείο. Φυσικά αυτή η διαδικασία είναι επίπονη και συχνά αντιμετωπίζονται προβλήματα που μπορεί να αφορούν μία από τις τέσσερις παρακάτω περιπτώσεις:

- Τα ύποπτα αρχεία μπορούν να έχουν ή να μην έχουν καθόλου κρυφή πληροφορία μέσα τους.
- Η κρυμμένη πληροφορία μπορεί να είναι ήδη κρυπτογραφημένη.
- Τα κρυμμένα αρχεία μπορεί να είναι τελικά άσχετα σε σχέση με τον σκοπό της έρευνας. Τα άσχετα αυτά αρχεία προκαλούν μεγάλα προβλήματα καθώς η εξαγωγή τους απαιτεί πολύ χρόνο (προκειμένου να διαπιστωθεί εάν σχετίζονται με την υπόθεση υπό διερεύνηση ή όχι), ο οποίος θα μπορούσε να έχει χρησιμοποιηθεί δημιουργικά για τον εντοπισμό διαφορετικού είδους στοιχείων. Κρυπτογράφηση των κρυφών αυτών αρχείων προκαλεί επιπλέον απώλεια πολύτιμου χρόνου.
- Εκτός και εάν ολοκληρωθεί πλήρως η διαδικασία της στεγανάλυσης (ανάκτηση, αποκρυπτογράφηση, αποκωδικοποίηση) του υπόπτου αρχείου, δεν μπορεί να διαπιστωθεί εάν το αρχείο που στεγαναλύσαμε είναι χρήσιμο για την έκβαση μίας συγκεκριμένης έρευνας.

Η βασική διαφορά μεταξύ της Στεγανάλυσης και της Κρυπτανάλυσης είναι ότι η πληροφορία που βρίσκεται υπό ανάλυση στην περίπτωση της Στεγανάλυσης είναι τεράστιος όγκος αρχείων από τα οποία δεν γνωρίζουμε εξαρχής πια είναι στεγανογραφημένα σε αντίθεση με την Κρυπτανάλυση στην οποία γνωρίζουμε ότι το αρχείο που αναλύουμε είναι σίγουρα κρυπτογραφημένο.

Η διαδικασία της Στεγανάλυσης ξεκινά πάντα με μία συλλογή υπόπτων αρχείων. Ο στεγαναλυτής, το άτομο δηλαδή που διενεργεί Στεγανάλυση δεν γνωρίζει εξαρχής ποιά ή πόσα αρχεία περιέχουν Στεγανογραφία. Βασικός σκοπός του είναι μέσω μαθηματικών και στατιστικών μεθόδων να μειώσει όσο το δυνατόν περισσότερο την συλλογή των υπόπτων στοιχείων εξαιρώντας αρχεία που θεωρεί πως σίγουρα δεν περιέχουν κρυφή πληροφορία.

Πολλές φορές είναι πιθανό σε μία έρευνα όλα τα αρχεία του υπολογιστή ενός υπόπτου να θεωρηθούν ύποπτα για Στεγανογραφία άρα απαιτούνται γρήγορες και αξιόπιστες τεχνικές στεγανάλυσης προκειμένου να μειωθεί ο αριθμός των υπόπτων αρχείων σε εκείνα που πραγματικά περιέχουν κρυφή πληροφορία. Αυτός ακριβώς είναι και ο σκοπός της εργασίας αυτής.

Η Στεγανογραφία / Στεγανάλυση στην Πράξη

SAO PAULO, 10 Μαρτίου, 2008 —Το διάσημο ανά τον κόσμο Ιαπωνικό καρτούν «Hello Kitty»[116], χρησιμοποιήθηκε από αδίστακτο Κολομβιανό Έμπορο ναρκωτικών για να κρύψει μηνύματα που έστελνε σε συνεργάτες του.

Ο Juan Carlos Ramirez Abadia, που κρατούνταν στην Βραζιλία μετά την σύλληψή του των Αύγουστο του 2008, στεγανογραφούσε μηνύματα ήχου και

κειμένου σε αθώες κατά τα άλλα εικόνες του καρτούν τις οποίες ύστερα επισύναπτε και προωθούσε με email σε συνεργάτες του.

Οι ερευνητές εντόπισαν εκατοντάδες τέτοιες φωτογραφίες στον υπολογιστή του υπόπτου, ενώ σε μερικές από αυτές υπήρχαν σαφείς εντολές προς τους συνεργάτες του για αγοραπωλησίες κοκαΐνης μεταξύ διαφόρων χωρών.

Ίδια τεχνική άλλωστε είχε χρησιμοποιηθεί και από την γνωστή τρομοκρατική[34] οργάνωση Al-Qaeda για την προετοιμασία της επίθεσης της 11 Σεπτεμβρίου το 2001 με χρήση στεγανογραφημένων εικόνων που δημοσιεύτηκαν σε ιστοσελίδα ηλεκτρονικών δημοπρασιών (EBAY).

3.5 Κατηγορίες / Είδη Στεγανάλυσης.

Στην Στεγανάλυση, ο στόχος είναι να μπορούμε να συγκρίνουμε το cover-object (μήνυμα κάλυψης), το stego-object (το μήνυμα κάλυψης με τα ενσωματωμένα κρυφά δεδομένα) καθώς και πιθανά τμήματα του stego-key (μέθοδος κάλυψης) σε μια προσπάθεια να διακόψουμε, αναλύσουμε η / και να καταστρέψουμε την κρυφή επικοινωνία. Όπως τονίζει στο βιβλίο του ο Fabien A.P. Petitcolas, υπάρχουν έξι (6) γενικά πρωτόκολλα που χρησιμοποιούνται κατά της Στεγανογραφίας[1]:

1) Stego only attack – Μόνο το μήνυμα κάλυψης με τα ενσωματωμένα κρυφά δεδομένα είναι διαθέσιμα για ανάλυση.

2) Known cover attack – Το αρχικό μήνυμα κάλυψης (cover object) αλλά και το μήνυμα κάλυψης με τα ενσωματωμένα κρυφά δεδομένα (stego object) είναι διαθέσιμα για ανάλυση

3) Known message attack – Το κρυφό μήνυμα είναι διαθέσιμο για σύγκριση με το μήνυμα κάλυψης με τα ενσωματωμένα κρυφά δεδομένα (stego object).

4) Chosen stego attack – Ο αλγόριθμος (stego tool) και το μήνυμα κάλυψης με τα ενσωματωμένα κρυφά δεδομένα (stego-object) είναι διαθέσιμα για ανάλυση.

5) Chosen message attack – Έχοντας ένα επιλεγμένο μήνυμα δημιουργούμε ένα μήνυμα κάλυψης με ενσωματωμένα κρυφά δεδομένα (stego-object) για μελλοντική ανάλυση.

6) Known stego attack – Στον αλγόριθμος (stego tool), το μήνυμα κάλυψης (cover object) και το μήνυμα κάλυψης με ενσωματωμένα κρυφά δεδομένα (stego-object) είναι διαθέσιμα για ανάλυση.

3.6 Εργαλεία Στεγανογραφίας & Στεγανάλυσης.

Σε αυτή την εργασία, ερευνούμε περισσότερα από 100 διαφορετικά εργαλεία Στεγανογραφίας και Στεγανάλυσης τα οποία μπορούν να χρησιμοποιηθούν για να κρύψουν, εντοπίσουν ή καταστρέψουν ενσωματωμένη κρυφή επικοινωνία

σε μια σειρά ψηφιακών μέσων όπως εικόνων, ήχων, βίντεο, κειμένων, βάσεων δεδομένων, συστημάτων αρχείων, σκληρών δίσκων, αρχείων εφαρμογών και ψηφιακών πακέτων. Η παρουσίαση των εργαλείων είναι δομημένη με τον ακόλουθο τρόπο :

- Ενότητα 3.6.1, περιγράφει Στεγανογραφικά εργαλεία εικόνων.
- Ενότητα 3.6.2, περιγράφει Στεγανογραφικά εργαλεία ήχων.
- Ενότητα 3.6.3, περιγράφει Στεγανογραφικά εργαλεία κειμένου.
- Ενότητα 3.6.4, περιγράφει Στεγανογραφικά εργαλεία συστημάτων και σκληρού δίσκου.
- Ενότητα 3.6.5, περιγράφει Στεγανογραφικά εργαλεία ποικίλης ύλης.
- Ενότητα 3.6.6, περιγράφει Στεγαναλυτικά εργαλεία.

3.6.1 Στεγανογραφικά Εργαλεία Εικόνων

Image Steganographic Tools	JPEG	BMP	Others	Embedding Approach	Production
Blindside		Yes		SDS	Yes
Camera Shy	Yes			SDS	Yes
dc-Steganograph			PCX	TDS	
F5	Yes	Yes	GIF	TDS	Yes
Gif Shuffle			GIF	Change the order of the color map	Yes
Hide4PGP		Yes		SDS	Yes
JP Hide and Seek	Yes			SDS	Yes
Jsteg Jpeg	Yes			SDS	Yes
Mandelsteg			GIF	SDS	Yes
OutGuess	Yes		PNG	TDS	Yes
PGM Stealth			PGM		Yes
Steghide		Yes		SDS	Yes
wbStego		Yes		SDS	Yes
WnStorm			PCX		Yes

TDS - Transform Domain Steganography

SDS - Spatial Domain Steganography (LSB Replacement and LSB Matching)

Εικόνα 3.1: Συγκριτικά στοιχεία για Στεγανογραφικά εργαλεία εικόνων. [33]

Σε αυτή την κατηγορία Στεγανογραφικών εργαλείων ερευνήσαμε 14 διαφορετικά προϊόντα ελεύθερης διακίνησης (παρουσιάζονται στην Εικόνα 3.1) και τριάντα τέσσερα προϊόντα εμπορίου (παρουσιάζονται στην Εικόνα 3.2). Για τα προϊόντα ελεύθερης διακίνησης οι τύποι αρχείων JPEG και BMP αποτελούν την προτιμώμενη επιλογή μέσου κάλυψης καθώς 9 προϊόντα προσφέρουν λειτουργικότητα για ενσωμάτωση σε αυτούς τους τύπους εικόνων. Ο επόμενος προτιμώμενος τύπος αρχείου είναι αυτός του GIF, όπου τα εργαλεία F5[15][109], GifShuffle[110] και Mandlesteg είναι χρήσιμα. Τα Wnstorm and dc-Steganograph[108] ενσωματώνουν πληροφορίες με αρχεία PCX ενώ τα OutGuess[42] και PGMStealth[43] χρησιμοποιούν τους τύπους

αρχείων PNG και PGM. Από τα 14 εργαλεία, τα περισσότερα εξ' αυτών ενσωματώνουν πληροφορίες σε "spatial domain", π.χ. αντικαθιστώντας ή αλλάζοντας τις τιμές των pixel, ενώ τα dc-Steganograph, F5[15] και OutGuess ενσωματώνουν στο "Transform Domain" π.χ. διαχειριζόμενα τους συντελεστές του "Transform Domain".
Και τα 3 αυτά εργαλεία μετατρέπουν τους συντελεστές "Discrete Cosine Transform (DCT)" για να ενσωματώσουν τα κρυφά δεδομένα.

Στο ελεύθερο υλικό αλλά και το υλικό διαμοιρασμού η πιο διαδεδομένη εικόνα κάλυψης είναι η BMP ακολουθούμενη από τις JPEG, GIF, PNG, TGA, TIF, PPM, PCX and DIB. Ένα σύνολο 20 εργαλείων μπορεί να ενσωματώσει σε εικόνες BMP ακολουθούμενα από 10 σε JPEG και 9 σε GIF. Από τα 34 εργαλεία, 17 από αυτά βρίσκονται στην παραγωγή π.χ. η ιστοσελίδα ήταν προσβάσιμη και η τελευταία έκδοση του εργαλείου ήταν διαθέσιμη. Αυτό περιλαμβάνει τα:

Contraband Hell[53], Contraband, Crypto123[45], Dound[54], Gif it Up, Camouflage[55], Hide and Seek[56], InThePicture[57], Hermetic Stego[46], IBM DLS[47], Invisible Secrets[48], S-Tools[58], Jpegx[59], Info Stego[49], Syscop[50], StegMark, Steganos[60].

Από αυτά το StegMark έχει ιδιαίτερο ενδιαφέρον καθώς μπορεί να ενσωματώσει σε τύπους αρχείων BMP, JPEG, GIF, PNG, TGA και TIF. Από τα τριάντα τέσσερα εργαλεία, τα οχτώ είναι διαμοιρασμού, τα δέκα ελεύθερα και τα υπόλοιπα εργαλεία είτε δεν βρίσκονται στην παραγωγή είτε οι πληροφορίες αδείας δεν ήταν διαθέσιμες την περίοδο της ερευνάς.

Image Steganographic Tools	BMP	JPEG	GIF	PNG	TGA	Other	Production	License
Crypto123	Yes	Yes					Yes	S
Hermetic Stego	Yes						Yes	S
IBM DLS	Yes	Yes	Yes	Yes			Yes	S
Invisible Secrets	Yes	Yes		Yes			Yes	S
Info Stego	Yes	Yes	Yes				Yes	S
Syscop		Yes					Yes	S
StegMark	Yes	Yes	Yes	Yes	Yes	TIF	Yes	S
Cloak	Yes							S
Contraband Hell	Yes						Yes	F
Contraband	Yes						Yes	F
Dound	Yes						Yes	F
Gif it Up			Yes				Yes	F
Camouflage				Yes	Yes		Yes	F
Hide and Seek	Yes		Yes				Yes	F
InThePicture	Yes						Yes	F
S-Tools	Yes						Yes	F
Jpegx		Yes					Yes	F
Steganos	Yes					DIB	Yes	F
BMP Secrets	Yes							
DCT-Steg		Yes						
Digital Picture Envelope	Yes							
EikonAmark		Yes						
Empty Pic			Yes					
Encrypt Pic	Yes							
EzStego			Yes					
BMP Embed	Yes							
BMPTable	Yes							
StegoTif					Yes	TIF		
Hide Unhide						TIF		
In Plain View	Yes							
Invisible Encryption			Yes					
JK-PGS						PPM		
Scytale						PCX		
appendX		Yes	Yes	Yes				
Total	20	10	9	5	3	6	17	

S – Shareware License
F – Freeware License

Εικόνα 3.2[33]

3.6.2 Στεγανογραφικά Εργαλεία Ήχου

Σε αυτή τη κατηγορία Στεγανογραφικών εργαλείων ερευνήσαμε δέκα (10) διαφορετικά προϊόντα.
Αυτά παρουσιάζονται στην Εικόνα 3.3. Υπάρχουν πέντε(5) προϊόντα ελεύθερης διακίνησης ενώ άλλα είναι εμπορικά ή προγράμματα διαμοιρασμού με χρονικό περιορισμό λειτουργικότητας. Η πλειοψηφία των προϊόντων αυτών κρύβει δεδομένα σε τύπο αρχείων WAV ενώ τρία (3) από αυτά κρύβουν δεδομένα σε τύπο αρχείων MP3. Προϊόντα όπως τα Hide4PGP[111] και Steganos παρέχουν Στεγανογραφική δυνατότητα για άλλους τύπους αρχείων

όπως VOC ενώ τα StegMark και StegHide[44] ενσωματώνουν κρυφά δεδομένα σε τύπους αρχείων .MIDI και .AU.

Audio Steganographic Tools	MP3	WAV	Others	Production	License
Info Stego	Yes			Yes	Shareware
ScramDisk		Yes		Yes	Shareware
MP3Stego	Yes			Yes	Open Source
StegoWav		Yes		Yes	Open Source
Hide4PGP	Yes		VOC	Yes	Open Source
Steghide		Yes	AU	Yes	Open Source
S-Tool		Yes		Yes	Open Source
Invisible Secrets		Yes		Yes	Commercial
Paranoid			Yes	Yes	Commercial
Steganos		Yes	VOC	Yes	Commercial

Εικόνα 3.3[33]

Το Invisible Secrets[48] είναι ένα εμπορικό προϊόν το οποίο μπορεί να ενσωματώσει σε διαφορετικά μέσα κάλυψης όπως σε εικόνες, ήχο και κείμενο. Στο σύνολο, τους αυτά τα δέκα εργαλεία βρίσκονται στην παραγωγή και η πιο πρόσφατη έκδοση τους ήταν διαθέσιμη κατά τη διάρκεια της έρευνας.

3.6.3 Στεγανογραφικά Εργαλεία Κειμένου

Στον τομέα της Στεγανογραφίας κειμένου ερευνήσαμε ένα σύνολο δεκαπέντε εργαλείων, στα οποία περιλαμβάνονται προϊόντα ελεύθερης χρήσης και διακίνησης, προϊόντα διαμοιρασμού αλλά και προϊόντα του εμπορίου.
Αυτά παρουσιάζονται στην Εικόνα 3.4. Η πλειοψηφία των προϊόντων αυτών ενσωματώνουν κρυφές πληροφορίες μέσα σε απλό κείμενο, ωστόσο με κάποιες εξαιρέσεις όπως τα wbStego, Steganos και Invisible Secrets τα οποία ενσωματώνουν σε σελίδες HTML. Η ευρωστία των εργαλείων αυτών πρέπει να δοκιμαστεί περαιτέρω, δεν είναι αυτή τη στιγμή ξεκάθαρο αν μπορούν να διατηρήσουν μορφοποίηση HTML από προγράμματα όπως το Dreamweaver.
Επίσης, το wbStego ενσωματώνει μηνύματα και σε έγγραφα PDF κάνοντας το με διαφορά το μοναδικό προϊόν που προσφέρει Στεγανογραφική δυνατότητα για τύπους αρχείων .PDF. Υπάρχουν πολλά άλλα προγράμματα που ενσωματώνουν "visible watermark" σε .PDF άλλα όχι Στεγανογραφικό υλικό. Τα περισσότερα από τα Στεγανογραφικά εργαλεία κειμένου είναι ελεύθερης χρήσης και διακίνησης εκτός από τα Steganos, Invisible Secrets και PGPn123.

Text Steganographic Tools	Plain Text	Other	Source Code	License	Production
PGPn123		Yes		Shareware	Yes
Nicetext	Yes		Yes	Open Source	Yes
Snow	Yes		Yes	Open Source	Yes
Texto	Yes		Yes	Open Source	Yes
Sam's Big Play Maker	Yes		Yes	Open Source	Yes
Steganosaurus	Yes		Yes	Open Source	Yes
FFEncode	Yes			Open Source	Yes
Mimic	Yes			Open Source	Yes
wbStego	Yes	HTML, PDF	Yes	Open Source	Yes
Spam Mimic	Yes			Not Specified	Yes
Secret Space	Yes			Not Specified	Yes
WitnesSoft	Yes	Yes		No longer in production	
MergeStreams		Hides excel file in word		Freeware	Yes
Steganos	Yes	HTML		Commercial	Yes
Invisible Secrets		HTML		Commercial	Yes

Εικόνα 3.4[33]

3.6.4 Στεγανογραφικά Εργαλεία Συστημάτων Αρχείων και Σκληρού Δίσκου

Στον τομέα Στεγανογραφίας που αφορά Συστήματα αρχείων και σκληρών δίσκων ερευνήσαμε ένα σύνολο δεκαεννέα (19) εργαλείων, τα οποία περιελάμβαναν προϊόντα ελεύθερης χρήσης, "open source" και προϊόντα διαμοιρασμού. Δεν βρήκαμε κάποια εμπορικά προϊόντα αναφορικά με αυτόν τον τομέα. Όλα τα εργαλεία αναγράφονται στην Εικόνα 3.5.
Η πλειοψηφία αυτών των εργαλείων ενσωμάτωσαν κρυφές πληροφορίες στο Σύστημα Αρχείων και το *Windows Registry*.
Εργαλεία όπως τα Magic Folders[82], Dark Files[83], bProtected 2000[84], BuryBury[85], StegFS[86], Folder Guard Jr[87], Dmagic[88] και BackYard χρησιμοποιούν το σύστημα αρχείων για ενσωμάτωση, ενώ τα Disk Hide και Drive Hider βασίστηκαν στην απόκρυψη μέσα στην *Windows Registry.*
Τα Easy File & Folder Protector και Anahtar είναι κατά κάποιο τρόπο μοναδικά καθώς, το πρώτο ενσωματώνει κρυφή πληροφορία σε driver VXD, ενώ το δεύτερο χρησιμοποιεί δισκέτα 3.5 ιντσών. Το Anahtar δεν βρίσκεται πια στην παραγωγή καθώς οι δισκέτες floppy είναι πια ξεπερασμένες, αλλά θα ήταν ενδιαφέρον να δούμε αν θα μπορούσαν στην θέση τους να χρησιμοποιηθούν μονάδες USB. Κανένα από αυτά τα προϊόντα δεν παρέχει πηγαίο κώδικα καθώς όλα είναι ελεύθερης διακίνησης η διαμοιρασμού αλλά κανένα δεν είναι "open source". Το Snow disk είναι επίσης ένα αξιόλογο πρόγραμμα, το οποίο ενσωματώνει κρυφές πληροφορίες στον χώρο του δίσκου, αν και δεν είναι πια στην παραγωγή.

File System Steganographic Tools	Location of Embedding	Source Code	License	Production
Disk Hide	Windows Registry	No		No
Drive Hider	Windows Registry	No		No
Easy File & Folder Protector	VXD driver. Windows Kernel	No	Shareware	Yes
Invisible Files 2000	Hard Disk	No	Shareware	Yes
Magic Folders	File System	No	Shareware	Yes
Dark Files	File system	No	Shareware	Yes
bProtected 2000	File system	No	Shareware	Yes
BuryBury	File system	No	Shareware	Yes
StegFS	File system	Yes	Open Source	Yes
Folder Guard Jr	File System	No	Freeware	Yes
Dmagic	File System	No	Freeware	Yes
BackYard	File System	No		No
Snowdisk	Disk space			No
Masker	Any file (Image, Text, Audio, Video)	No	Shareware	Yes
Anahtar	3.5-inch diskette	No		No
Hide Folders		No	Shareware	Yes
Hidden		No		No
Paranoid		No		No
Diskhide		No		No

Εικόνα 3.5[33]

3.6.5 Άλλα Ποικίλα Εργαλεία Στεγανογραφίας

Κατηγοριοποιήσαμε τα ακόλουθα εργαλεία σε ποικίλες κατηγορίες διότι είναι κατά κάποιο τρόπο μοναδικά Στεγανογραφικά εργαλεία. Το gzSteg ενσωμάτωσε κρυφές πληροφορίες σε αρχεία gz.
Τα S-Mail και Hydan[94] ενσωμάτωσαν πληροφορίες σε εκτελέσιμα (.exe) και Dynamic link libraries (dll files). Συγκεκριμένα, το Hydan ενσωμάτωσε σε binaries. Τα Hiderman[93], StegMark και InfoSteg ενσωμάτωσαν αρχεία εικόνων, ήχου και βίντεο.
Εν τέλει, το KPK File[92] ήταν κατά κάποιο τρόπο μοναδικό καθώς ενσωμάτωσε πληροφορίες σε αρχεία Word όπως επίσης και σε αρχεία BMP.

Miscellaneous Steganographic Tools	Cover Media	Source Code	License
GZSteg	.gz files	Yes	
InfoStego	Image, audio, video		Shareware
KPK File	Word, BMP		Shareware
S-Mail	.exe and .dll files		
Hiderman	Many different media		Shareware
StegMark	Image, audio, video		
Steghide	JPEG, BMP, WAV, AU	Yes	
S-Tools	BMP, GIF, WAV	Not sure	
Hydan	Program Binaries	Yes	Open Source
Covert.tcp	TCP/IP Packets	Yes	Open Source

Εικόνα 3.6[33]

3.6.6 Στεγαναλυτικά Εργαλεία

Σε αυτή την ενότητα θα εξετάσουμε κάποια από τα Στεγαναλυτικά Εργαλεία που ερευνήσαμε, όπως φαίνεται στην Εικόνα 3.7.

Hard Disk Steganographic Tools	Tools Analyzed	Detection Approach	Extraction Approach	Destruction Approach
2Mosaic	Removes stego content from any images			Break Apart
StirMark Benchmark	Removes stego content from any images			Resample
Phototile	Removes stego content from any images			Break Apart
Steganography Analyzer Real-Time Scanner	Analyzes Network Packets	Signature		
StegBreak	Jsteg-shell, JPhide, and Outguess 0.13b		Dictionary	
StegDetect	Jsteg, JPhide, Invisible Secrets, Outguess 01.3b, F5, appendX, Camouflage	Statistical		
StegSpy	Hiderman, JPHIde and Seek, Masker, JPegX, Invisible Secrets			
Stego-Suite	Detects Stego Image and Audio file		Dictionary	

Εικόνα 3.7[33]

Ερευνήθηκε ένα σύνολο οχτώ (8) Στεγαναλυτικών εργαλείων. Τα 2Mosaic[96], StirMark Benchmark[97] και PhotoTitle[98] είναι τα τρία Στεγαναλυτικά Εργαλεία τα οποία μπορούν να αφαιρέσουν Στεγανογραφικό υλικό από κάθε

εικόνα.
Αυτό επιτυγχάνεται καταστρέφοντας το κρυφό μήνυμα χρησιμοποιώντας δύο τεχνικές τις: **break apart** και **resample**. Τα StegDetect[101], StegBreak[100] και StegSpy[102] αναγνωρίζουν ενσωματωμένες πληροφορίες των ακολούθων εργαλείων:

Jsteg-shell[41], JPhide, and Outguess 0.13b, Invisible Secrets, F5, appendX, Camouflage, Hiderman, JPHIde and Seek[40], Masker, JPegX[59].

Τέλος το Steganography Analyzer Real-Time Scanner[99] είναι το καλύτερο λογισμικό Στεγανάλυσης Δικτύου που είναι διαθέσιμο στη αγορά αυτή τη στιγμή.Μπορεί να αναλύσει όλη την δικτυακή κίνηση με σκοπό την ανίχνευση Στεγανογραφικής επικοινωνίας.

4.1 Ο νόμος του Benford και η επαλήθευση αυτού.

4.1.1 Ιστορικά Στοιχεία

Το 1881 ο Newcomb[3][4] συνέβαλε στην ανακάλυψη της στατιστικής αρχής που είναι σήμερα γνωστή ως «Νόμος του Benford», όταν είδε ότι οι πρώτες σελίδες των πινάκων λογαρίθμων, που χρησιμοποιούνταν τότε για την εκτέλεση υπολογισμών, ήταν πολύ πιο φθαρμένες από τη χρήση από ότι οι τελευταίες. Αυτό τον οδήγησε στο να διατυπώσει την αρχή ότι σε οποιονδήποτε κατάλογο αριθμών λαμβανόμενων από ένα τυχαίο σύνολο δεδομένων, περισσότεροι αριθμοί αρχίζουν από το ψηφίο ένα (1) παρά από οποιοδήποτε άλλο.

Πιο πρόσφατα, η συγκεκριμένη συμπεριφορά παρατηρήθηκε και τυποποιήθηκε από τον Benford το 1938, για περισσότερες μορφές τυχαίων συνόλων δεδομένων. Προς τιμήν της εργασίας του ο Νόμος που διατυπώθηκε, ονομάστηκε Νόμος του Benford. Μια ικανοποιητική εξήγηση του Νόμου αυτού δόθηκε πρόσφατα μέσα από την δουλειά του Hill[6] το 1996. Το συγκεκριμένο φαινόμενο είναι σήμερα πολύ διαδεδομένο καθώς μπορεί να χρησιμοποιηθεί σε πολλές εφαρμογές με τεχνητά ή ακόμη και φυσικά τυχαία σύνολα αριθμών τα οποία και το επαληθεύουν.

4.1.2 Ο Νόμος και η επαλήθευση

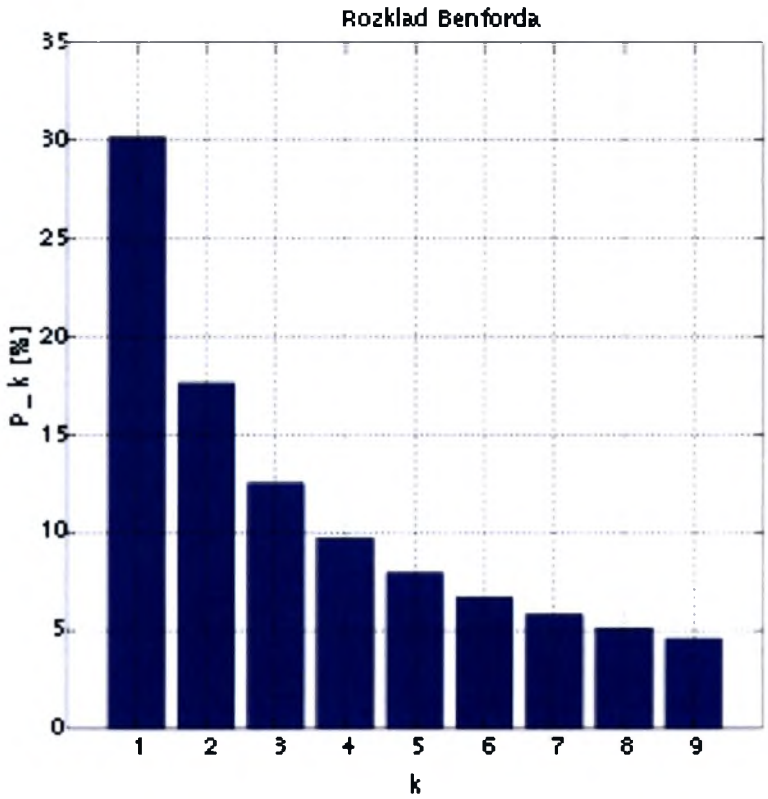
Για να γίνουμε πιο ακριβείς ο Νόμος του Benford δηλώνει ότι το σημαντικότερο ψηφίο d (d Ε {1, …, b − 1}) ενός αριθμού με βάση b ($b \geq 2$) προκύπτει με πιθανότητα που υπολογίζεται ως:

$$P(d) = \log_b(d+1) - \log_b(d) = \log_b\left(1 + \frac{1}{d}\right).$$

Η ποσότητα αυτή αποτελεί το χώρο μεταξύ των αριθμών d και $d + 1$ στην λογαριθμική κλίμακα.

Με βάση το 10 τώρα, τα πρώτα ψηφία έχουν την παρακάτω κατανομή σύμφωνα με τον Νόμο του Benford, όπου d είναι το πρώτο ψηφίο και p η πιθανότητα εμφάνισης του:

d	1	2	3	4	5	6	7	8	9
p	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%



Εικόνα 4.1: Γραφική απεικόνιση του Νόμου του Benford για το πρώτο ψηφίο. [115]

Η επαλήθευση του Νόμου

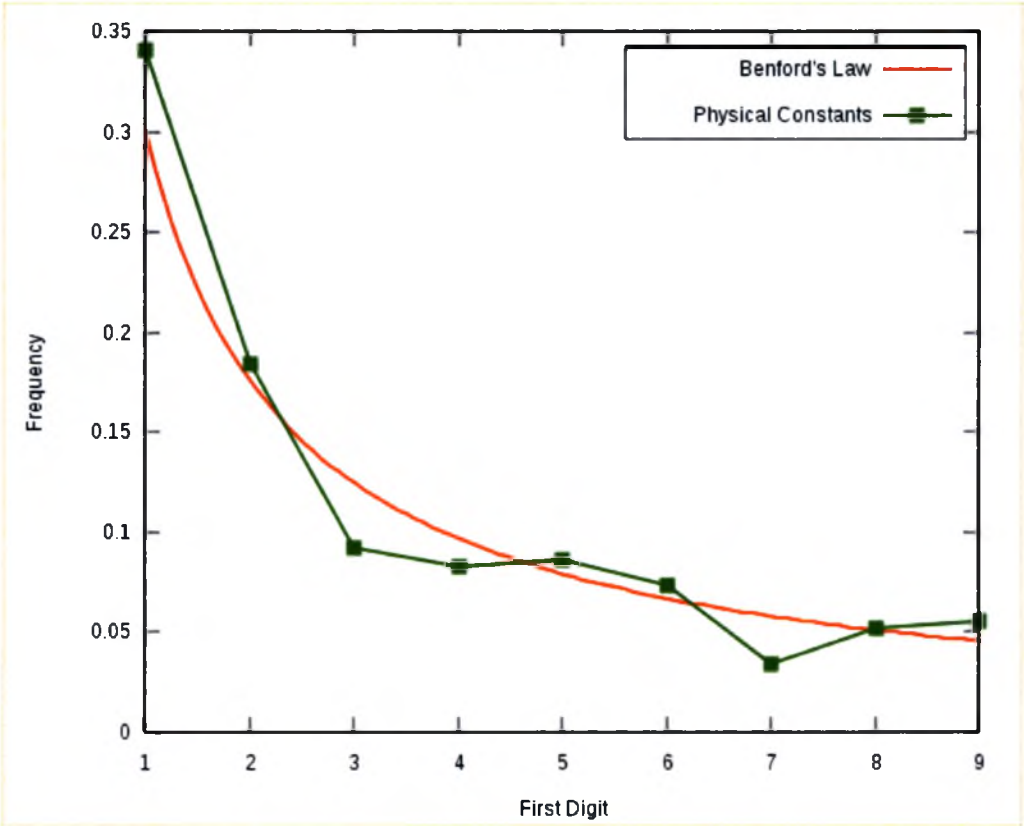
Ο νόμος του Benford μπορεί να αποδειχθεί με διαφορετικές προσεγγίσεις μερικές από τις οποίες παρουσιάζονται παρακάτω.

Αποτέλεσμα διαδικασίας εκθετικής αύξησης.

Ο ακριβής τύπος του Νόμου του Benford μπορεί να αποδειχθεί εάν υποθέσουμε ότι οι λογαριθμικές τιμές των αριθμών κατανέμονται ομοιόμορφα. Για παράδειγμα[115] είναι γνωστό πως η τιμή ενός αριθμού είναι το ίδιο πιθανό να βρίσκεται μεταξύ του 100 και του 1000 (λογάριθμοι μεταξύ του 2 και 3) και μεταξύ του 10.000 και 100.000 (λογάριθμοι μεταξύ του 4 και 5).Για πολλά σύνολα αριθμών, ειδικά για αυτά που αυξάνονται εκθετικά, όπως εισοδήματα και τιμές μετοχών, το παραπάνω αποτελεί μια λογική υπόθεση.

Σαν παράδειγμα, εάν μία ποσότητα διπλασιάζεται κάθε χρόνο τότε θα είναι διπλάσια σε ένα χρόνο, τετραπλάσια σε δύο χρόνια και οχταπλάσια σε τρία. Όταν η τιμή φτάσει το 100 θα έχει σαν πρώτο ψηφίο το ένα (1) για ένα χρόνο, τον επόμενο χρόνο θα φτάσει από 200 σε 400 και θα έχει πρώτο ψηφίο το δύο (2) για λίγο παραπάνω από 7 μήνες ενώ θα έχει πρώτο ψηφίο το τρία (3) για τους υπόλοιπους 5 μήνες .Τον τρίτο χρόνο το πρώτο ψηφίο θα αλλάξει τιμές μεταξύ 4,5,6 και 7 για όλο και λιγότερο διάστημα ανά ψηφίο και στην αρχή του 4 χρόνου η τιμή του πρώτου ψηφίου θα αλλάξει από 8 σε 9. Όταν η τιμή θα έχει φτάσει 1000 η διαδικασία θα ξεκινήσει από την αρχή με πρώτο ψηφίο πάλι το ένα (1).

Ο νόμος του Benford λοιπόν ισχύει ξεκάθαρα για τιμές που αυξάνονται εκθετικά, αλλά υπάρχουν παραδείγματα για τα οποία ο νόμος ισχύει χωρίς να υπάρχει εμφανής εκθετική αύξηση.

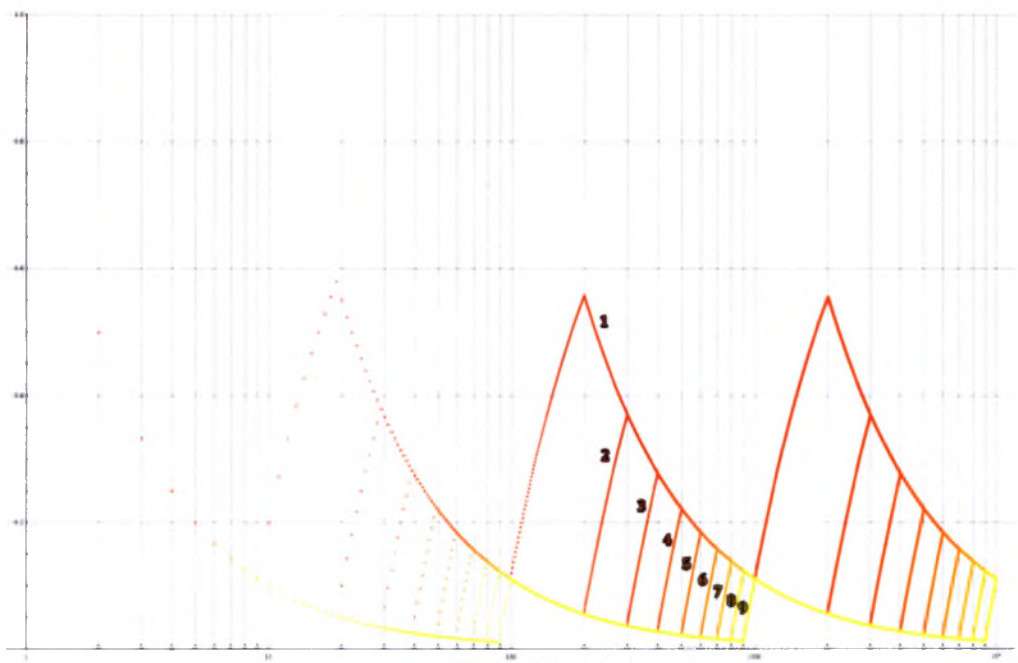


Εικόνα 4.2: Συχνότητα Φυσικών σταθερών σε αντιπαραβολή με τον Νόμο του Benford. [115]

Ανεξαρτησία μέτρησης

Ο νόμος μπορεί να αποδειχθεί εναλλακτικά από το γεγονός ότι εάν όντως το πρώτο ψηφίο κάθε αριθμού έχει μία συγκεκριμένη κατανομή συχνότητας εμφάνισης τότε αυτή θα πρέπει είναι ανεξάρτητη της μονάδας μέτρησης που χρησιμοποιείται.

Για παράδειγμα εάν μετατρέψουμε ένα σύνολο αριθμών από πόδια σε γιάρδες (πολλαπλασιασμός με σταθερά) η κατανομή των πρώτων ψηφίων δεν πρέπει να αλλοιωθεί. Η μόνη κατανομή για την οποία θα ισχύει αυτή η υπόθεση είναι αυτή για την οποία ο λογάριθμος της ήταν ομοιόμορφα κατανεμημένος.



Εικόνα 4.3[115]

Για κάθε θετικό ακέραιο n, το παραπάνω γράφημα δείχνει την πιθανότητα ένας τυχαίος ακέραιος μεταξύ του 1 και n να ξεκινά με ένα από τα πιθανά εννιά ψηφία. Για κάθε τιμή του n, η πιθανότητα δεν ικανοποιεί ακριβώς τον Νόμο του Benford, αλλά εξετάζοντας την ποικιλία των τιμών που μπορεί να πάει το n και βρίσκοντας τον μέσο όρο για κάθε μία από αυτές τις πιθανότητες οδηγούμαστε σε αποτελέσματα ίδια με αυτά του Νόμου του Benford.

Για παράδειγμα, το πρώτο (μη μηδενικό) ψηφίο των μηκών και αποστάσεων μεταξύ αντικειμένων θα πρέπει να έχει την ίδια κατανομή ανεξαρτήτως του αν μετράται σε πόδια, γιάρδες ή οτιδήποτε άλλο.

Αλλά υπάρχουν 3 πόδια σε μία γιάρδα και η πιθανότητα το πρώτο ψηφίο ενός μήκους σε γιάρδες να είναι 1 πρέπει να είναι ίδια με αυτή του να είναι το πρώτο ψηφίο 3,4,5 σε πόδια. Εφαρμόζοντας τα παραπάνω σε όλες τις δυνατές κλίμακες μέτρησης μήκους και σε συνδυασμό με το γεγονός ότι $\log_{10}(1) = 0$ και $\log_{10}(10) = 1$ οδηγούμαστε στον νόμο του Benford.

Πολλαπλές Κατανομές Πιθανότητας

Εάν επιλέξουμε τυχαία ένα σύνολο αριθμών με δική του πιθανότητα κατανομής και μετά τυχαία διαλέξουμε μια τιμή από αυτό το σύνολο και έπειτα επαναλάβουμε την διαδικασία αυτή όσες φορές θέλουμε, το σύνολο αριθμών που θα παραχθεί θα υπακούει στον Νόμο του Benford.

Από την παραπάνω διαπίστωση καταλαβαίνει κανείς πως σε πραγματικά τυχαία σύνολα αριθμών ο Νόμος του Benford επαληθεύεται.

4.2 Χρήσεις του Νόμου του Benford μέχρι σήμερα σε διάφορους τομείς.

1. Το 1972, ο Hal Varian πρότεινε την χρήση του Νόμου του Benford για τον εντοπισμό πιθανής απάτης σε κοινωνικό-οικονομικά δεδομένα που κατατέθηκαν για να παρθούν αποφάσεις που αφορούσαν την δημόσια οργάνωση. Βασισμένος στην λογική υπόθεση ότι οι άνθρωποι που θα άλλαζαν τις τιμές για να επιτύχουν την απάτη δεν θα χρησιμοποιούσαν πραγματικά τυχαίες τιμές, οδηγήθηκε στο συμπέρασμα ότι μια απλή σύγκριση με τα αποτελέσματα του Νόμου του Benford θα καταδείκνυε την απάτη.
2. Ακολουθώντας την παραπάνω ιδέα, ο Mark Nigrini[8] έδειξε ότι ο Νόμος του Benford μπορεί να χρησιμοποιηθεί για τον εντοπισμό οικονομικής απάτης.
3. Στοιχεία βασισμένα στον Νόμο του Benford μπορούν νομικά να κατατεθούν σε ομοσπονδιακό, πολιτειακό και τοπικό επίπεδο στις Ηνωμένες Πολιτείες.
4. Ο Νόμος του Benford χρησιμοποιήθηκε επίσης για να αποδείξει την νοθεία των Ιρανικών εκλογών[115] του 2009.

Με βάση τις παραπάνω χρήσεις του Νόμου του Benford, θα υποθέσουμε ότι ένα οποιοδήποτε αρχείο αποτελούμενο από ένα σύνολο bytes μπορεί δυνητικά να ακολουθεί τον Νόμο του Benford κάτι που μπορεί να διαταραχθεί εάν κάποιος προσπαθήσει να εισάγει σε αυτό κρυφή πληροφορία.

Περιορισμοί

Μεγάλη προσοχή πρέπει να δοθεί στις εφαρμογές για τις οποίες ισχύει ο Νόμος του Benford. Μπορεί ο νόμος να ισχύει για ένα σύνολο φυσικών αριθμών, δεν ισχύει όμως και για ένα περιορισμένο υποσύνολο του.

Για παράδειγμα, ο πληθυσμός των πόλεων της Ελλάδας το όνομα των οποίων ξεκινά με 'Α' μπορεί να ακολουθεί τον Νόμο του Benford, δεν ισχύει όμως το ίδιο εάν περιορίσουμε το δείγμα στις πόλεις με πληθυσμό μεταξύ 3000 και 9999 ανθρώπων.

4.3 Χρήσεις του Νόμου του Benford για Στεγανάλυση μέχρι σήμερα.

Ενδιαφέρουσα σχετικά με τους σκοπούς μας είναι η εργασία του J.M. Jolion, ο οποίος έδειξε ότι ο νόμος του Benford λειτουργεί αρκετά καλά σε απεικονίσεις υπό κλίση και σε πυραμιδικές διασπάσεις βασισμένες στη μετατροπή Laplace. Από όσο γνωρίζουμε η μοναδική άλλη εργασία που σχετίζεται με το νόμο του Benford είναι των E.Acebo και M. Sbert[30] οι οποίοι πρότειναν την χρήση του νόμου του Benford για να καθοριστεί εάν συνθετικές απεικονίσεις δημιουργούνταν χρησιμοποιώντας φυσικές μεθόδους, αν και το γεγονός ότι πολλές πραγματικές απεικονίσεις δεν ακολουθούν το νόμο του Benford κάνει αυτή την θεωρία αμφισβητήσιμη. Ακόμη στην εργασία των Dongdong Fu, Yun Q. Shi και Wei Sub[30] βλέπουμε ότι ενώ οι απεικονίσεις στο pixel domain δεν

δείχνουν να υπακούν στο νόμο Benford, η κατάσταση αλλάζει δραματικά όταν αυτές μετατρέπονται χρησιμοποιώντας το Discrete Cosine Transform (DCT). Παρουσιάζεται δε μια γενίκευση του νόμου του Benford βασισμένη στην ανάλυση Fourier που οδηγεί σε μια καλύτερη προσαρμογή των συχνοτήτων των στοιχείων που έχουμε παρατηρήσει. Ακόμη δίνεται μια θεωρητική εξήγηση του γιατί οι απεικονίσεις στο DCT domain ικανοποιούν το γενικευμένο νόμο. Αυτή η εξήγηση βασίζεται σε γνωστά και ελεγμένα στατιστικά στοιχεία των συντελεστών DCT. Η εργασία καταλήγει σε μία προσπάθεια χρήσης της γενίκευσης στον εντοπισμό υδατογραφημένων εικόνων με εφαρμογή στον τομέα του Computer Forensics.

Παρατηρήθηκε ότι:

- Σε κανένα από τα παραπάνω παραδείγματα χρήσης του Νόμου του Benford, δεν οδηγούμαστε σε ένα ολοκληρωμένο εργαλείο εντοπισμού κρυμμένης πληροφορίας, αλλά υπάρχουν μόνο θεωρητικές υποδείξεις της χρήσης του γι αυτό τον σκοπό.

- Σε καμία από τις θεωρητικές προτάσεις χρήσης του Νόμου του Benford, ο εντοπισμός κρυφής πληροφορίας δεν επεκτείνεται σε παραπάνω από ένα τύπο αρχείου.

Σε αυτήν την εργασία παρουσιάζεται λοιπόν μια ξεχωριστή προσέγγιση της χρήσης του Νόμου του Benford για εντοπισμό κρυφού περιεχομένου εφαρμόζοντας την γενίκευση του νόμου στο byte stream του αρχείου.
Με την μέθοδο αυτή και σε συνδυασμό με την αναδόμηση αρχείων δίνεται η δυνατότητα σύγκρισης των αποτελεσμάτων της γενίκευσης του νόμου με πολύ ικανοποιητικά αποτελέσματα για εφαρμογές εντοπισμού Στεγανογραφίας σε διαφορετικούς τύπους αρχείων μέσα από πραγματικά εργαλεία Στεγανάλυσης που αναπτύχθηκαν στα πλαίσια της διπλωματικής.

ΜΕΡΟΣ ΙΙ - ΠΡΟΤΕΙΝΟΜΕΝΗ ΜΕΘΟΔΟΛΟΓΙΑ ΔΙΕΡΕΥΝΗΣΗΣ

5.1 Στεγανογραφία και αλλοίωση του file structure

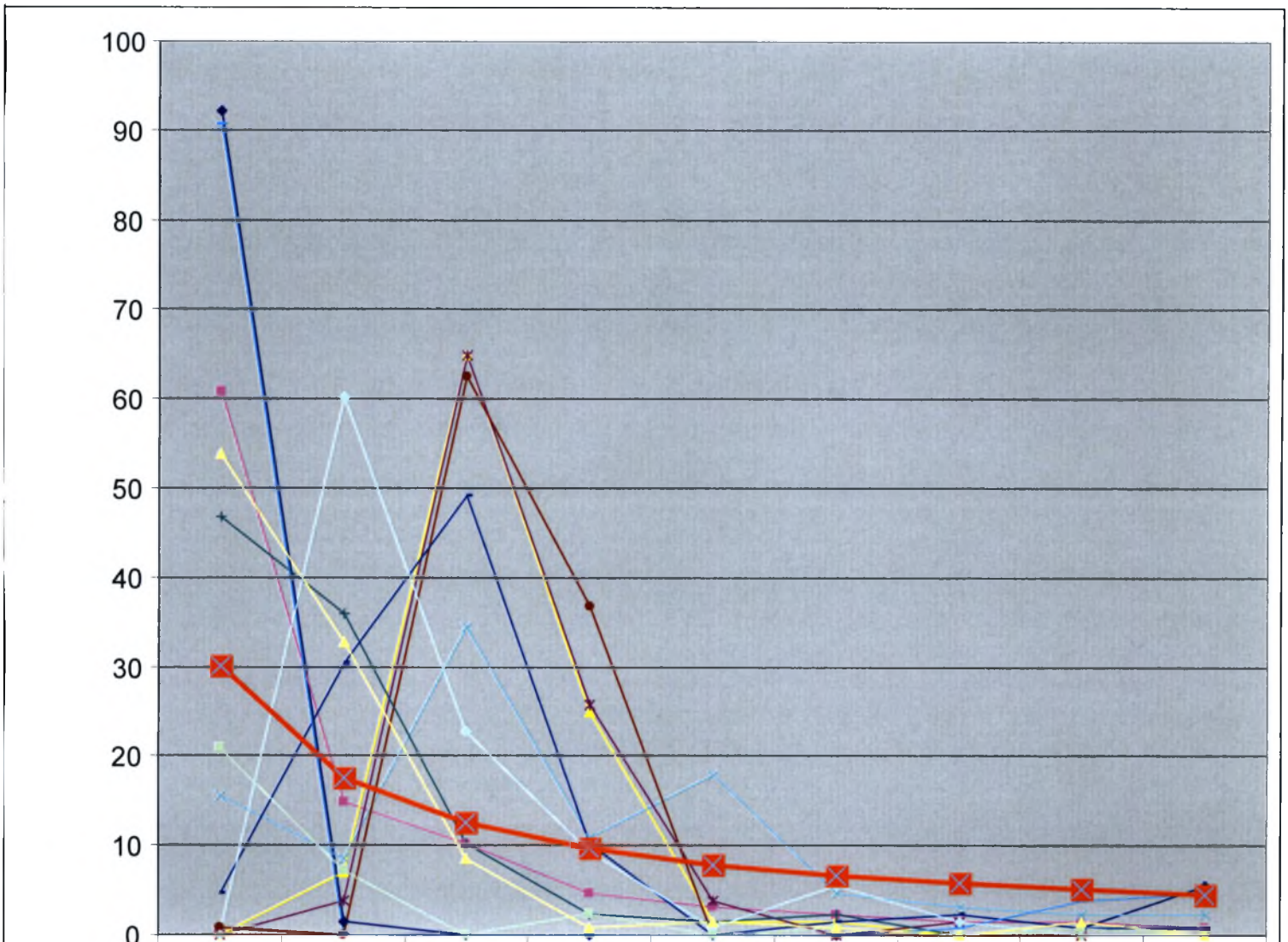
Ύστερα από εκτεταμένο πειραματισμό με την χρήση εργαλείων Στεγανογραφίας για διαφορετικούς τύπους αρχείων, παρατηρήθηκε αλλοίωση της δομής του byte array, σε σχέση με αυτό των αρχικά μη Στεγανογραφημένων αρχείων.
Η αλλοίωση / διαφορά του byte array παρατηρήθηκε ότι:

1. Ήταν μετρήσιμη για κρυφά αρχεία μικρού μεγέθους.
2. Αυξανόταν όσο αυξανόταν το μέγεθος του κρυφού αρχείου.
3. Ήταν εμφανής για όλους τους τύπους αρχείων που Στεγανογραφήθηκαν ανεξαρτήτως εργαλείου Στεγανογραφίας ή αλγορίθμου Στεγανογραφίας που χρησιμοποιήθηκε.

Παράλληλα με την παραπάνω παρατήρηση, η απεικόνιση ύστερα από επεξεργασία του σημαντικότερου ψηφίου του συνόλου των τιμών του byte array για κάποιους τύπους αρχείων φάνηκε να ακολουθεί με μεγάλη προσέγγιση των Νόμο του Benford.
Σε κάποιες περιπτώσεις τύπου αρχείων (πχ. .WAV) η επαλήθευση του Νόμου του Benford ήταν ξεκάθαρη, ενώ για άλλους τύπους αρχείων (.BMP), ήταν μερική.
Συνολικά αναλύθηκαν πολλοί διαφορετικοί τύποι αρχείων με έμφαση σε αρχεία που υπόκεινται πιο συχνά σε Στεγανογραφία ή χρησιμοποιούνται για διακίνηση κρυφής πληροφορίας.

Στην Εικόνα 5.1 παρουσιάζεται ένα γενικό γράφημα για τους δώδεκα (12) πιο δημοφιλείς τύπους αρχείων που χρησιμοποιούνται για απόκρυψη πληροφορίας και διακινούνται ελεύθερα στο Διαδίκτυο. Το γράφημα αυτό απεικονίζει την κατανομή Benford για το byte array representation του κάθε τύπου αρχείου.

Για την συγκομιδή των αποτελεσμάτων που χρησιμοποιήθηκαν στο παρακάτω γράφημα, εφαρμόστηκε ο νόμος του Benford σε δέκα (10) τυχαίου μεγέθους αρχεία για κάθε τύπο αρχείου. Έπειτα υπολογίστηκε ο μέσος όρος για κάθε τύπο αρχείου ο οποίος και παρουσιάζεται στο παρακάτω γράφημα σε αντιπαραβολή με αναμενόμενα αποτελέσματα του Νόμου του Benford (κόκκινη γραμμή).

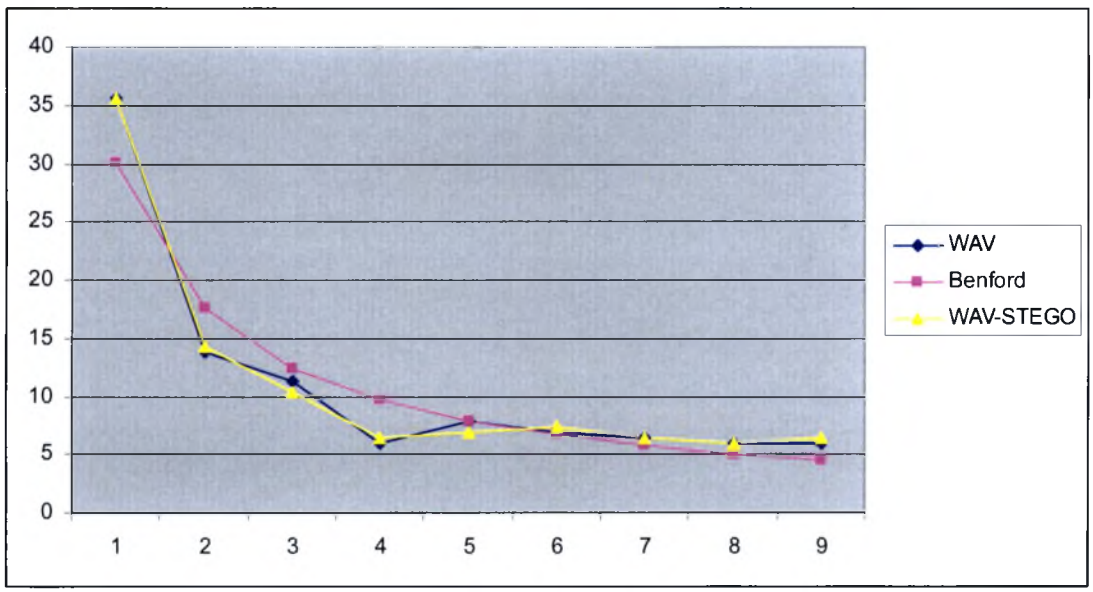


	1	2	3	4	5	6	7	8	9
WMV	92,19	1,56	0	0	0	0	0	0,78	5,47
WAV	61	15	10	5	3	2	2	2	1
PNG	0	7	65	25	2	2	0	0	0
BMP	16	9	34	11	18	5	3	2	2
MP3	0	4	65	26	4	0	2	0	0
ZIP	1	0	63	37	0	0	0	0	0
EXE	47	36	10	2	2	2	0	1	0
PDF	5	30	49	10	0	2	2	1	1
JPEG	91	0	0	0	0	0	1	4	5
GIF	0	60	23	9	1	5	2	0	0
RAR	21	7	0	2	0	1	0	1	0
DOC	54	33	9	1	2	1	0	2	0
Benford	30	18	12	10	8	7	6	5	5

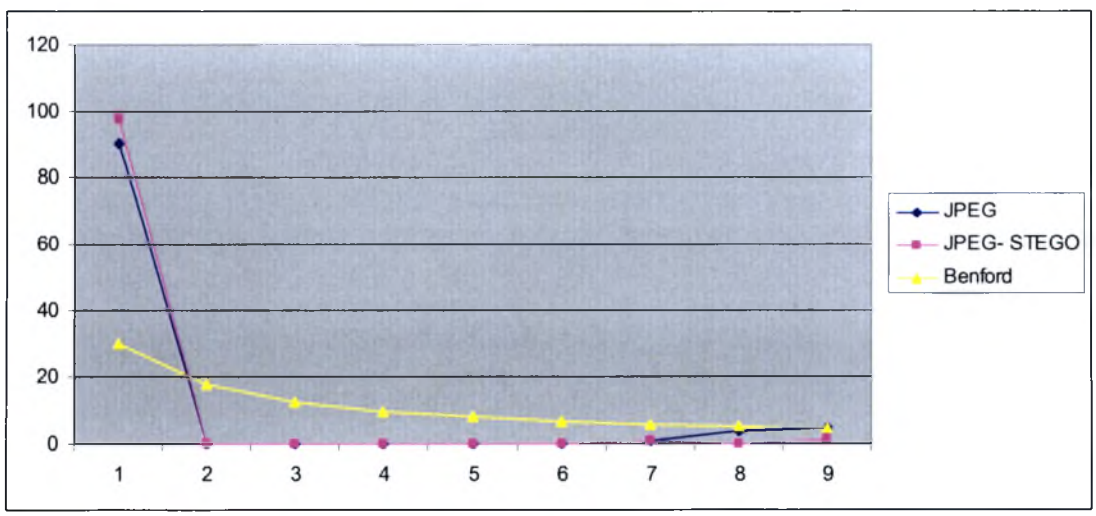
Εικόνα 5.1

5.2 Στατιστικά αλλοιώσεων για διαφορετικούς τύπους αρχείων

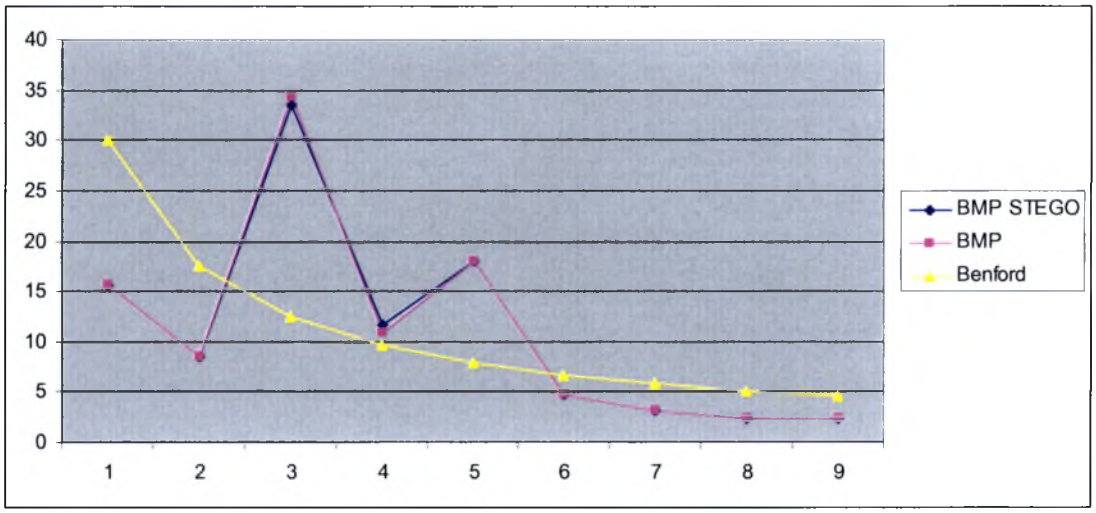
Είναι φανερό πως το επόμενο λογικό βήμα, για την εξακρίβωση του εάν η Στεγανογραφία και γενικά η απόκρυψη πληροφορίας αλλοιώνουν την δομή ενός αρχείου, ήταν η αντιπαραβολή της κατανομής Benford του αρχικού και του στεγανογραφημένου αρχείου. Παρακάτω παρουσιάζονται τέσσερα δείγματα αρχείων σε αντιπαραβολή με τα στεγανογραφημένα τους που περιέχουν την ελάχιστη δυνατή κρυφή πληροφορία (<=1kb). Και στα τέσσερα παραδείγματα εμφανίζονται διαφορές μικρού μεγέθους που όμως μπορούν να εκτιμηθούν. Σκοπός είναι αυτή η πληροφορία να χρησιμοποιηθεί κατάλληλα ώστε να αποτελέσει μέσω εντοπισμού στεγανογραφίας μέσα από ένα εργαλείο forensics.



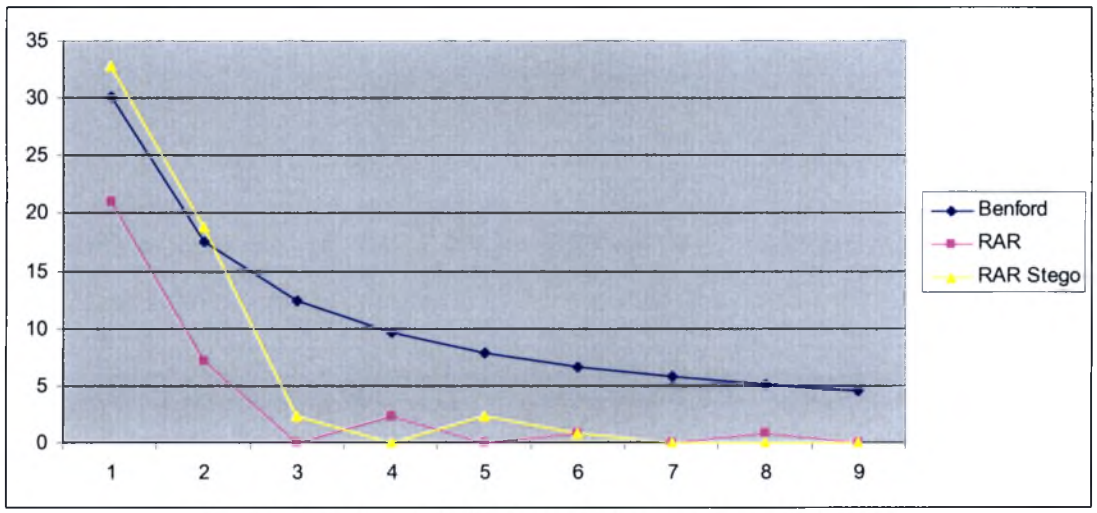
Εικόνα 5.2: Αρχείο WAV



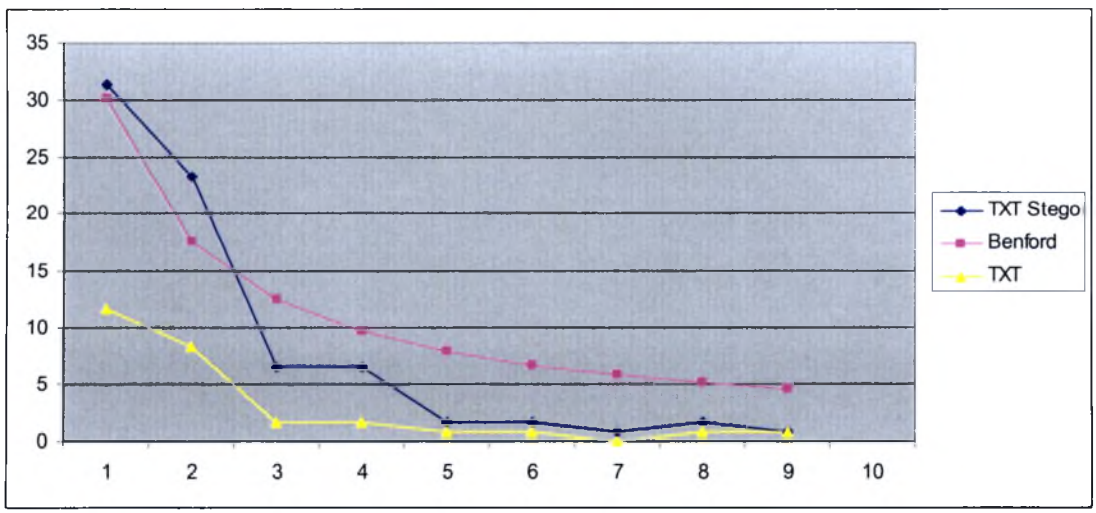
Εικόνα 5.3: Αρχείο εικόνας JPEG



Εικόνα 5.4:Αρχείο εικόνας BMP



Εικόνα 5.5: Αρχείο RAR (κρυφή πληροφορία με C.A.D.S.)



Εικόνα 5.6: Αρχείο κειμένου TXT

5.3 Συμπεράσματα που οδηγούν στην Γενίκευση του Νόμου του Benford

Σε όλους τους τύπους αρχείων που χρησιμοποιήθηκαν για Κρυπτογραφία χρησιμοποιώντας διαφορετικά εργαλεία Στεγανογραφίας, παρουσιάστηκε αλλοίωση της δομής τους σε σχέση με το αρχικό αρχείο σε ότι αφορά τις τιμές των πρώτων ψηφίων των bytes που τα αναπαριστούν.
Σε κάποια είδη αρχείων η γραφική αναπαράσταση των πρώτων ψηφίων των bytes που αποτελούν το σύνολο των αρχείων ακολουθούσε πιστά την κατανομή Benford ενώ σε άλλα αρχεία όχι.
Η χρήση στεγανογραφίας αλλοίωνε σε όλες τις περιπτώσεις την κατανομή Benford είτε επαληθεύοντας τις αναμενόμενες τιμές είτε όχι.
Εντοπίζοντας λοιπόν την αλλοίωση στα bytes που αντιπροσωπεύουν ένα αρχείο ύστερα από χρήση στεγανογραφίας ανεξαρτήτως τύπου αρχείου (file type) αποφασίσαμε να εφαρμόσουμε τον Νόμο του Benford στα bytes του αρχικού αρχείου αλλά και αυτού που περιέχει κρυμμένη (Στεγανογραφημένη πληροφορία).
Οι διαφορές μας οδήγησαν στην απόφαση να παράγουμε μία μεθοδολογία ασφαλούς εντοπισμού στεγανογραφημένων αρχείων με βάση την διαφορά πού είχαν οι στατιστικές αναπαραστάσεις του νόμου του Benford για το πρωτότυπο αρχείο σε σχέση με το στεγανογραφημένο. Για να επιτευχθεί το αναμενόμενο αποτέλεσμα φυσικά έπρεπε να λυθούν κάποια προβλήματα όπως:

1. Στις περισσότερες περιπτώσεις στεγανάλυσης δεν υπάρχει το αρχικό αρχείο για να γίνει η σύγκριση με το στεγανογραφημένο.
2. Η κατανομή του Benford δεν ισχύει για όλα τα είδη αρχείων.

Για την επιτυχία του σκοπού της εργασίας αυτής τέθηκαν οι παρακάτω στόχοι :

1. Δημιουργία μεθοδολογίας και υλοποίησης χειροκίνητης (manual) αναδόμησης για έναν τουλάχιστον τύπο αρχείου. (.TXT)
2. Δημιουργία μεθοδολογίας και υλοποίησης αυτοματοποιημένης αναδόμησης αρχείου για τουλάχιστον ένα τύπο αρχείου. (.JPEG)
3. Γενίκευση του νόμου του Benford ώστε να λειτουργήσει σαν μέτρο σύγκρισης (metric) παρόλο που δεν ικανοποιείται για όλα τα είδη αρχείων.

Λαμβάνοντας υπόψη την ύπαρξη σοβαρών αλλοιώσεων στην δομή των αρχείων που υπόκεινται Στεγανογραφία, οδηγηθήκαμε στην ανάγκη να δημιουργήσουμε μία πρωτότυπη μεθοδολογία προκειμένου να εντοπίσουμε, χωρίς την ύπαρξη αρχικού αρχείου, πιθανά στεγανογραφημένα αρχεία. Βασικό εμπόδιο υπήρξε η αναδόμηση του αρχικού αρχείου από το στεγανογραφημένο. Στο παρόν κεφάλαιο θα περιγραφεί η θεωρητική προσέγγιση της αναδόμησης αρχείων που θα επεξηγηθεί μέσα από την αναδόμηση αρχείων κειμένου (.txt), ενώ σε επόμενο κεφάλαιο θα παρουσιαστεί μία ρεαλιστική προσέγγιση τεχνικής αυτοματοποιημένης αναδόμησης αρχείων εικόνας τύπου JPEG.

6.1 Γενίκευση του Νόμου του Benford για εντοπισμό Κρυμμένου Περιεχομένου σε Forensics Investigation

Σε αυτήν την ενότητα περιγράφεται η προτεινόμενη μεθοδολογία για τον εντοπισμό Κρυμμένου Περιεχομένου με χρήση της Γενίκευσης του Νόμου του Benford. Η μεθοδολογία αυτή δεν έχει χρησιμοποιηθεί ξανά και θεωρούμε ότι μπορεί να εξελιχθεί σε μία αξιόπιστη μέθοδο εντοπισμού κρυμμένου περιεχομένου ανεξαρτήτως τύπου αρχείου.

Δεδομένου ενός Υπόπτου Αρχείου *'F':*

1. Υπολογίζουμε την αλληλουχία byte array που αποτελεί το Ύποπτο Αρχείο *F[file.length()]*.
2. Υπολογίζουμε την πιθανότητα *x=1,2,…,9, για* το πρώτο και το τελευταίο ψηφίο της αλληλουχίας byte array *F[file.length()]* για το Ύποπτο Αρχείο.
3. Αναδομούμε το Αρχικό Αρχείο 'O' από το Ύποπτο.
4. Υπολογίζουμε την αλληλουχία byte array που αποτελεί το Αναδομημένο Αρχείο *O[file.length()]*.
5. Υπολογίζουμε την πιθανότητα *x=1,2,…,9,για* το πρώτο και το τελευταίο ψηφίο της αλληλουχίας byte array *O[file.length()]* για το Αναδομημένο Αρχείο.
6. Συγκρίνουμε τα αποτελέσματα των βημάτων 2 και 5 και χρησιμοποιώντας ένα προκαθορισμένο κατώφλι αποφασίζουμε εάν το ύποπτο αρχείο αποτελεί στεγανογραφημένο αρχείο (carrier file) ή όχι.

Η διαδικασία καθορισμού του κατωφλίου που χρησιμοποιείται για την σύγκριση των δύο αρχείων θα παρουσιαστεί με λεπτομέρεια σε επόμενη ενότητα.

Γιατί να χρησιμοποιήσουμε την Γενίκευσης του Νόμου του Benford;

Η αξία της Γενίκευσης του Νόμου του Benford για τον εντοπισμό κρυφού περιεχομένου μπορεί εύκολα να προσδιοριστεί.

Η Γενίκευσης του Νόμου του Benford, αποτελεί ταυτόχρονα :

1. Μία αρκετά γενική μέθοδο στον τρόπο εφαρμογής της, ώστε να αγνοεί μικρές διαφορές στην αλληλουχία byte array του Αναδομημένου Αρχικού σε σχέση με το Ύποπτο Αρχείο, που προήλθαν από την αναδόμηση.

2. Μια αρκετά ακριβή μέθοδο για τον εντοπισμό σημαντικών αλλαγών στην αλληλουχία byte array του Αναδομημένου Αρχικού σε σχέση με το Ύποπτο αρχείο που προήλθαν από την εφαρμογή Στεγανογραφίας.

Τα παραπάνω συμπεράσματα προήλθαν από συγκριτικά αποτελέσματα τιμών της Γενίκευσης του Νόμου του Benford σε :

1. Απλά αρχεία καθ' αυτά.
2. Αρχεία που προήλθαν από αναδόμηση απλών αρχείων.
3. Αρχεία που προήλθαν από Στεγανογραφία σε απλά αρχεία.
4. Αρχεία που προήλθαν από Αναδόμηση αρχείων που προήλθαν από Στεγανογραφία.

Σύγκριση 1

Για να γίνουμε πιο ακριβείς όταν μετρήθηκαν οι τιμές αλληλουχίας byte array απλών αρχείων (1) και αρχείων που προήλθαν από αναδόμηση απλών αρχείων (2) παρατηρήθηκε ότι οι τιμές ήταν ίδιες στα σημαντικότερα ψηφία τους σε πολύ μεγάλο ποσοστό ~96% σε αντίθεση με τα λιγότερο σημαντικά ψηφία τους τα οποία έμοιαζαν σε ένα ποσοστό της τάξεως του 20-30%.

Σύγκριση 2

Τώρα όταν η ίδια σύγκριση έγινε για Απλά αρχεία (1) σε σχέση με Αρχεία που προήλθαν από Στεγανογραφία (3) τα ποσοστά άλλαξαν δραματικά καθώς παρατηρήθηκε ότι οι τιμές στα σημαντικότερα ψηφία τους ήταν ίδιες κατά ~70% ενώ τα λιγότερο σημαντικά ψηφία τους έμοιαζαν σε ένα ποσοστό της τάξεως του 10%.

Σύγκριση 3

Σε σύγκριση τώρα Αρχείων που προήλθαν από Στεγανογραφία (3) με αρχεία προήλθαν από αναδόμηση αρχείων που προήλθαν από Στεγανογραφία (4). Παρατηρήθηκε ότι οι τιμές στα σημαντικότερα ψηφία τους ήταν ίδιες κατά ~60% ενώ τα λιγότερο σημαντικά ψηφία τους έμοιαζαν σε ένα ποσοστό της τάξεως του 10%.

Σύγκριση 4

Τέλος στην σημαντικότερη ίσως σύγκριση για την πρώτη επαλήθευση της χρησιμότητας της Γενίκευσης του Νόμου του Benford στον εντοπισμό κρυφού περιεχομένου, συγκρίθηκαν Αρχεία που προήλθαν από αναδόμηση αρχείων που προήλθαν από Στεγανογραφία (4) με απλά αρχεία (τα αρχικά τους)(1).

Παρατηρήθηκε λοιπόν ότι οι τιμές ήταν ίδιες στα σημαντικότερα ψηφία τους σε πολύ μεγάλο ποσοστό ~80% σε αντίθεση με τα λιγότερο σημαντικά ψηφία τους τα οποία έμοιαζαν σε ένα ποσοστό της τάξεως του 10-20%.

Από τα παραπάνω μπορούμε να φτάσουμε στο συμπέρασμα ότι :

- Η επιτυχημένη αναδόμηση αρχείων οδηγεί σε σχεδόν όμοια αρχεία για την Σύγκριση 1 όπου δεν υπάρχει Στεγανογραφία πράγμα αναμενόμενο. Η διαδικασία βέβαια δεν μπορεί να παράγει πανομοιότυπα αρχεία με αποτέλεσμα να υπάρχουν μικρές διαφορές οι οποίες αργότερα πρέπει να διακριθούν προκειμένου να μην συσχετιστούν λανθασμένα με ύπαρξη Στεγανογραφίας.
- Η Σύγκριση 3 και Σύγκριση 4 οδηγούν στο συμπέρασμα ότι η αναδόμηση αρχείων αγνοεί / εξαλείφει την Στεγανογραφία οδηγώντας σε αρχεία κοντά στο αρχικό. Αυτό μπορεί να εξηγηθεί καθώς τα περισσότερα εργαλεία στεγανογραφίας προσθέτουν πληροφορία η οποία δεν ταυτίζεται με την ορθή δομή του κλασσικού προτύπου των τύπων αρχείων και άρα στην διαδικασία επαναδόμησης εξαλείφεται.

Πλεονεκτήματα σε σχέση με άλλους Αλγορίθμους Εντοπισμού

Τα πλεονεκτήματα της εφαρμογής της Γενίκευσης του Νόμου του Benford στον εντοπισμό κρυφού περιεχομένου μέσω του προτεινόμενου αλγόριθμου είναι:

-Υψηλά επίπεδα αναγνώρισης στεγανογραφημένων αρχείων.(Ενότητα 8.4)
- Δεν υπάρχει ανάγκη προηγούμενης βάσης δεδομένων με ύποπτα αρχεία για εκπαίδευση του συστήματος.
-Επεκτασιμότητα, ο συγκεκριμένος αλγόριθμος μπορεί να χρησιμοποιηθεί για την ανίχνευση Στεγανογραφίας ανεξαρτήτως τύπου αρχείου και φυσικά χωρίς το αρχικό αρχείο.

6.2 Γενικευμένη Αναδόμηση αρχείων

Με τον όρο γενικευμένη αναδόμηση αναφερόμαστε στην διαδικασία επαναδημιουργίας ενός αρχείου με ιδιότητες πολύ κοντά σε αυτές του αρχικού. Με τον όρο ιδιότητες αναφερόμαστε:

- Στην ποιότητα του περιεχομένου
- Στην δομή του αρχείου η οποία πρέπει να συνάδει με αυτή του προτύπου του τύπου αρχείου και έχει προδιαγραφεί από αναγνωρισμένους οργανισμούς

Οι διαδικασίες που μπορεί να αλλάξουν τις ιδιότητες ενός αρχείου μπορεί να είναι :

1. Χρήση προγραμμάτων για τροποποίηση ή αποθήκευση.

2. Απλή αντιγραφή ενός αρχείου μπορεί να οδηγήσει σε απώλεια ποιότητας λόγω των ίδιων των ιδιοτήτων των τύπου αρχείου (π.χ. JPEG).

3. Χρήση εργαλείων Στεγανογραφίας ή άλλων εργαλείων τα οποία παραποιούν τη αρχική δομή.

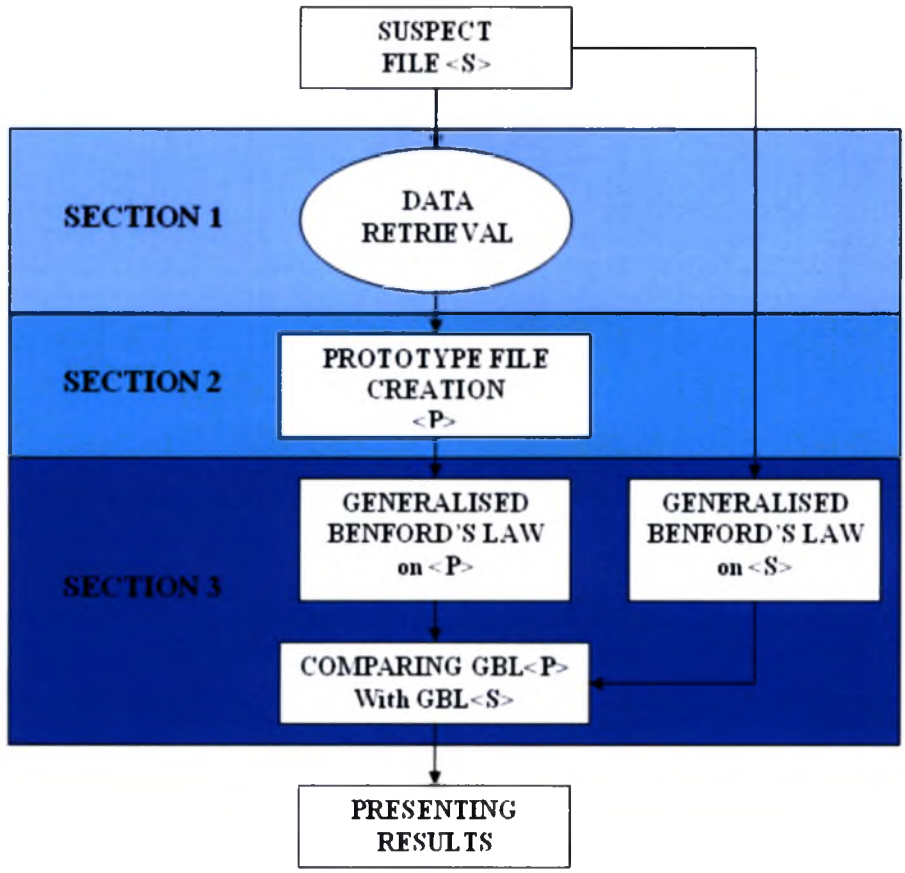
Εικόνα 6.1 : Απεικόνιση διαδοχικής συμπίεσης με lossy compression από αριστερά προς τα δεξιά. [113]

Δεδομένου του τύπου αρχείων που χρησιμοποιούνται στην Στεγανογραφία, όπως αρχεία εικόνων και ήχου, μια τέτοια διαδικασία μπορεί να είναι αρκετά επίπονη και όχι πάντα επιτυχημένη.

6.3 Προτεινόμενη μέθοδος Στεγανάλυσης με χρήση της Γενίκευσης του Νόμου του Benford.

Η μέθοδος που προτείνουμε μπορεί να διαιρεθεί σε τρεις αυτόνομες διαδικασίες, κάθε μία από τις οποίες παίζει σημαντικό ρόλο στην επιτυχή προσπάθεια να αναγνωρίσουμε ένα ύποπτο αρχείο το οποίο περιέχει Στεγανογραφία.

Οι τρεις βασικές διαδικασίες μαζί με κάποιες άλλες βοηθητικές περιγράφονται και απεικονίζονται παρακάτω.



Εικόνα 6.2: Στάδια Προτεινόμενης Στεγανάλυσης.

Αρχικό Στάδιο

Σε αυτό το στάδιο επιλέγουμε το ύποπτο αρχείο.

Διαδικασία 1

Εξακριβώνουμε την δομή / τύπο του αρχείου, χρησιμοποιώντας κατάλληλα προγράμματα.

Διαδικασία 2

Σε αυτό το στάδιο αναδομούμε το αρχείο, όπως περιγράφηκε σε προηγούμενη ενότητα.

Διαδικασία 3

Εφαρμόζουμε την γενίκευση του νόμου του Benford, τόσο για το αρχείο που δημιουργήσαμε όσο και για το αρχικό αρχείο. Έπειτα συγκρίνουμε τα αποτελέσματα.

Τελικό Στάδιο

Παρουσιάζουμε τα αποτελέσματα μας. Το υπό εξέταση αρχείο μπορεί να είναι Στεγανογραφημένο ή μη Στεγανογραφημένο.

Παράδειγμα απλής Στεγανάλυσης αρχείου κειμένου (.TXT)

Σε ένα απλουστευμένο παράδειγμα, θα προσπαθήσουμε να εφαρμόσουμε όσα περιγράφηκαν στην προτεινόμενη μέθοδο Στεγανάλυσης.
Το αρχείο που θα στεγαναλύσουμε θα είναι αρχείο κειμένου, τύπου .TXT, με όνομα **"free.txt"**, που δημιουργήθηκε με το πρόγραμμα notepad.exe των Windows και στο οποίο έχει κρυφτεί πληροφορία με χρήση του εργαλείου Στεγανογραφίας Camouflage[55].Το μέγεθος του αρχείου **"free.txt"**,ήταν 634 bytes πριν την Στεγανογραφία και 1,53 kb μετά. Η κρυφή πληροφορία είναι επίσης ένα αρχείο κειμένου με όνομα **"secret.txt"** και μέγεθος 85 bytes.

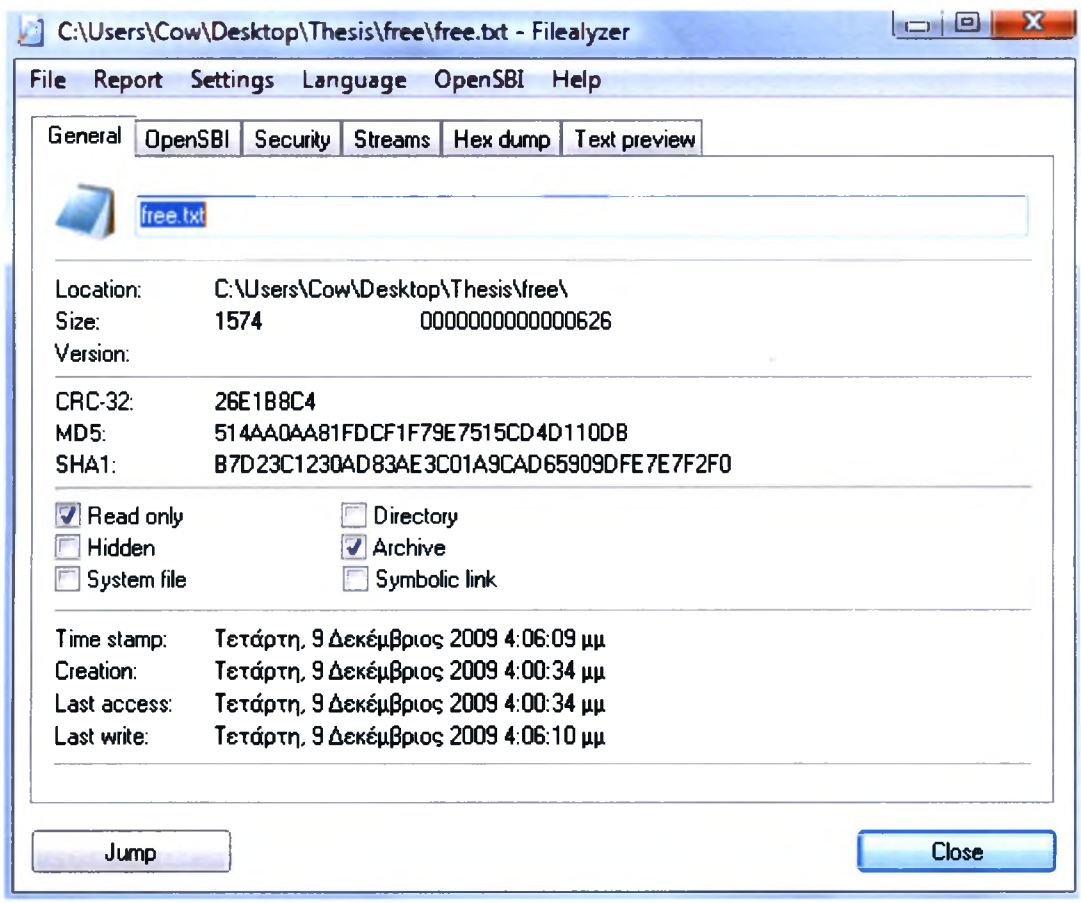
Εικόνα 6.3: Τα τρία αρχεία txt, και τα μεγέθη τους.

Αρχικό Στάδιο

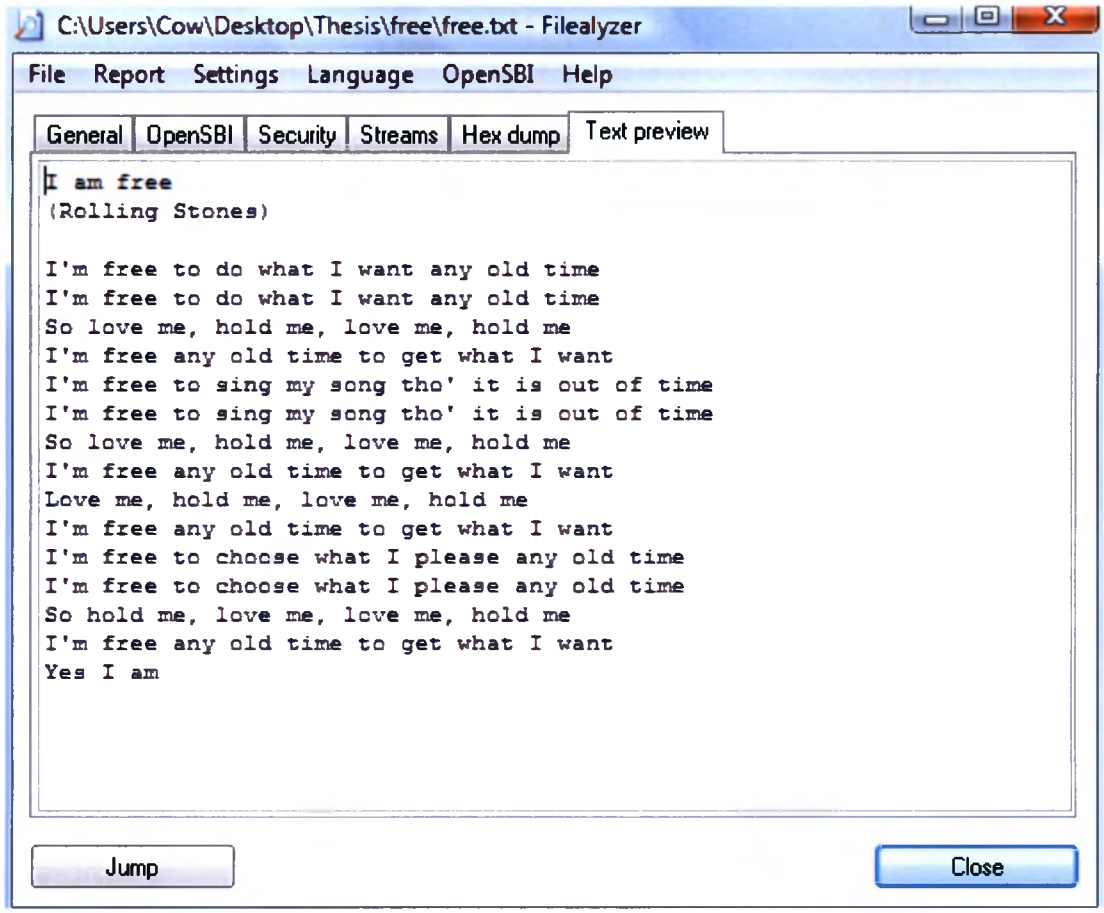
Επιλέγουμε το Αρχείο **"free.txt"**, προκειμένου να το αναλύσουμε.

Διαδικασία 1

Εξακριβώνουμε την δομή και το περιεχόμενο του αρχείου, χρησιμοποιώντας κατάλληλα προγράμματα όπως το FileAlyzer [105].



Εικόνα 6.4: Χαρακτηριστικά του υπόπτου αρχείου.

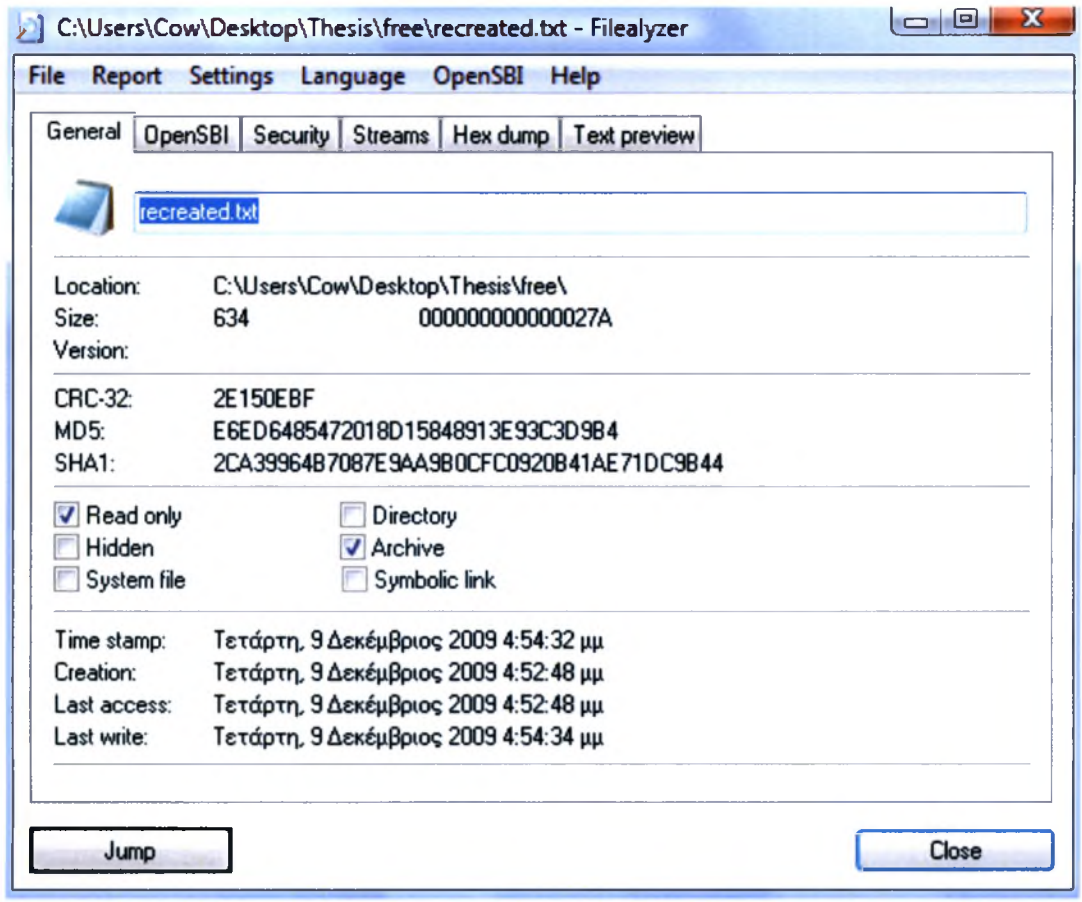


Εικόνα 6.5: Περιεχόμενο του Υπόπτου αρχείου.

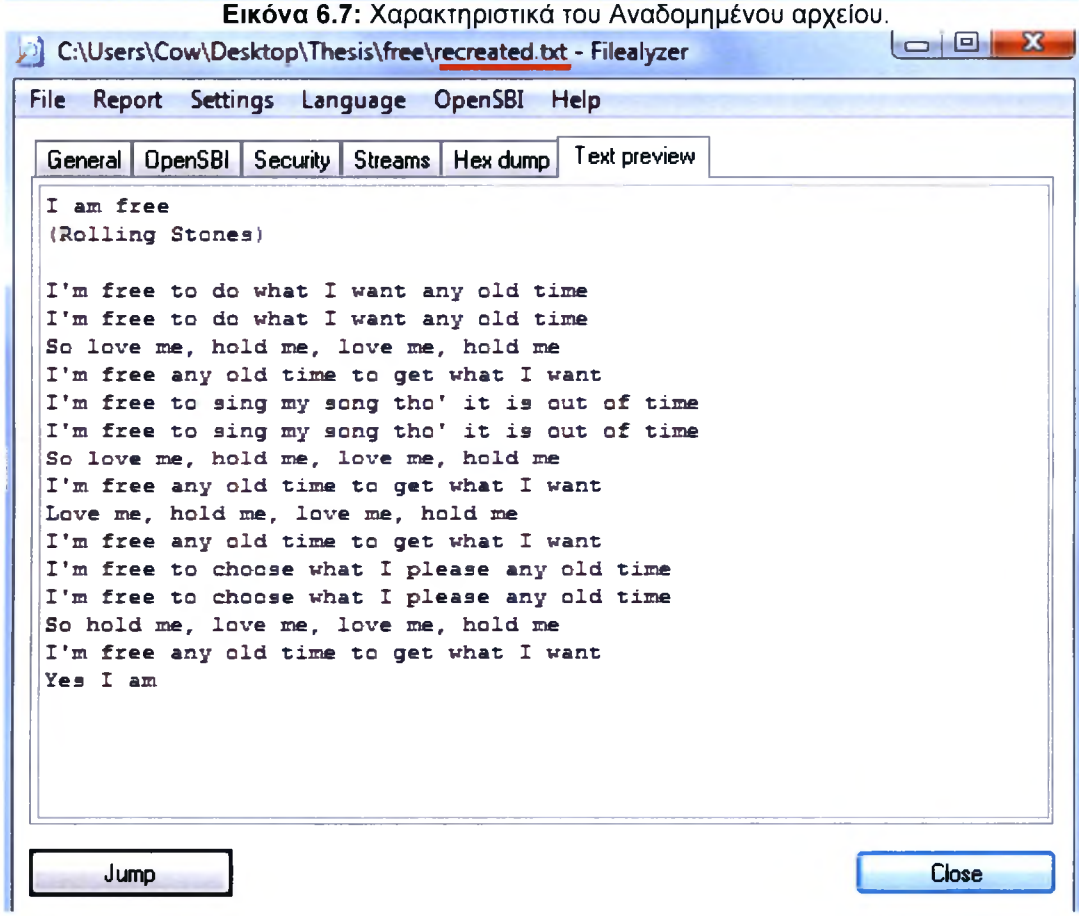
Διαδικασία 2

Σε αυτό το στάδιο αναδομούμε το αρχείο χρησιμοποιώντας το περιεχόμενο αλλά και τα χαρακτηριστικά του αρχείου που ανακτήθηκαν από την προηγούμενη διαδικασία. Έτσι παράγουμε το "recreated.txt" με μέγεθος 634 bytes (Εικόνα 6.6).

Εικόνα 6.6



Εικόνα 6.7: Χαρακτηριστικά του Αναδομημένου αρχείου.



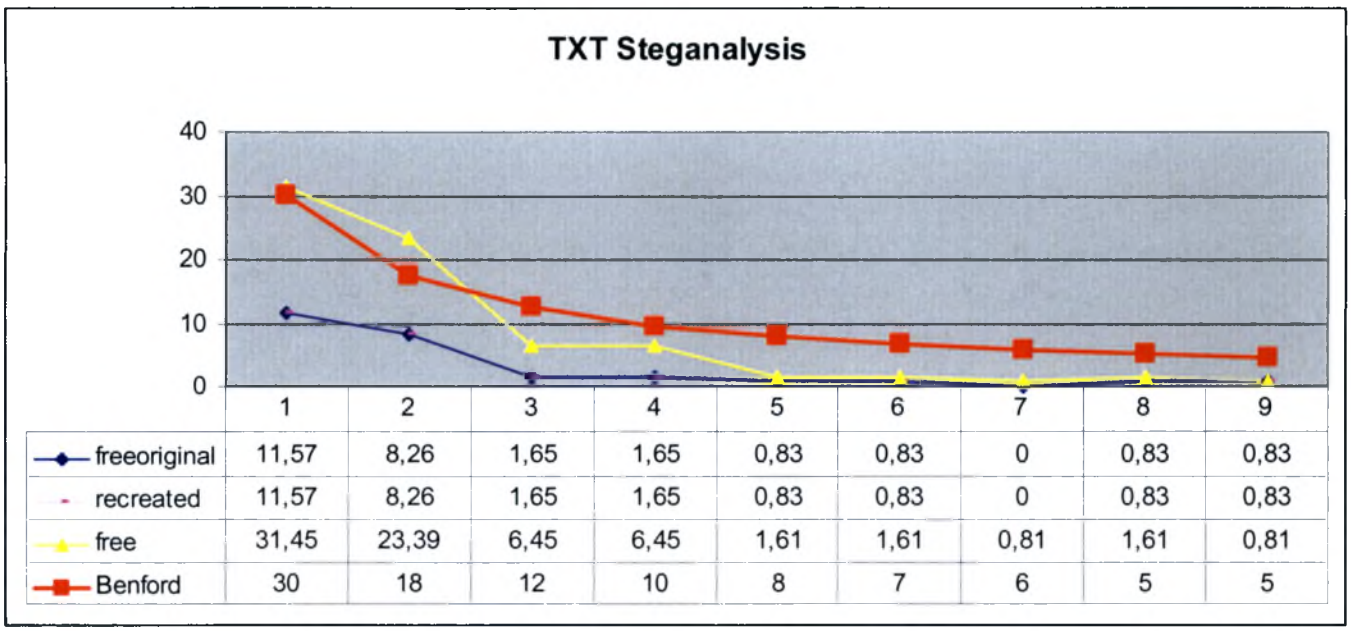
Εικόνα 6.8: Περιεχόμενο του Αναδομήμένου αρχείου.

Διαδικασία 3

Εφαρμόζουμε την γενίκευση του νόμου του Benford, τόσο για το αρχείο που δημιουργήσαμε όσο και για το ύποπτο αρχείο.

1. Υπολογίζουμε την αλληλουχία byte array που αποτελεί το Ύποπτο Αρχείο *F[file.length()]*.
2. Υπολογίζουμε την πιθανότητα *x=1,2,…,9, για* το πρώτο ψηφίο της αλληλουχίας byte array *F[file.length()]* για το Ύποπτο Αρχείο.
3. Υπολογίζουμε την αλληλουχία byte array που αποτελεί το Αναδομημένο Αρχείο *O[file.length()]*.
4. Υπολογίζουμε την πιθανότητα *x=1,2,…,9,για* το πρώτο ψηφίο της αλληλουχίας byte array *O[file.length()]* για το Αναδομημένο Αρχείο.

Συγκρίνοντας τα αποτελέσματα:



	1	2	3	4	5	6	7	8	9
freeoriginal	11,57	8,26	1,65	1,65	0,83	0,83	0	0,83	0,83
recreated	11,57	8,26	1,65	1,65	0,83	0,83	0	0,83	0,83
free	31,45	23,39	6,45	6,45	1,61	1,61	0,81	1,61	0,81
Benford	30	18	12	10	8	7	6	5	5

Παρατηρούμε ότι το αναδομημένο αρχείο έχει μεγάλες διαφορές για όλα τα ψηφία της αλληλουχίας byte array. Αντίθετα (για να επαληθεύσουμε την εγκυρότητα της μέθοδου) εάν συγκρίνουμε τις τιμές του αρχικού αρχείου (free original) με το δικό του αναδομημένο, αλλα και το αναδομημένο του free.txt θα δούμε ότι είναι ακριβώς ίδιες.Αυτό μας οδηγεί σε μία 100% ανίχνευση ύποπτου αρχείου για κρυφό περιεχόμενο.

Τελικό Στάδιο

Το υπό εξέταση αρχείο είναι Στεγανογραφημένο.

ΜΕΡΟΣ ΙΙΙ - ΕΦΑΡΜΟΓΗ

Προκειμένου να επαληθεύσουμε την μέθοδο ανακάλυψης κρυφού περιεχομένου με χρήση της γενίκευσης του Νόμου του Benford, θα την εφαρμόσουμε σε αρχεία εικόνων. Ύστερα από μελέτη πληθώρας εργαλείων και μεθόδων Στεγανογραφίας καταλήξαμε στο συμπέρασμα ότι η Στεγανογραφία σε αρχεία εικόνων είναι περισσότερο διαδεδομένη από κάθε άλλη μορφή Στεγανογραφίας.
Ακόμη, από όλους τους τύπους εικόνων (format) που διακινούνται στο Ίντερνετ και υποστηρίζονται από εργαλεία Στεγανογραφίας, θεωρήθηκε επικρατέστερο το JPEG.

- Οι μέθοδοι Στεγανογραφίας που πρόκειται να εξεταστούν λοιπόν κρύβουν πληροφορία μέσα σε διαφορετικά σημεία μιας εικόνας τύπου JPEG.
- Η κάθε Στεγανογραφική μέθοδος / εργαλείο που θα εξεταστεί κρύβει την πληροφορία σε διαφορετικό μέρος της JPEG εικόνας.
- Το κάθε εργαλείο που επιλέχθηκε δοκιμάσθηκε εκτενώς έτσι ώστε τα παραγόμενα αποτελέσματα να είναι ασφαλή.

Οι τεχνικές και εργαλεία που θα χρησιμοποιηθούν και έχουν περιγραφεί εκτενώς σε προηγούμενη ενότητα είναι:

Μέθοδος Στεγανογραφίας : Least Significant Bit (LSB)
Λογισμικό Στεγανογράφησης : JPHSWin [40]

Μέθοδος Στεγανογραφίας: Fuse
Λογισμικό Στεγανογράφησης: Invisible Secrets [48]

Μέθοδος Στεγανογραφίας: Fuse
Λογισμικό Στεγανογράφησης: Camouflage [55]

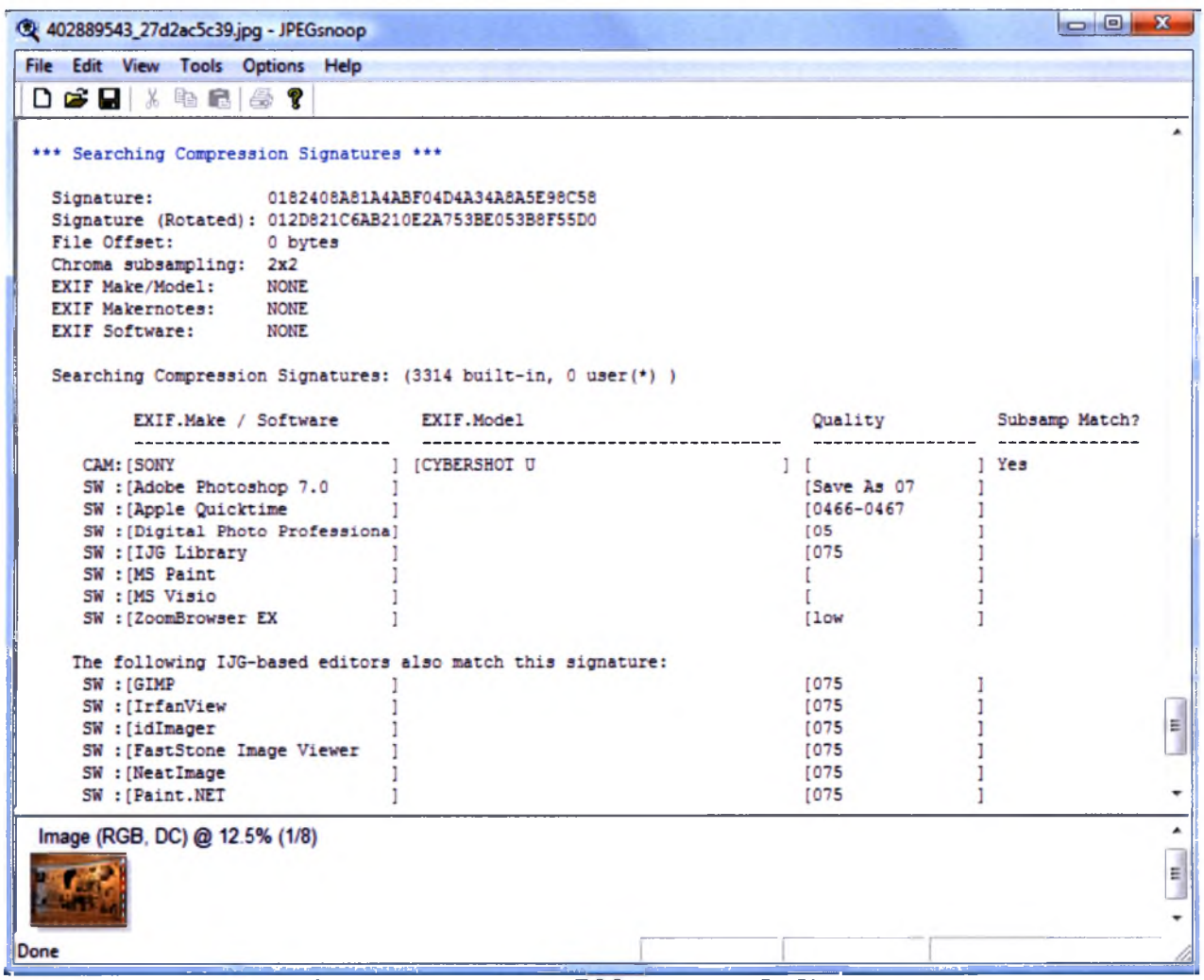
Για χάριν της εργασίας αυτής θα διακρίνουμε τέσσερις (4) τύπους εικόνων:

1. **Το Αυθεντικό (Αρχικό) αρχείο**, στην δικής μας περίπτωση θα είναι μία εικόνα τύπου JPEG που δημιουργήθηκε με MS Paint.
2. **Το Αρχείο Φορέας (carrier file)**, που στην δική μας περίπτωση θα είναι το αρχείο εικόνας JPEG που θα παραχθεί από την εφαρμογή στεγανογραφίας στο Αυθεντικό αρχείο μέσω ενός από τα προγράμματα που αναφέρθηκαν παραπάνω.
3. **Το Αναδομημένο αρχείο**, που στην περίπτωσή μας θα παραχθεί από λογισμικό που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής και παράγει αρχεία εικόνων JPEG με δομή παρόμοια με αυτή του MS Paint.
4. **Το Ύποπτο αρχείο**, το οποίο μπορεί να είναι είτε Αυθεντικό αρχείο είτε αρχείο Φορέας.

7.1 Αναδόμηση αρχείων

Σημαντικό τμήμα της εργασίας αποτελεί η ορθή αναδόμηση των αρχείων η οποία δρα σαν βάση για την σύγκριση μας και παίζει καταλυτικό ρόλο στον εντοπισμό Στεγανογραφημένων αρχείων.
Σκοπός μας λοιπόν είναι η όσο το δυνατόν πιο ακριβής αναδόμηση του αρχείου / εικόνας ώστε να μοιάζει τόσο σε δομή όσο σε περιεχόμενο με το αρχικό αρχείο πριν από οποιαδήποτε δηλαδή παρέμβαση τρίτου(π.χ. Στεγανογραφία).

Για να επιτευχθεί λοιπόν η σωστή αναδόμηση της εικόνας JPEG πρέπει πρώτα να αναγνωριστεί το λογισμικό ή ο αλγόριθμος κωδικοποίησης (encoding algorithm) της Ύποπτης εικόνας (EXIF Make).Για τον σκοπό αυτό χρησιμοποιήθηκε με επιτυχία ένα εργαλείο ανοικτού λογισμικού που διενεργεί ανάλυση εικόνας με όνομα "*JPEGSnoop*" [119].Το συγκεκριμένο εργαλείο έχει την δυνατότητα να αναγνωρίζει εκτός από το λογισμικό που δημιούργησε μία εικόνα πιθανά μοντέλα φωτογραφικών μηχανών (EXIF metadata, IPTC) και άλλες πολλές πληροφορίες που θα είναι χρήσιμες στην συνέχεια όπως ανάλυση (quality) εικόνας ή αλγόριθμο συμπίεσης για την αποθήκευση της εικόνας.



Εικόνα 7.1:Το εργαλείο JPEGSnoop εντοπίζει Signatures

Η παραπάνω διαδικασία αναγνώρισης και εξέτασης των χαρακτηριστικών της εικόνας είναι γνωστή στην βιβλιογραφία με τον όρο Digital Image Ballistics / Forensics.

Στην προσπάθειά μας να αναδομήσουμε μία εικόνα χρησιμοποιώντας τον αλγόριθμο κωδικοποίησης που εντοπίσαμε πως είχε αρχικά [28][31] επιλέξαμε να αναπτύξουμε εργαλείο σε JAVA το οποίο μιμείται την κωδικοποίηση του εργαλείου ζωγραφικής MS Paint.
Αυτό το εργαλείο λοιπόν θα λειτουργήσει σαν εξομοιωτής του MS Paint με σκοπό την όσο πιο πιστή αναδόμηση των αρχείων που είχαν υποστεί αλλοιώσεις. Το εργαλείο αυτό μπορεί να επεκταθεί εύκολα ώστε να μιμείται την κωδικοποίηση JPEG και άλλων εργαλείων (π.χ. Adobe Photoshop), με αλλαγές διαφόρων παραμέτρων που δεν θα εξεταστούν στα πλαίσια αυτής της εργασίας.

Διαδικασία Αναδόμησης

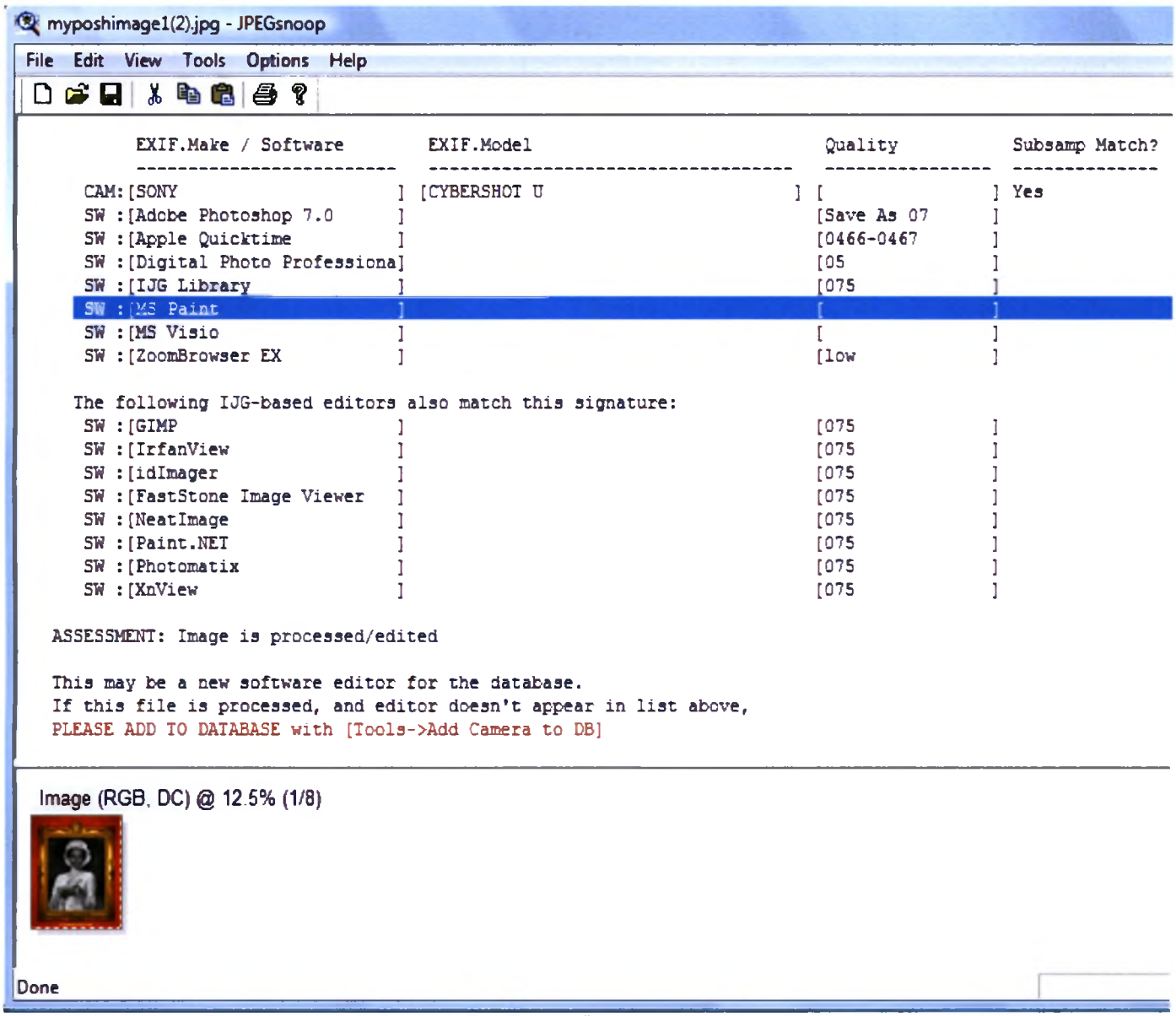
Για να επιτύχουμε γρήγορη και αποδοτική αναδόμηση της Αρχικής εικόνας από μία Ύποπτη, χρησιμοποιήσαμε γνωστές βιβλιοθήκες επεξεργασίας εικόνας της JAVA .
Οι διαδικασίες που θα περιγραφούν στην συνέχεια έχουν να κάνουν με την δημιουργία της αναδομημένης εικόνας που θα παραγόταν από το MS Paint και του πώς αυτή θα χρησιμοποιηθεί για να στηρίξει την Υπόθεσή μας.

Βήμα 1 – Χρήση JPEGSnoop

Δεδομένου ενός Υπόπτου Αρχείου < *S* >,η διαδικασία αναδόμησης ξεκινά με την ανάκτηση ορισμένων δεδομένων και πληροφοριών για την δομή και τα χαρακτηριστικά της εικόνας, με χρήση του εργαλείου JPEGSnoop .
Οι πληροφορίες που ανακτώνται είναι:

1. Ο πίνακας Κβάντισης (Quantization Table ή Quality Factor)
2. Το “EXIF Make” ή Υπογραφή Λογισμικού (Software Signature)από το οποίο δημιουργήθηκε η υπό εξέταση εικόνα.

Είναι σημαντικό να υπογραμμιστεί ότι τα τρία υπό εξέταση εργαλεία (*JPHSWin, Camouflage, Invisible Secrets*) δεν επανακωδικοποιούν (re-encode) την JPEG εικόνα. Μόνο τμήματα της εικόνας αλλάζουν προκειμένου να κρυφτεί η πληροφορία αλλά το “EXIF Make” ή Υπογραφή Λογισμικού δεν αλλοιώνονται. Λαμβάνοντας τα παραπάνω υπόψη μπορούμε εύκολα να υποθέσουμε ότι η χρήση του εργαλείου JPEGSnoop σε μία εικόνα μετά την χρήση Στεγανογραφίας με ένα από τα παραπάνω προγράμματα θα έχει τα ίδια αποτελέσματα, δηλαδή αυτά του αρχικού “EXIF Make” ή Υπογραφής Λογισμικού. Γνωρίζοντας λοιπόν το αρχικό λογισμικό που δημιούργησε την εικόνα μπορούμε να μιμηθούμε τον αλγόριθμο κωδικοποίησης του για να φτάσουμε στην αρχική εικόνα υπερπηδώντας τις αλλαγές του εργαλείου Στεγανογραφίας.



Εικόνα 7.2: "EXIF Make" ή Υπογραφή Λογισμικού

```
Precision=8 bits
Destination ID=0 (Luminance)
  DQT, Row #0:   8   6   5   8  12  20  26  31
  DQT, Row #1:   6   6   7  10  13  29  30  28
  DQT, Row #2:   7   7   8  12  20  29  35  28
  DQT, Row #3:   7   9  11  15  26  44  40  31
  DQT, Row #4:   9  11  19  28  34  55  52  39
  DQT, Row #5:  12  18  28  32  41  52  57  46
  DQT, Row #6:  25  32  39  44  52  61  60  51
  DQT, Row #7:  36  46  48  49  56  50  52  50
  Approx quality factor = 74.75 (scaling=50.51 variance=0.81)
```

Εικόνα 7.3: Luminance Quality Factor

```
Precision=8 bits
Destination ID=1 (Chrominance)
  DQT, Row #0:   9   9  12  24  50  50  50  50
  DQT, Row #1:   9  11  13  33  50  50  50  50
  DQT, Row #2:  12  13  28  50  50  50  50  50
  DQT, Row #3:  24  33  50  50  50  50  50  50
  DQT, Row #4:  50  50  50  50  50  50  50  50
  DQT, Row #5:  50  50  50  50  50  50  50  50
  DQT, Row #6:  50  50  50  50  50  50  50  50
  DQT, Row #7:  50  50  50  50  50  50  50  50
  Approx quality factor = 74.74 (scaling=50.52 variance=0.19)
```

Εικόνα 7.4: Chrominance Quality Factor

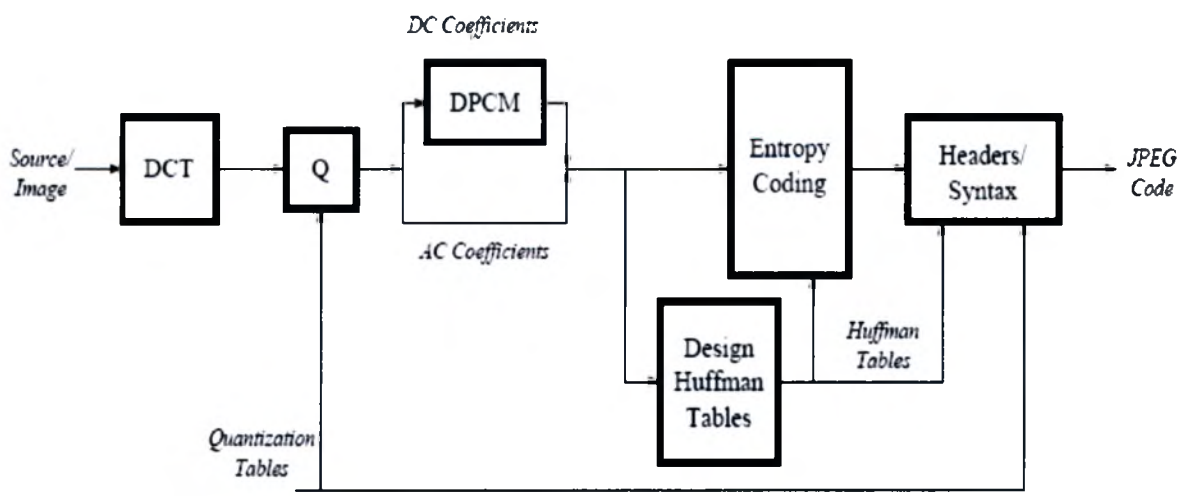
```
*** Searching Compression Signatures ***

Signature:           0182408A81A4ABF04D4A34A8A5E98C58
Signature (Rotated): 012D821C6AB210E2A753BE053B8F55D0
```

Εικόνα 7.5 Η Μοναδική Υπογραφή του λογισμικού MS Paint

Αφού εξάγουμε την Υπογραφή Λογισμικού (στην περίπτωσή μας MS Paint) και τον Πίνακα Κβάντισης προχωράμε στο Βήμα *2*.

Βήμα 2 – Κωδικοποίηση Αναδομημένης Εικόνας

Σε αυτό το βήμα σκοπός μας είναι να επιτύχουμε εξομοίωση της κωδικοποίησης του λογισμικού που αναγνωρίσθηκε ότι χρησιμοποιήθηκε στο πρώτο βήμα .
Έχοντας καθορίσει σε προηγούμενο βήμα τους πίνακες κβάντισης και γνωρίζοντας τους πίνακες Huffman αλλά και την κεφαλίδα που χρησιμοποιεί το λογισμικό MS Paint μπορούμε να προχωρήσουμε στην κωδικοποίηση.
Το τελευταίο κομμάτι που υπολείπεται είναι ή ίδια η εικόνα δηλαδή η χρωματική πληροφορία που την αποτελεί (source / image RGB / YCC values) η οποία είναι απαραίτητη για την αναπαραγωγή της εικόνας.



Εικόνα 7.6: Το JPEG Σύστημα Κωδικοποίησης

Διαφορετικές Μέθοδοι έχουν αναπτυχθεί για την ανάκτηση και ορθή αναδόμηση των τιμών RGB / YCC [14] με την μεγαλύτερη δυνατή ακρίβεια

ώστε η επανακωδικοποίηση να γίνει χωρίς απώλειες (lossless).

Θα χρησιμοποιηθούν δύο μέθοδοι επανα-κωδικοποίησης της εικόνας:

1. Χρήση του JAVA Advance Imaging API (JAI). (λιγότερο ακριβής)
2. Χρήση του Independent JPEG Group[29] (IJG) Software. (ακριβής) (δεν απαιτεί το βήμα 1).

Η αναδομημένη εικόνα που δημιουργείται σαν αποτέλεσμα της παραπάνω διαδικασίας θα χρησιμοποιηθεί σαν την βάση της σύγκρισής μας με γνώμονα τα αποτελέσματα της γενίκευσης του Νόμου του Benford.

7.2 Εφαρμογή της προτεινόμενης μεθόδου Στεγανάλυσης σε επιλεγμένα format αρχείων

Οι διαδικασίες Στεγανάλυσης που θα περιγραφούν στην ενότητα αυτή αφορούν Ύποπτα αρχεία εικόνων JPEG και ο αναλυτής που τα χειρίζεται:

1. Δεν γνωρίζει εάν είναι Στεγανογραφημένα ή όχι.
2. Δεν γνωρίζει το κρυφό περιεχόμενο ή τον αλγόριθμο Στεγανογραφίας (εφόσον το αρχείο είναι Στεγανογραφημένο).

Η διαφορά των δύο μέθοδος βρίσκεται στην μέθοδο αναδόμησης της εικόνας. Στην πρώτη περίπτωση χρησιμοποιείται κώδικας αναδόμησης που αναπτύχθηκε στα πλαίσια αυτής της εργασίας με χρήση βιβλιοθηκών J.A.I., ενώ στην δεύτερη περίπτωση ολόκληρη η διαδικασία αναδόμησης γίνεται αυτόματα από ένα ανοικτού κώδικα (open source) εργαλείο για αναδόμηση εικόνων JPEG, το Independent JPEG Group (IJG)[29] Software που υλοποιεί παρόμοια διαδικασία με αυτή που περιγράφηκε προηγουμένως αλλά είναι black box, δεν μπορούμε δηλαδή να παρέμβουμε στον αλγόριθμό κωδικοποίησης του.

Μέθοδος 1 : Custom encoding (J.A.I.)

Δεδομένου ενός Υπόπτου Αρχείου <*S*> :

Διαδικασία 1 – Ανάκτηση Δεδομένων (Data Retrieval)

1. Εξάγουμε τον πίνακα κβάντισης (quantization table /quality factor) από το <*S*>.
2. Εξάγουμε την Υπογραφή Λογισμικού, (“EXIF Make”/Software Signature) για να εντοπίσουμε το πρόγραμμα που δημιούργησε την εικόνα.
3. Εξάγουμε τις κεφαλίδες του αρχείου <*S*>.
4. Οι χρωματικές τιμές RGB / YCC εξάγονται από το αρχείο <*S*> με την βοήθεια του JAVA Advanced Imaging API.

Διαδικασία 2 – Αναδόμηση Εικόνας (Image Reconstruction)

1. Γνωρίζοντας το πρόγραμμα που δημιούργησε την εικόνα <S> καθορίζεται ο τύπος κωδικοποίησης που θα χρησιμοποιηθεί για την δημιουργία της αναδομημένης εικόνας
2. Ανάλογα με τον τύπο κωδικοποίησης καθορίζονται οι τιμές διαφόρων παραμέτρων που θα μας φανούν χρήσιμες για την JPEG κωδικοποίηση της εικόνας, τέτοιες παράμετροι είναι:

   -Κεφαλίδες
   -Πίνακες κβάντισης
   -Πίνακες Huffman

3. Η αναδομημένη εικόνα <*P*> κωδικοποιείται.

Διαδικασία 3 – Αλγόριθμος Εντοπισμού Κρυφού Περιεχομένου (Detection Algorithm)

1. Εφαρμόζουμε την γενίκευση του Νόμου του Benford στην ύποπτη εικόνα <*S*> *(Ενότητα 6.1)*.
2. Καταγράφουμε το μέγεθος και τιμή κατακερματισμού (hash value) του Ύποπτου αρχείου <*S*>.
3. Εφαρμόζουμε την γενίκευση του Νόμου του Benford στην *Αναδομημένη εικόνα <P> (Ενότητα 6.1)*.
4. Καταγράφουμε το μέγεθος και τιμή κατακερματισμού (hash value) *της αναδομημένης εικόνας <P>*.
5. Συγκρίνουμε τα αποτελέσματα των βημάτων 1 & 3 Διαδικασία *3*, ανανεώνουμε την τιμή του παράγοντα ομοιότητας (similarity factor).
6. Συγκρίνουμε τα αποτελέσματα των βημάτων 2 & 4 Διαδικασία *3*, ανανεώνουμε την τιμή του παράγοντα ομοιότητας (similarity factor).
7. Αποφασίζουμε εάν ένα ύποπτο αρχείο είναι Στεγανογραφημένο ή όχι με βάση τον παρακάτω αλγόριθμο:

```
if (similarity_factor > similarity_ threshold){
        "Suspect File <S> is not Carrier File"
}
 else {"Suspect File <S> is  Carrier File" }
```

Ο Παράγοντα Ομοιότητας (similarity factor) αποτελεί μία παράμετρο που καθορίζει πόσο όμοια τόσο σε δομή όσο και σε περιεχόμενο είναι δυο αρχεία. Οι τιμές του Παράγοντα Ομοιότητας κυμαίνονται από μηδέν μέχρι εννιά βαθμούς ομοιότητας (0-9) με το εννιά (9) να χαρακτηρίζει την μεγαλύτερη ομοιότητα μεταξύ δύο αρχείων.
Ανάλογα με τον αλγόριθμο Στεγανογραφίας που εξετάζουμε υπάρχει και διαφορετικός τρόπος εντοπισμού του Παράγοντα Ομοιότητας.

Έτσι για αρχεία στα οποία μπορεί να έχουμε LSB Στεγανογραφία ο Παράγοντας Ομοιότητας υπολογίζεται μεταξύ του ύποπτου και του αναδομημένου αρχείου από τον παρακάτω αλγόριθμο:

```
Similarity_factor=0;
for (int i=1;i<=9;i++){
        if(Math.abs(<S>GBL(i) - <P>GBL(i))==0)
                Similarity_factor++;
}

if(Math.abs(<S>.size()-<P>.size())<=800){
        Similarity_factor++;
}else {
        Similarity_factor--;
}
```

Επεξηγώντας κάποιες από τις παραπάνω παραμέτρους:

- Ως **<S>GBL(x)** θεωρούμε το αποτέλεσμα της γενίκευσης του Νόμου του Benford στην αλληλουχία των bytes ενός Υπόπτου Αρχείου, για το σημαντικότερο και το λιγότερο σημαντικό ψηφίο με τιμή x όπου x = 1,2…9.

- Ως **<P>GBL(x)** θεωρούμε το αποτέλεσμα της γενίκευσης του Νόμου του Benford στην αλληλουχία των bytes ενός Αναδομημένου Αρχείου, για το σημαντικότερο και το λιγότερο σημαντικό ψηφίο με τιμή x όπου x = 1,2…9.

Για αρχεία στα οποία μπορεί να έχουμε Στεγανογραφία τύπου FUSE ο Παράγοντας Ομοιότητας υπολογίζεται μεταξύ του ύποπτου και του αναδομημένου αρχείου από τον παρακάτω αλγόριθμο:

```
Similarity_factor=9;
Size_dif=Math.abs(<S>.size()-<P>.size());

for (int i=1;i<=9;i++){
        if(Math.abs(<S>GBLp(i) - <P>GBLp(i))==0)
                Similarity_factor++;
 }
```

Επεξηγώντας κάποιες από τις παραπάνω παραμέτρους:

Ως **<S>GBLp(x)** θεωρούμε το αποτέλεσμα της γενίκευσης του Νόμου του Benford στην αλληλουχία των bytes *τμήματος* ενός Υπόπτου Αρχείου, για το σημαντικότερο και το λιγότερο σημαντικό ψηφίο με τιμή x όπου x = 1,2…9.

Ως **<P>GBLp(x)** θεωρούμε το αποτέλεσμα της γενίκευσης του Νόμου του Benford στην αλληλουχία των bytes *τμήματος* ενός Αναδομημένου Αρχείου,

για το σημαντικότερο και το λιγότερο σημαντικό ψηφίο με τιμή x όπου x = 1,2...9.

Το τμήμα που εξετάζεται κάθε φορά σε αυτό τον αλγόριθμο έχει μέγεθος ίσο με την απόλυτή τιμή της διαφοράς μεγέθους του Ύποπτου και του Αναδομημένου Αρχείου :

```
Size_dif = Math.abs( <S>.size()-<P>.size() )
```

Υπολογισμός Κατωφλίου Ομοιότητας

Ως ***Κατώφλι Ομοιότητας*** (*Similarity Threshold*) ορίζουμε μία σταθερά που έχει υπολογιστεί μετά από εφαρμογή της γενίκευσης του Νόμου του Benford σε μεγάλο αριθμό κανονικών και στεγανογραφημένων αρχείων. Η τιμή της σταθεράς αυτής είναι άμεσα εξαρτώμενη από τον τύπο κωδικοποίησης της εικόνας. Έτσι για παράδειγμα το MS Paint έχει διαφορετικό Κατώφλι ομοιότητας από το Photoshop 9

Ύστερα από εκτενή μελέτη μεγάλου αριθμού εικόνων διευκρινίστηκε η ακριβής τιμή του ***Κατωφλίου Ομοιότητας*** του λογισμικού MS Paint που χρησιμοποιήθηκε στα πειράματά μας για την κωδικοποίηση της εικόνας.
Η τιμή αυτή υπολογίσθηκε με απόκλιση κοντά στο πέντε (~5).
Ο παρακάτω πίνακας απεικονίζει τα πειραματικά αποτελέσματα από την εφαρμογή της γενίκευσης του νόμου σε 1500 εικόνες JPEG διαφορετικών διαστάσεων και μεγέθους.

JPEG Image Size	320 x 240	600 x 320	800 x 600
Number of Original Image Files used	500	500	500
Average similarity factor	7.86	6.98	5.66
Minimum/Maximum	6 / 9	5 / 9	5 / 9
Number of Carrier Image Files used	1500	1500	1500
Average similarity factor	1.23	3.25	4.57
Minimum/Maximum	0 / 3	1 / 4	2 / 5

Εικόνα 7.7: Στατιστικά Μέγιστου, Μέσου και Ελάχιστου Παράγοντα Ομοιότητας

Για την δημιουργία των στεγανογραφημένων αρχείων, τα κανονικά αρχεία χωρίστηκαν σε τρεις διαφορετικές ομάδες των 1500 εικόνων. Για κάθε ομάδα εφαρμόστηκε διαφορετικός αλγόριθμος στεγανογραφίας από τους *JPHSWin, Camouflage και Invisible Secrets* προκειμένου να στεγανογραφηθεί η μικρότερη δυνατή πληροφορία (1kb ASCII .txt file).
Όπως γίνεται εμφανές και στην παραπάνω εικόνα, ο βαθμός ομοιότητας 5 μπορεί να θεωρηθεί σαν το κατάλληλο ***Κατωφλίου Ομοιότητας.***

Κάτω από την τιμή 5 το Ύποπτο Αρχείο που εξετάζεται μπορεί να θεωρηθεί Στεγανογραφημένο ενώ τιμές μεγαλύτερες του 5 δείχνουν «καθαρό» αρχείο.

Ένα ποσοστό της τάξης του 1.5% βρέθηκε *false positive* στεγανογραφημένο αλλά το ποσοστό αυτό μειώθηκε αισθητά για κρυμμένα αρχεία μεγαλύτερου μεγέθους (5KB ASCII .txt file). Σε επόμενο κεφάλαιο θα παρουσιαστούν τεχνικές που μειώνουν ακόμη περισσότερο τις *false-positive προβλέψεις.*

Η παραπάνω μεθοδολογία μπορεί να επεκταθεί σε πολλά διαφορετικά είδη (format) κωδικοποίησης εικόνας (όπως BMP, TIFF, PNG), με σκοπό την μεγαλύτερη δυνατή κάλυψη εντοπισμού στεγανογραφημένων αρχείων εικόνας.

Μέθοδος 2 : Independent JPEG Group (IJG) Software

Δεδομένου ενός Υπόπτου Αρχείου <*S*> :

Διαδικασία 1 – Αναδόμηση Εικόνας (Image Reconstruction)

1. Εισαγωγή του αρχείου <S> στο Independent JPEG Group (IJG) Software.
2. Η αναδομημένη εικόνα <*P*> κωδικοποιείται.

Διαδικασία 2 – Αλγόριθμος Εντοπισμού Κρυφού Περιεχομένου (Detection Algorithm)

1. Εφαρμόζουμε την γενίκευση του Νόμου του Benford στην ύποπτη εικόνα <*S*> *(Ενότητα 6.1)*.
2. Καταγράφουμε το μέγεθος και τιμή κατακερματισμού (hash value) του Ύποπτου αρχείου <*S*>.
3. Εφαρμόζουμε την γενίκευση του Νόμου του Benford στην *Αναδομημένη εικόνα <P> (Ενότητα 6.1)*.
4. Καταγράφουμε το μέγεθος και τιμή κατακερματισμού (hash value) *της αναδομημένης εικόνας <P>*.
5. Συγκρίνουμε τα αποτελέσματα των βημάτων 1 & 3 Διαδικασία *2*, ανανεώνουμε την τιμή του παράγοντα ομοιότητας (similarity factor).
6. Συγκρίνουμε τα αποτελέσματα των βημάτων 2 & 4 Διαδικασία *2*, ανανεώνουμε την τιμή του παράγοντα ομοιότητας (similarity factor).
7. Αποφασίζουμε εάν ένα ύποπτο αρχείο είναι Στεγανογραφημένο ή όχι με βάση τον παρακάτω αλγόριθμο:

```
if (similarity_factor > similarity_ threshold){
        "Suspect File <S> is not Carrier File"
}
 else {"Suspect File <S> is  Carrier File" }
```

Ο Παράγοντα Ομοιότητας (similarity factor) αποτελεί μία παράμετρο που καθορίζει πόσο όμοια τόσο σε δομή όσο και σε περιεχόμενο είναι δυο αρχεία. Οι τιμές του Παράγοντα Ομοιότητας κυμαίνονται από μηδέν μέχρι εννιά βαθμούς ομοιότητας (0-9) με το εννιά (9) να χαρακτηρίζει την μεγαλύτερη ομοιότητα μεταξύ δύο αρχείων.
Ανάλογα με τον αλγόριθμο Στεγανογραφίας που εξετάζουμε υπάρχει και διαφορετικός τρόπος εντοπισμού του Παράγοντα Ομοιότητας.

Έτσι για αρχεία στα οποία μπορεί να έχουμε LSB Στεγανογραφία ο Παράγοντας Ομοιότητας υπολογίζεται μεταξύ του ύποπτου και του αναδομημένου αρχείου από τον παρακάτω αλγόριθμο:

```
Similarity_factor=0;
for (int i=1;i<=9;i++){
        if(Math.abs(<S>GBL(i) - <P>GBL(i))==0)
                Similarity_factor++;
}

if(Math.abs(<S>.size()-<P>.size())<=800){
        Similarity_factor++;
        } else{
         Similarity_factor--;
         }
```

Για αρχεία στα οποία μπορεί να έχουμε Στεγανογραφία τύπου FUSE ο Παράγοντας Ομοιότητας υπολογίζεται μεταξύ του ύποπτου και του αναδομημένου αρχείου από τον παρακάτω αλγόριθμο:

```
Similarity_factor=9;
Size_dif=Math.abs(<S>.size()-<P>.size());

for (int i=1;i<=9;i++){
        if(Math.abs(<S>GBLp(i) - <P>GBLp(i))==0)
                Similarity_factor++;
 }
```

Υπολογισμός Κατωφλίου Ομοιότητας

Ως ***Κατώφλι Ομοιότητας*** (*Similarity Threshold*) ορίζουμε μία σταθερά που έχει υπολογιστεί μετά από εφαρμογή της γενίκευσης του Νόμου του Benford σε μεγάλο αριθμό κανονικών και στεγανογραφημένων αρχείων. Η τιμή της σταθεράς αυτής είναι άμεσα εξαρτώμενη από τον τύπο κωδικοποίησης της εικόνας. Έτσι για παράδειγμα το MS Paint έχει διαφορετικό Κατώφλι ομοιότητας από το Photoshop 9.

Ύστερα από εκτενή μελέτη μεγάλου αριθμού εικόνων διευκρινίστηκε η ακριβής τιμή του ***Κατωφλίου Ομοιότητας*** του λογισμικού MS Paint που χρησιμοποιήθηκε στα πειράματά μας για την κωδικοποίηση της εικόνας.
Η τιμή αυτή υπολογίσθηκε κοντά στο εννιά (~9).

Για την δημιουργία των στεγανογραφημένων αρχείων, τα κανονικά αρχεία χωρίστηκαν σε τρεις διαφορετικές ομάδες των 1500 εικόνων. Για κάθε ομάδα εφαρμόστηκε διαφορετικός αλγόριθμος στεγανογραφίας από τους *JPHSWin, Camouflage και Invisible Secrets* προκειμένου να στεγανογραφηθεί η μικρότερη δυνατή πληροφορία (1kb ASCII .txt file).
Κάτω από την τιμή 9 το Ύποπτο Αρχείο που εξετάζεται μπορεί να θεωρηθεί Στεγανογραφημένο ενώ τιμές μεγαλύτερες του «9» δείχνουν «καθαρό» αρχείο.

Για την ακρίβεια οι τιμές συνοψίζονται στον παρακάτω πίνακα ανεξαρτήτως είδους στεγανογράφησης και μεγέθους εικόνας:

	Στεγανογραφημένο Αρχείο	***Μη Στεγανογραφημένο αρχείο***	***False positive***
Τιμές Παράγοντα Ομοιότητας	0-8	9	0%

Σχολιάζοντας τον παραπάνω Πίνακα:

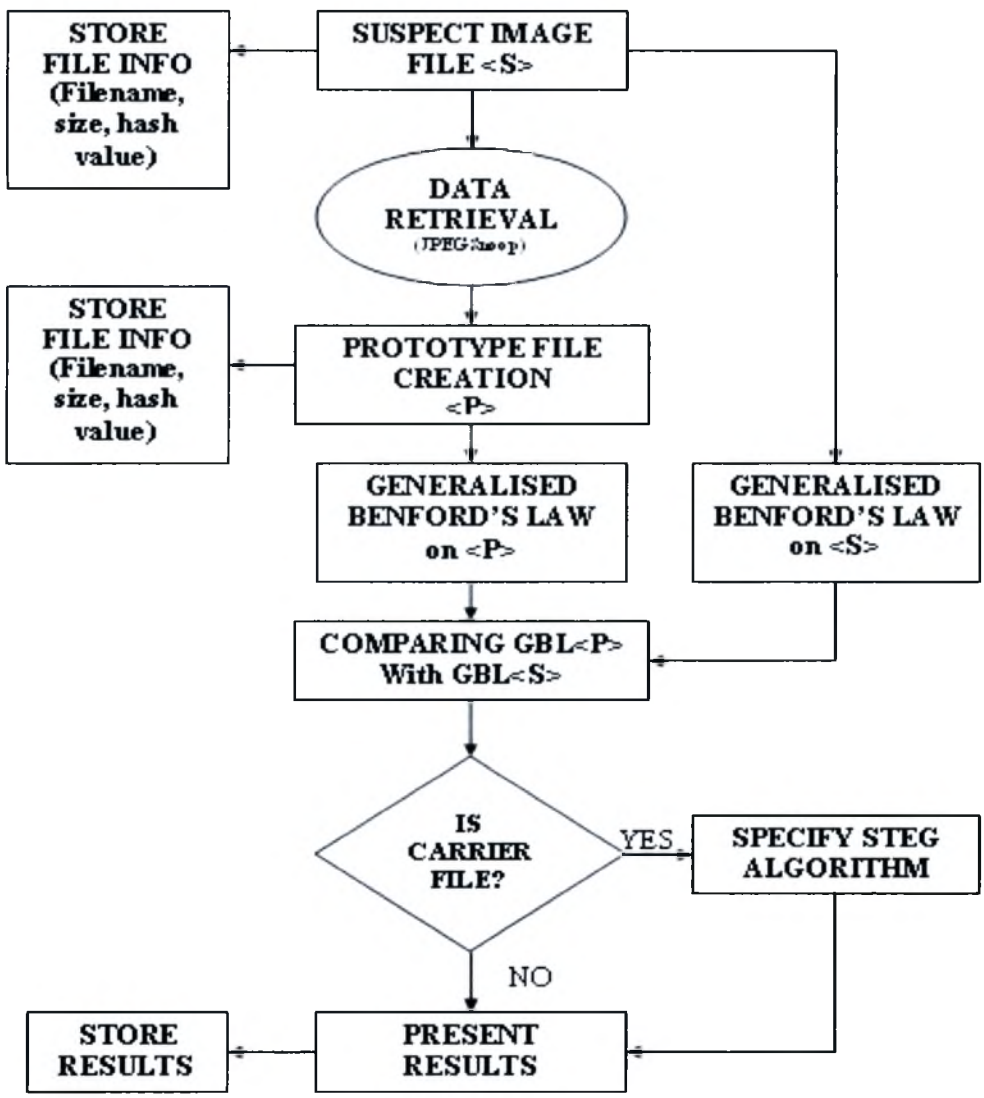
1. Δεν βρέθηκε *false positive* στεγανογραφημένο αρχείο, λόγω μεγάλης ακρίβειας της μεθόδου αναδόμησης αρχείων.
2. Τα στεγανογραφημένα αρχεία δεν ήταν ποτέ απολύτως όμοια με το αναδομημένο αρχείο γι αυτό και οι τιμές του παράγοντα ομοιότητας βρίσκονταν μεταξύ μηδέν και οχτώ.
3. Τα μη στεγανογραφημένα αρχεία που αναδομήθηκαν ήταν ακριβώς όμοια με τα αρχικά σε όλες τις δοκιμές, έτσι δικαιολογείται και η τιμή 9/9 του παράγοντα ομοιότητας.

Τα παραπάνω στοιχεία αποδεικνύουν την αποδοτικότητα και την αξιοπιστία του προτεινόμενου αλγορίθμου εντοπισμού κρυφού περιεχομένου και δεν μπορούν να συγκριθούν σε απόδοση με κανένα παρόμοιο εργαλείο Στεγανάλυσης για τους συγκεκριμένους τύπους στεγανογραφίας σε εικόνες JPEG που εντοπίζουν με 100% επιτυχία μέσω της ***Μεθόδου 2 : Independent JPEG Group (IJG) Software.***

8.1 Υλοποίηση Λογισμικού που χρησιμοποιεί την προτεινόμενη μέθοδο Στεγανάλυσης για JPEG format

Το εργαλείο που θα παρουσιαστεί, χρησιμοποιεί τον προτεινόμενο αλγόριθμο Στεγανάλυσης που παρουσιάστηκε σε προηγούμενη ενότητα σε συνδυασμό με αλγόριθμους εντοπισμού συγκεκριμένου λογισμικού (Stego-tool specific) έτσι ώστε να επιτευχθούν μεγαλύτερα ποσοστά εντοπισμού στο συντομότερο δυνατό χρονικό διάστημα.

Το εργαλείο ονομάστηκε "*Ben-4D*" , από την μέθοδο στεγανάλυσης που χρησιμοποιήθηκε για τον εντοπισμό του κρυμμένου περιεχομένου σε Ύποπτα Αρχεία. Το εργαλείο αναπτύχθηκε σε γλώσσα προγραμματισμού JAVA και διενεργεί "Stego only attack".



Εικόνα 8.1: Διάγραμμα Διαδικασιών του Εργαλείου "BEN-4D".

Stego-only attack: *Μόνο το μήνυμα κάλυψης με τα ενσωματωμένα κρυφά δεδομένα είναι διαθέσιμα για ανάλυση.*

Το παραπάνω διάγραμμα απεικονίζει τις βασικές μεθόδους / διεργασίες που υλοποιήθηκαν από το εργαλείο που αναπτύχθηκε στα πλαίσια της εργασίας και θα παρουσιαστούν αναλυτικά σε επόμενη ενότητα *(Ενότητα 8.2)*

8.2 Επιμέρους τμήματα λογισμικού

Τα τμήματα του λογισμικού εκτελούν διαφορετικές ενέργειες καθεμία από τις οποίες συμβάλει στην εξαγωγή ορθών αποτελεσμάτων και παράλληλα πληροί όλες τις αυστηρές προϋποθέσεις ενός εργαλείου Forensics. Παρακάτω σχολιάζονται τα βασικότερα τμήματα του εργαλείου.

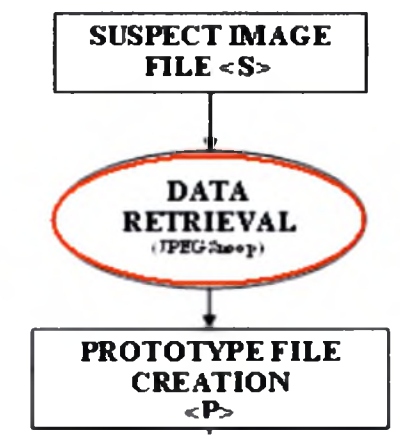
Αποθήκευση Στοιχείων Αρχείου (Store File Info)

Διαδικασία κατά την οποία βασικά χαρακτηριστικά του υπό εξέταση αρχείου καταγράφονται. Χαρακτηριστικά τέτοια θεωρούνται:

1. Το όνομα του αρχείου (filename)
2. Το Μέγεθος του αρχείου
3. Η τιμή κατακερματισμού (hash value)

Χρήση κώδικα σε Java. (Παράρτημα)

Ανάκτηση Δεδομένων (Data Retrieval)



Εικόνα 8.2: Ανάκτηση Δεδομένων

Διαδικασία κατά την οποία η ύποπτη εικόνα αναλύεται με χρήση του εργαλείου JPEGSnoop και διάφορα δεδομένα εξάγονται προκειμένου να ανατροφοδοτηθούν στο σύστημα για την δημιουργία της αναδομημένης εικόνας. Τα δεδομένα που εξάγονται είναι :

1. Ο πίνακας κβάντισης (quantization table /quality factor)
2. Η Υπογραφή Λογισμικού (“EXIF Make”/Software Signature)
3. Οι κεφαλίδες
4. Οι χρωματικές τιμές (RGB / YCC)

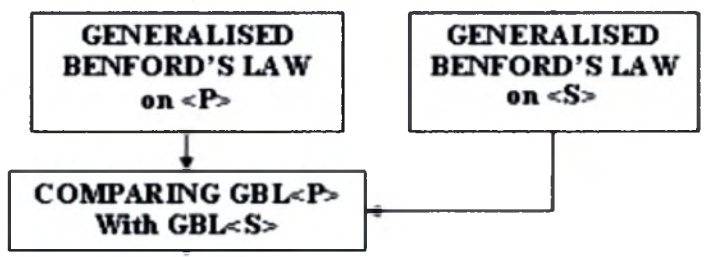
Αναδόμηση Αρχείου / Εικόνας

Αναδόμηση της εικόνας βάση της μεθοδολογίας που περιγράφεται στην *Ενότητα 7.1* με χρήση Java Advance Imaging API (JAI).
Χρήση κώδικα σε Java. (Παράρτημα)

Γενίκευση του Νόμου του Benford (Generalized Benford's Law)

Χρήση ειδική μεθόδου σε Java που υλοποιεί τον αλγόριθμο που περιγράφηκε στην Ενότητα 7.2. (Παράρτημα)

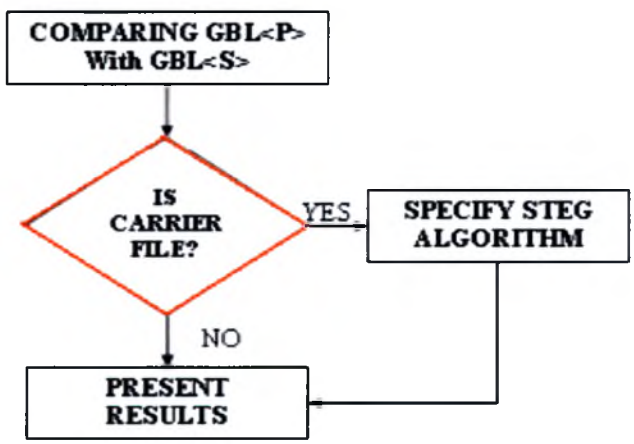
Σύγκριση



Εικόνα 8.3: Σύγκριση

Χρήση ειδική μεθόδου σε Java που υλοποιεί τον αλγόριθμο σύγκρισης που περιγράφηκε στην Ενότητα 7.2. (Παράρτημα)

Έλεγχος για ύπαρξη Στεγανογραφίας



Εικόνα 8.4: Έλεγχος για ύπαρξη Στεγανογραφίας

Υλοποίηση μεθόδου που αποφασίζει για την ύπαρξη Στεγανογραφίας βάση του ***Κατωφλίου Ομοιότητας.***

Εξακρίβωση Αλγορίθμου Στεγανογραφίας

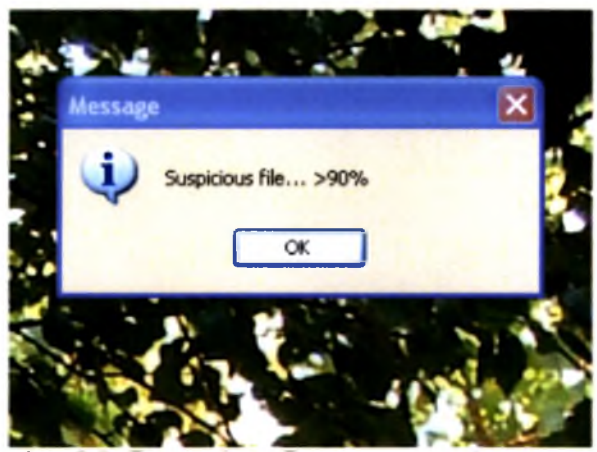
Σε περίπτωση θετικής αναγνώρισης ύπαρξης Στεγανογραφίας γίνεται προσπάθεια εντοπισμού συγκεκριμένων υπογραφών / αποτυπωμάτων του λογισμικού Στεγανογραφίας για να διευκολυνθεί η εξαγωγή της κρυμμένης πληροφορίας.

Υπογραφές / Αποτυπώματα που χρησιμοποιεί το λογισμικό προκειμένου να προσδιορίσει επακριβώς το λογισμικό Στεγανογραφίας είναι:

1. Αλλοιωμένοι ή μη κανονικοί πίνακες Huffman(JPHSWin).
2. Αισθητή διαφορά μεγέθους, Μεταξύ του Υπόπτου και του Αναδομημένου Αρχείου (Camouflage,Invisible Secrets).
3. Μη συμβατικές κεφαλίδες(Invisible Secrets).
4. Διαφοροποιημένα bits που ακολουθούν συγκεκριμένη αλληλουχία.
5. Μη συμβατικός τερματισμός αρχείου (Camouflage).

Ο συνδυασμός της Γενίκευσης του Νόμου του Benford με την χρήση Υπογραφών / Αποτυπωμάτων οδηγεί σε βελτίωση της απόδοσης εντοπισμού υπόπτων αρχείων μειώνοντας ταυτόχρονα το ποσοστό του *False – Positive* εντοπισμού Στεγανογραφημένων Αρχείων από **1,5%** σε **0.1%** σε μέσω όρο, όπως παρουσιάζεται και σε επόμενο πίνακα αναλυτικών αποτελεσμάτων.

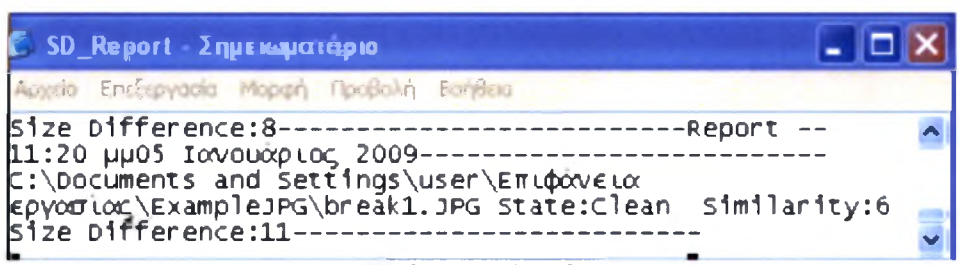
Παρουσίαση Αποτελεσμάτων



Εικόνα 8.5: Παρουσίαση Στεγανογραφημένου αρχείου

Τα αποτελέσματα από τον έλεγχο κάθε Υπόπτου αρχείου παρουσιάζονται στον χρήστη μέσα από μία εύχρηστη διεπαφή. Τόσο ένα αρχείο όσο και συλλογή αρχείων μπορούν να εξεταστούν ταυτόχρονα και τα αποτελέσματά τους παρουσιάζονται συνολικά μέσα από την χρήση Java Swing ενημερωτικών Παραθύρων μαζί με ποσοστό πιθανότητας ύπαρξης Στεγανογραφίας.

Αποθήκευση Αποτελεσμάτων



Εικόνα 8.6: Log file

Εκτός από την παρουσίαση αποτελεσμάτων, το λογισμικό δίνει την δυνατότητα αποθήκευσης των αποτελεσμάτων σε εξωτερικό αρχείο (SD_Report.txt) προκειμένου να ανατρέχει ο χρήστης για μία πιο αναλυτική ενημέρωση των ελέγχων που έχει κάνει.

8.3 Παραδείγματα χρήσης λογισμικού

Το εργαλείο BEN-4D αποτελεί ένα φιλικό προς τον χρήστη λογισμικό με γραφικό περιβάλλον (GUI) που απλοποιεί την διαδικασία της Στεγανάλυσης.
Έχει προγραμματιστεί με χρήση Java και είναι συμβατό με λογισμικό Windows 98 / Xp / Vista.

Ο χρήστης έχει την δυνατότητα να διαλέξει μεταξύ των παρακάτω επιλογών:

1. Έλεγχος μοναδικού Αρχείου JPEG.
2. Έλεγχος φακέλου που περιέχει συλλογή εικόνων JPEG.Η επιλογή αυτή είναι κατάλληλη για μεγάλο αριθμό αρχείων.
3. Καθορισμός και προβολή αρχείου καταγραφής (log file).
4. Καθορισμός τύπου ελέγχου (**fast** ή **full**).
5. Βοήθεια για την χρήση του προγράμματος.

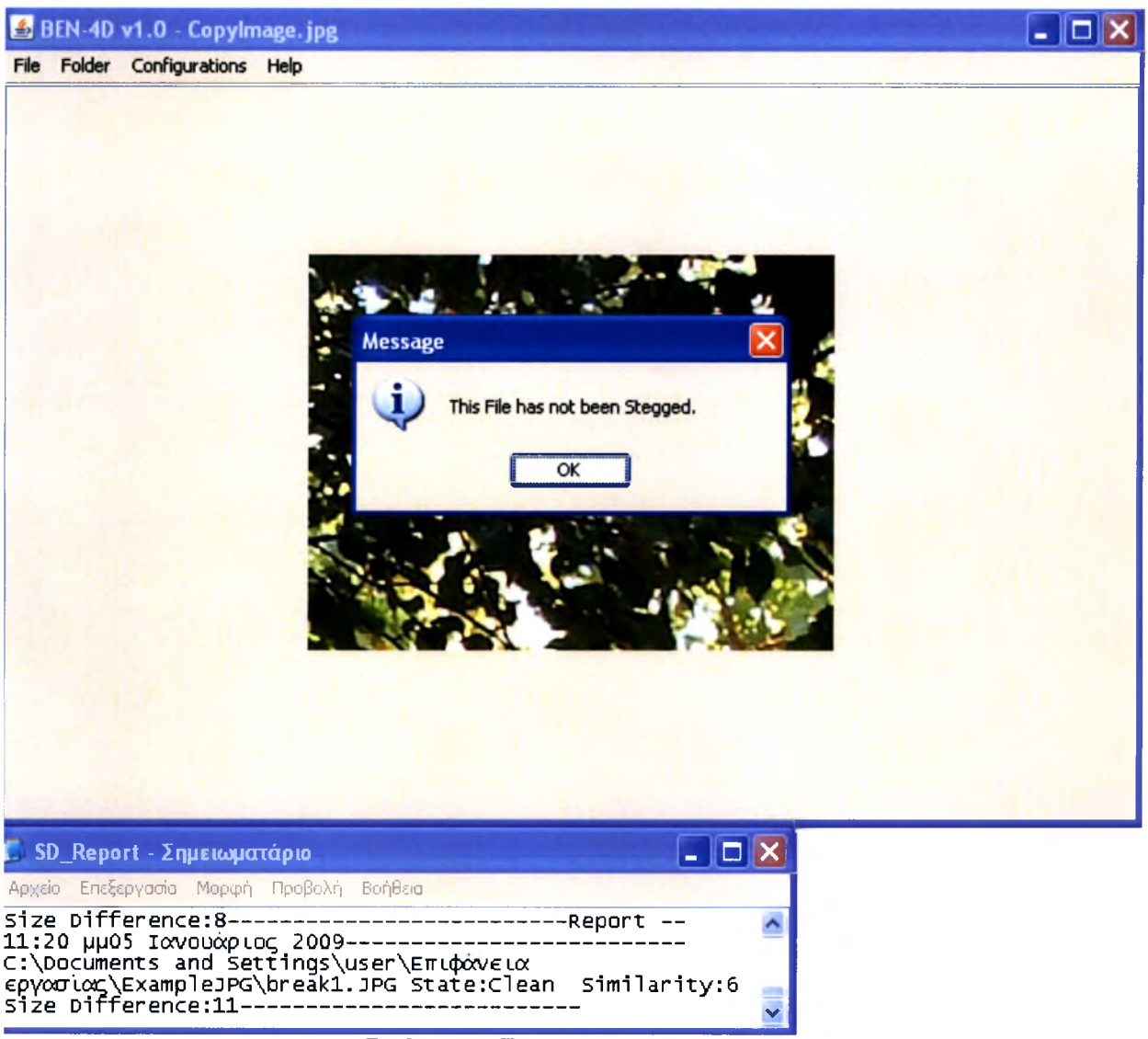
Ο τύπος ελέγχου **«fast»** εκτελεί μόνο έλεγχο με χρήση της Γενίκευσης του Νόμου του Benford ενώ ο τύπος ελέγχου **«full»** εκτελεί ταυτόχρονα και την Γενίκευση του Νόμου του Benford αλλά και ελέγχει για Υπογραφές / Αποτυπώματα εργαλείων Στεγανογραφίας. Γενικά για έλεγχο ενός ή μικρού αριθμού αρχείων ενδείκνυται χρήση ελέγχου τύπου **«full scan»** αφού οδηγεί σε πολύ αξιόπιστα αποτελέσματα. Από την άλλη όταν υπάρχει ανάγκη ελέγχου μεγάλου αριθμού εικόνων η χρήση του ελέγχου τύπου **«fast scan»** οδηγεί σε γρήγορα αποτελέσματα.

Όταν επιλέγουμε ένα αρχείο για Έλεγχο, εμφανίζεται η επιλεγμένη εικόνα στον χρήστη υπό μορφή preview.Την ίδια στιγμή ξεκινά ο υπολογισμός μέσω του προτεινόμενου αλγόριθμου Στεγανάλυσης για να εξακριβωθεί αν το υπό εξέταση αρχείο είναι Στεγανογραφημένο ή όχι.
Για κάθε αρχείο που εξετάζεται εμφανίζεται ενημερωτικό μήνυμα με το αποτέλεσμα της εξέτασης καθώς και καταγράφονται αρκετά χρήσιμα στοιχεία στο αρχείο καταγραφής όπως:

1. Ημερομηνία ελέγχου
2. Τοποθεσία αρχείου
3. Κατάσταση αρχείου (clean, suspicious x%)
4. Ομοιότητα (0-9)
5. Διαφορά μεγέθους (σε bytes)
6. Το Εργαλείο Στεγανογραφίας,

Όπως απεικονίζεται και στην Εικόνα 8.7 το αρχείο που επιλέχθηκε για έλεγχο, βρέθηκε καθαρό. Ο Παράγοντας Ομοιότητας (Similarity Factor) είχε τιμή (6) η οποία είναι μεγαλύτερη από το κατώφλι ομοιότητας *(5)*.
Ακόμη, η διαφορά μεγέθους μεταξύ του Υπόπτου και του Αναδομημένου αρχείου ήταν πολύ μικρή (11 bytes).

Η εξέταση του αρχείου έγινε με χρήση του ελέγχου τύπου **«full scan»** για μεγαλύτερη ακρίβεια και εντοπισμό συγκεκριμένου τύπου Στεγανογραφίας που χρησιμοποιήθηκε.

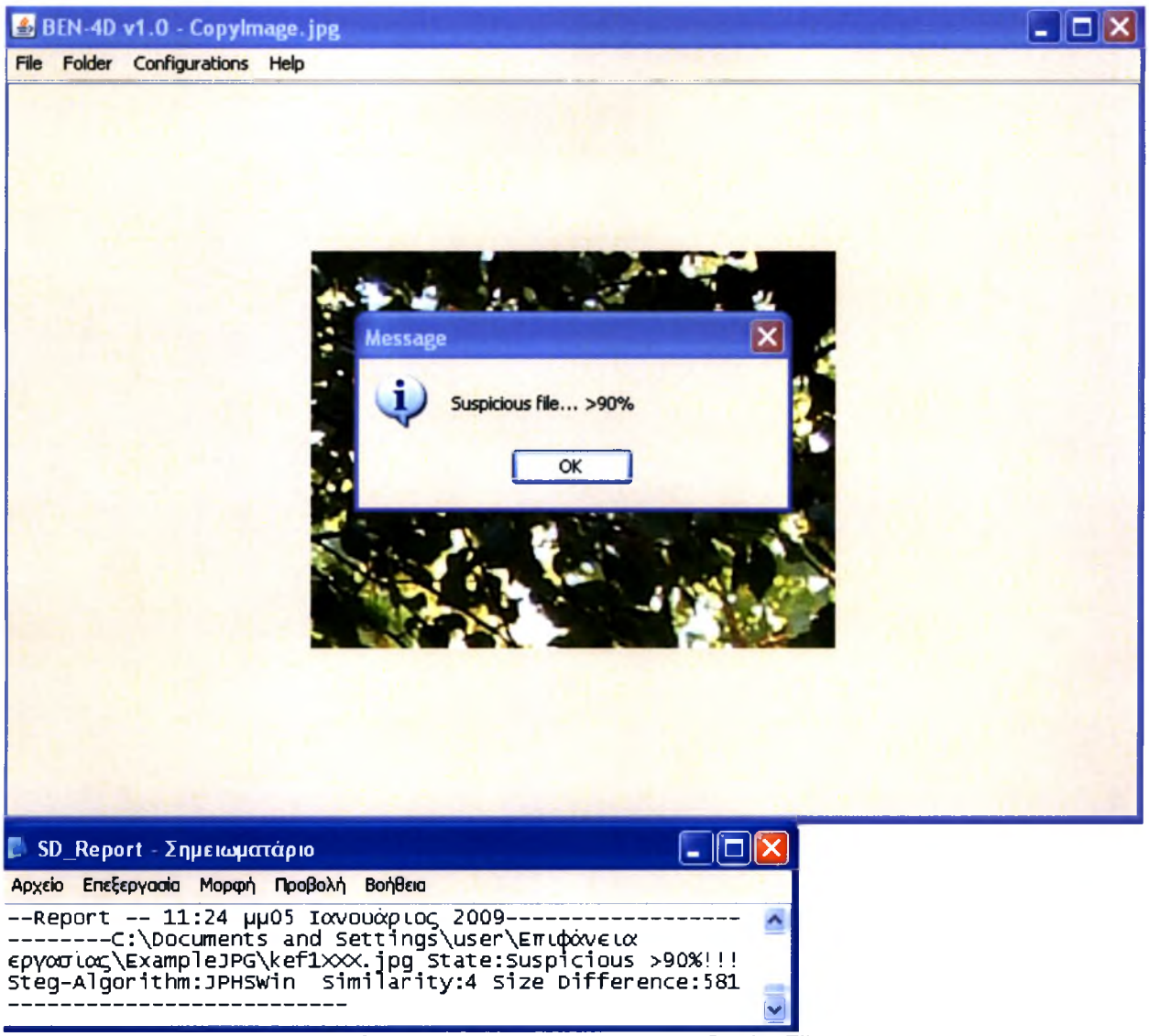


Εικόνα 8.7: Το εργαλείο "BEN-4D"

Στην Εικόνα 8.8 βλέπουμε το αρχείο που επιλέχθηκε για έλεγχο και βρέθηκε ύποπτο για Στεγανογραφία με ποσοστό Πάνω από 90%. Ο Παράγοντας Ομοιότητας (Similarity Factor) είχε τιμή (4) η οποία είναι μεγαλύτερη από το κατώφλι ομοιότητας *(5)*.

Ακόμη, η διαφορά μεγέθους μεταξύ του Υπόπτου και του Αναδομημένου αρχείου ήταν σχετικά μεγάλη (581 bytes).Ο τύπος Στεγανογραφίας εντοπίστηκε και προσδιορίστηκε ως: ***JPHSWin***

Η εξέταση του αρχείου έγινε με χρήση του ελέγχου τύπου **«full scan»** για μεγαλύτερη ακρίβεια και εντοπισμό συγκεκριμένου τύπου Στεγανογραφίας που χρησιμοποιήθηκε.



Εικόνα 8.8: "BEN-4D", εντοπισμός Carrier File.

8.4 Αποτελέσματα / Στατιστικά Στοιχεία Επιτυχημένου εντοπισμού

Για να αποδειχθεί η αποδοτικότητα, αποτελεσματικότητα και ταχύτητα του αλγορίθμου εντοπισμού κρυφού περιεχομένου με χρήση της Γενίκευσης του Νόμου του Benford, ο αλγόριθμος δοκιμάστηκε μέσω του λογισμικού που αναπτύχθηκε γι αυτό τον σκοπό. Για την δοκιμή του λογισμικού δημιουργήθηκε μια μεγάλη συλλογή εικόνων τύπου JPEG διαφορετικών αναλύσεων που Στεγανογραφήθηκαν με τρία διαφορετικά εργαλεία Στεγανογραφίας. Οι αναλύσεις / διαστάσεις που επιλέχθηκαν αποτελούν τις πιο διαδεδομένες που μπορούν να βρεθούν στο Διαδίκτυο (320x240, 600x320, 800x600).Το μέγεθος των αρχείων ποικίλει. Ακόμη καταγράφηκαν, ο χρόνος ελέγχου ανά αρχείο αλλά και το ποσοστό επιτυχίας καθώς και το ποσοστό false-positive εντοπισμού του αλγορίθμου ανά περίπτωση. Τέλος γίνεται ανάλυση των αποτελεσμάτων τόσο του τύπου ελέγχου **«fast»** που εκτελεί μόνο έλεγχο με χρήση της Γενίκευσης του Νόμου του Benford όσο και του τύπου ελέγχου **«full»** που εκτελεί ταυτόχρονα Γενίκευση του Νόμου του Benford και ελέγχει για Αποτυπώματα εργαλείων Στεγανογραφίας.

JPEG Image Dimensions	320 x 240				600 x 320				800 x 600			
File Type	Original	JPHSWin	Camouflage	Invisible Secrets	Original	JPHSWin	Camouflage	Invisible Secrets	Original	JPHSWin	Camouflage	Invisible Secrets
Number of Files	500	500	500	500	500	500	500	500	500	500	500	500
Average Size (kb)	30	34	31	31	100	104	101	101	200	205	201	201
GBL Hit rate	89%	89.1%	99.6%	99.7%	88.5%	86.7%	99.6%	99.7%	88%	82.2%	99.6%	99.7%
False Positive Steg. Detection	10%				15%				20%			
Scan time (sec) per item	~1	~2	~1	~1	~1	~2	~1	~1	~1	~2	~1	~1
Total time (min)	8.3	16.6	8.3	8.3	8.3	16.6	8.3	8.3	8.3	16.6	8.3	8.3
GBL + Sig. Hit rate	99.9%	99.8%	100%	100%	99.9%	98.1%	100%	100%	99.9%	97.4%	100%	100%
False Positive Steg. Detection	0.1%				0.1%				0.1%			
Scan time (sec) per item	~2	~3	~2	~2	~2	~3	~2	~2	~3	~4	~3	~3
Total time (min)	16.6	25	16.6	16.6	16.6	25	16.6	16.6	25	33.3	25	25

Εικόνα 8.9: Στατιστικά Επιτυχίας του εργαλείου "BEN-4D" (1kb κρυφής πληροφορίας).

Παρατηρώντας τον παραπάνω Πίνακα μπορούμε να συμπεράνουμε τα εξής:

1. Όσο μεγαλώνουν οι διαστάσεις και το μέγεθος της εικόνας τόσο πιο δύσκολος ο έγκυρος εντοπισμός Στεγανογραφίας για την πρώτη μέθοδο.
2. Αυτό δικαιολογείται λόγω του ότι όσο μεγαλύτερο είναι το μέγεθος της εικόνας τόσο μικρότερη είναι στατιστικά η αλλοίωση από την Στεγανογραφία και άρα δυσκολότερος ο εντοπισμός.
3. Το κρυφό αρχείο έχει μέγεθος 1KB (μικρότερο δυνατό μέγεθος αρχείου), άρα τα αποτελέσματα του πίνακα αποτελούν το χείριστο σενάριο εντοπισμού.
4. Όσο μεγαλώνει σε μέγεθος του κρυφού αρχείου, τόσο μεγαλώνει και το ποσοστό επιτυχούς αναγνώρισης της Στεγανογραφίας.
5. Ο χρόνος εντοπισμού διπλασιάζεται όταν χρησιμοποιείται ο τύπος ελέγχου **«full»** που εκτελεί ταυτόχρονα Γενίκευση του Νόμου του Benford και ελέγχει για Αποτυπώματα εργαλείων Στεγανογραφίας σε σχέση με τον τύπο ελέγχου **«fast»** που εκτελεί μόνο έλεγχο με χρήση της Γενίκευσης του Νόμου του Benford.

Γενικά συμπεραίνουμε ότι τα αποτελέσματα του αλγορίθμου εντοπισμού κρυφού περιεχομένου με χρήση της Γενίκευσης του Νόμου του Benford είναι ικανοποιητικά και μπορούν να λειτουργήσουν σαν έναν ασφαλή τρόπο εντοπισμού Στεγανογραφίας.
Συγκριτικά με άλλα εργαλεία / μεθόδους Στεγανάλυσης η προτεινόμενη μέθοδος είναι ανταγωνιστικά αποτελεσματική[] στην εξεύρεση Στεγανογραφημένων αρχείων για συγκεκριμένα εργαλεία Στεγανογραφίας αλλά ταυτόχρονα είναι γρήγορη πράγμα το οποίο εκλείπει από άλλα ανάλογα εργαλεία στα οποία ο χρόνος εντοπισμού είναι σαφώς μεγαλύτερος λόγω του χρόνου εκπαίδευσης που απαιτείται.

8.5 Συγκριτικά αποτελέσματα σε σχέση με άλλα εργαλεία Στεγανάλυσης

Σε αυτή την ενότητα συνοψίζονται τα συγκεντρωτικά συγκριτικά αποτελέσματα εντοπισμού του εργαλείου στεγανάλυσης BEN-4D για συγκεκριμένους τύπους Στεγανογραφίας σε σχέση με άλλα εργαλεία Στεγανάλυσης αλλά και η σύγκριση όλων των παραπάνω με την ιδανική υλοποίηση ενός εργαλείου Στεγανάλυσης χρησιμοποιώντας τέλεια αναδόμηση αρχείου (*Ενότητα 7.2, Μέθοδος 2*).

Τα εργαλεία Στεγανάλυσης που συγκρίνονται, εντοπίζουν Στεγανογραφία εικόνων JPEG και έχουν δοκιμασθεί στο ίδιο σύνολο αρχείων με αυτό που δοκιμάσθηκε το εργαλείο BEN-4D και περιγράφεται στην *Ενότητα 8.4*.

File Type	Original	JPHSWin	Camouflage	Invisible Secrets
Number of Files	1500	1500	1500	1500
DETECTION RATES				
BEN-4D (full)	99%	98%	100%	100%
STEGDETECT	99.5%	99.2%	100%	100%
STEGSPY	98%	99%	100%	100%
Μέθοδος 2: Independent JPEG Group (IJG)	100%	100%	100%	100%
FALSE POSITIVE DETECTION RATES				
BEN-4D (full)	1%	2%	0%	0%
STEGDETECT	0.5%	0.8%	0%	0%
STEGSPY	2%	1%	0%	0%
Μέθοδος 2: Independent JPEG Group (IJG)	0%	0%	0%	0%

Πίνακας Α

	ΧΡΟΝΟΣ ΕΚΤΕΛΕΣΗΣ (min)	ΕΙΔΗ / ΕΡΓΑΛΕΙΑ ΣΤΕΓΑΝΟΓΡΑΦΙΑΣ ΠΟΥ ΕΝΤΟΠΙΖΟΝΤΑΙ	TRAINING	ΕΠΕΚΤΑΣΙΜΟΤΗΤΑ
Number of Files	6000	-	-	-
BEN-4D (full)	299.6	LSB,FUSE	ΟΧΙ	ΝΑΙ
STEGDETECT	290	LSB,FUSE	ΝΑΙ	ΟΧΙ
STEGSPY	320	LSB,FUSE	ΟΧΙ	ΟΧΙ
Μέθοδος 2: Independent JPEG Group (IJG)	-	LSB,FUSE	ΟΧΙ	ΟΧΙ

Πίνακας Β

ΚΕΦΑΛΑΙΟ 9 – ΕΠΙΛΟΓΟΣ

9.1 Συμπεράσματα

Υστέρα από μία γενικευμένη παρουσίαση των βασικών όρων, αξιών αλλά και μεθοδολογιών και τεχνικών της επιστήμης της Στεγανογραφίας και Στεγανάλυσης από την οπτική της επιστήμης του Computer Forensics, η διπλωματική αυτή εργασία επικεντρώθηκε στην παρουσίαση μιας πρωτότυπης μεθόδου Στεγανάλυσης.Η μέθοδος αυτή χρησιμοποίησε την γενίκευση του νόμου του Benford για Εντοπισμό Κρυμμένου Περιεχομένου ανεξαρτήτως τύπου αρχείου.Ακολούθησε υλοποίηση της μεθόδου για εικόνες JPEG.

Έπειτα από εκτενή μελέτη των ήδη υπαρχόντων εργαλείων (λογισμικού), μεθοδολογιών και τεχνικών Στεγανάλυσης και Στεγανογραφίας μέσω της Επιστήμης του Computer Forensics, αναπτύχθηκε πρότυπο επεκτάσιμο λογισμικό το οποίο εφαρμόζει με επιτυχία την προτεινόμενη μεθοδολογία Στεγανάλυσης και τηρεί όλα τα πρότυπα εργαλείων Computer Forensics.
Παράλληλα με το λογισμικό αυτό, ο Forensic Investigator έχει στα χέρια του ένα απλό αλλά μοναδικό στο είδος του από άποψη μεθοδολογίας εργαλείο Στεγανάλυσης.

9.2 Στόχοι που επιτεύχθηκαν

Στα πλαίσια της εργασίας αυτής επιτεύχθηκαν οι παρακάτω γενικοί στόχοι:

Εξετάστηκε σε βάθος η Ηλεκτρονική Απόκρυψη πληροφορίας (Data Hiding) και η σχέση της με την Επιστήμη του Computer Forensics.Βάρος δόθηκε στην κατανόηση της φύσης της Στεγανογραφίας αλλά και της Στεγανάλυσης μέσα από ορισμούς τεχνικές αλλά και εφαρμογές.

Διενεργήθηκε εξαντλητική μελέτη και δοκιμή πολλών εργαλείων Στεγανάλυσης και Στεγανογραφίας για την παρουσίαση του «State of The Art» στον τομέα αυτό. Μέσα από την διαδικασία αυτή κατανοήθηκαν οι τεχνικές που χρησιμοποιούνται στην πράξη για την απόκρυψη αλλά και τον εντοπισμό κρυφού περιεχομένου. Εντοπίσθηκε επίσης η έλλειψη εργαλείων Στεγανάλυσης τα οποία λειτουργούν ανεξαρτήτως τύπου αρχείου αλλά και έλλειψη εργαλείων που να μην απαιτούν βήμα εκπαίδευσης για τον εντοπισμό κρυφού περιεχομένου.

Παρουσιάστηκε εκτενώς ο νόμος του Benford τόσο στην θεωρία όσο και στην πράξη μέσα από τις πολλές χρήσης του. Έμφαση δόθηκε στην χρήση του νόμου για εντοπισμό κρυφού περιεχομένου στην Επιστήμη του Computer Forensics. Εντοπίσθηκε έλλειψη λογισμικού που να εφαρμόζει τις διάφορες προτεινόμενες μεθόδους εντοπισμού κρυφού περιεχομένου.

Παρουσιάστηκε πρωτότυπη θεωρητική χρήση της γενίκευσης του νόμου του Benford για τον Εντοπισμό Κρυμμένου Περιεχομένου ανεξαρτήτως τύπου αρχείου, καθώς και χρήση της γενίκευσης του νόμου του Benford για

Εντοπισμό Κρυμμένου Περιεχομένου σε αρχεία εικόνας τύπου .JPEG και κειμένου .TXT.
Αναπτύχθηκε και παρουσιάστηκε λογισμικό που χρησιμοποιεί την προτεινόμενη μέθοδο Στεγανάλυσης με συνδυαστική χρήση γενίκευσης του νόμου του Benford και άλλων τεχνικών εντοπισμού κρυφού περιεχομένου.

Τέλος παρουσιάστηκαν συγκριτικά αποτελέσματα του προτεινόμενου λογισμικού σε σχέση με ήδη υπάρχοντα προγράμματα εντοπισμού κρυφού περιεχομένου σε εικόνες JPEG. Μέσα από τα συγκριτικά στοιχεία διαφαίνεται η ανωτερότητα τόσο της μεθόδου όσο και του εργαλείου εντοπισμού κρυφού περιεχομένου που αναπτύχθηκε και δεν έχει να ζηλέψει τίποτα σε σχέση με άλλα συστήματα εντοπισμού κρυμμένης πληροφορίας.

9.3 Μελλοντικές Εφαρμογές

Σαν βελτίωση στην θεωρητική εργασία που έχει γίνει πρέπει να μελετηθούν ζητήματα όπως:

1. Η απώλεια δεδομένων που παρατηρείται κατά την διαδικασία αναδόμησης των αρχείων και η οποία μπορεί να ελαχιστοποιηθεί με βελτίωση της τεχνικής Αναδόμησης μέσω lossless transcoding[28].
2. Η χρήση πιο σύνθετων εργαλείων Στεγανογραφίας όπως το Outguess ή το F5, πρέπει να εντοπίζεται με επιτυχία. Αυτοί οι αλγόριθμοι επανακωδικοποιούν την αρχική εικόνα για να παράγουν την εικόνα φορέα. Ο κάθε ένας αλγόριθμος από αυτούς έχει δικό του μοντέλο κωδικοποίησης με αποτέλεσμα να παράγει μια μοναδική υπογραφή (“EXIF Make” ή Software Signature). Αυτό το σημείο μπορεί να αποτελεί την αρχή για βελτιώσεις της ήδη υπάρχουσας εργασίας.

Σαν επέκταση στο εργαλείο που σχεδιάστηκε, πρέπει να μελετηθούν ζητήματα όπως:

1. Ποιο πολλά εργαλεία Στεγανογραφίας θα πρέπει να μπορούν να εντοπιστούν.
2. Ποιο πολλοί Αλγόριθμοι κωδικοποίησης JPEG πρέπει να αναγνωρίζονται και να υποστηρίζονται. Το Threshold για κάθε έναν από αυτούς τους αλγορίθμους πρέπει να υπολογισθεί.
3. Ποιο πολλά είδη εικόνων όπως BMP,GIF πρέπει να υποστηρίζονται για ανάλυση και εντοπισμό κρυφού περιεχομένου.
4. Άλλα ήδη αρχείων πρέπει να υποστηρίζονται για ανάλυση και εντοπισμό κρυφού περιεχομένου.
5. Μέθοδος εξαγωγής της κρυφής πληροφορίας πρέπει να υλοποιηθεί .

9.4 Επίλογος

Με την αυξημένη χρήση των ηλεκτρονικών υπολογιστών στην σημερινή κοινωνία η ανάγκη ύπαρξης και βελτίωσης της επιστήμης της Σήμανσης των Ηλεκτρονικών Υπολογιστών γίνεται ολοένα και πιο αισθητή. Η χρησιμότητά της δεν περιορίζεται μόνο στην εξαγωγή ακεραίων και αδιάβλητων δικαστικών ηλεκτρονικών στοιχείων, αλλά κατά κύριο λόγω διευκολύνει την επίλυση

καθημερινών ζητημάτων που ολοένα και επηρεάζουν τις ζωές περισσοτέρων ανθρώπων όπως η διακίνηση κρυμμένης πληροφορίας. Βελτιώνοντας την μέθοδο Στεγανάλυσης αλλά και το λογισμικό που την υλοποιεί, ένα νέο εργαλείο βρίσκεται στα χέρια των ειδικών ώστε να βοηθήσει τόσο στον γρήγορο όσο και αποδοτικό εντοπισμό κρυφού περιεχομένου.

Όσο πιο πολύ λοιπόν εξελίσσεται και ενημερώνεται η επιστήμη του Computer Forensics για τις νέες απειλές και νέα προβλήματα τόσο πιο ασφαλής και προστατευμένος είναι ο απλός χρήστης.

Benford function

```
double[]  benford (byte[] b) throws IOException{

            int max=0;
//αρχικοποίηση των τιμών της κατανομής Benford για κάθε ψηφίο
            double benford[] = new double[9];

            benford[0]=30.103;
            benford[1]=17.609;
            benford[2]=12.494;
            benford[3]=9.691;
            benford[4]=7.918;
            benford[5]=6.695;
            benford[6]=5.799;
            benford[7]=5.115;
            benford[8]=4.576;

//
            for (int i = 0; i < b.length; i++) {

                  if(Math.abs(b[i])>max){

                        max = Math.abs(b[i]);

                  }
            }


            int distribution[]=new int[max];
// αρχικοποίηση κατανομής με τιμή 0 για κάθε ψηφίο
            for(int k = 0;k<max;k++){

                  distribution[k] = 0;

            }

//Υπολογισμός συχνότητας εμφάνισης του πρώτου στοιχείου

            for (int i = 0; i < b.length; i++) {

                  for(int l = 0;l<max;l++){

                  if(Math.abs(b[i])!=0){

distribution[Math.abs(b[i])-1]=distribution[Math.abs(b[i])-1]+1;

                        }
                  }
            }

//Υπολογισμός Benford κατανομής

      double mybenford1[] = new double[9];
      double benfordstats1[] = new double [9];
```

```
      for(int m = 0; m<max ;m++){

            if(distribution[m]!=0){

                  try{

                  String str = "" + distribution[m];
                  String str2= Character.toString(str.charAt(0));
                  int i = Integer.parseInt(str2);
                  mybenford1[i-1]=mybenford1[i-1]+1;

                        } catch (NumberFormatException nfe)
{System.out.println("NumberFormatException: " + nfe.getMessage());
                        }
                  }
            }

            for (int var=0;var<9;var++){
                  benfordstats1[var] =  (mybenford1[var]*100) / max;
            }

//Καταγραφή αποτελεσμάτων
FileOutputStream fout2 = new FileOutputStream ("Stats.txt");

new PrintStream(fout2).println("Stats 1st\tBenford\tDifference");
new PrintStream(fout2).println("--------------------------");

for (int var=0;var<9;var++){

      new PrintStream(fout2).println( String.format("%.2f",
benfordstats1[var])+"\t"+benford[var]+"\t"+String.format("%.2f",Math.
abs(benfordstats1[var]-benford[var])));

            }

System.out.println("Mean
Deviation"+String.format("%.2f",findDeviation(b)));
            fout2.close();

//επιστροφή πίνακα  με τις ανακτημένες τιμές
            return benfordstats1;
      }
```

Βιβλία

[1] Investigator's Guide to Steganography by Greg Kipper, ISBN:0849324335, Auerbach Publications © 2004
[2] R. C. Gonzalez, R.E. Woods,"Digital Image Processing", 2nd Edition.

Άρθρα

[3] Benford, F. 1938. The law of anomalous numbers. Proceedings of the American Philosophical Society 78:551-572.
[4] Benford's law. Journal of the American Taxation Association 18:72-91.
[5] Hill, T. 1995. Base-invariance implies Benford's law. Proceedings of the American Mathematical Society 12:887-895.
[6] Hill, T. 1996. A statistical derivation of the significant-digit law. Statistical Science 10:354-363.
[7] Ley, E. 1996. On the peculiar distribution of the U.S. stock indices digits. American Statistician 50:311-313.
[8] Nigrini, M. 1996. A tax payer compliance application of
[9] Raimi, R. 1969. The peculiar distribution of first digits. Scientific American (December) pp. 109-119.
[10] Benford's Law in Image Proccessing *Fernando P´erez-Gonz´alez, Greg L. Heileman, and Chaouki T. Abdallah*
[11] Chen Ming Zhang Ru Niu Xinxin Yang Yixian, "Analysis of Current Steganography Tools: Classifications & Features", in International Conference on Intelligent Information Hiding and Multimedia pp. 384-387, 2006
[12] N. Provos and P. Honeyman, "Hide and seek: An introduction to steganography," *IEEE Security & Privacy Magazine*, May-June 2003.
[13] R. Chandramouli, M. Kharrazi, and N. Memon, "Image steganography and steganalysis: concepts and practice." in *Proc. Digital Watermarking, Second International Workshop, IWDW 2003*, ser. Lecture Notes in Computer Science, T. Kalker, I. J. Cox, and Y. M. Ro, Eds., vol. 2939. Seoul, Korea: Springer, October 20-22 2003, pp. 35–49, ISBN: 3-540-21061-X.
[14] N. Provos, "Defending against statistical steganalysis." in *10th USENIX Security Symposium*, Washington, DC, USA, 2001.
[15] A. Westfeld, "F5-a steganographic algorithm." in *Proc. Information Hiding, 4th International Workshop, IHW 2001*, ser. Lecture Notes in Computer Science, I. Moskowitz, Ed., vol. 2137. Pittsburgh, PA, USA: Springer, April 25-27 2001, pp. 289–302, ISBN: 3-540-42733-3.
[16] A. Latham, "Steganography: JPHIDE AND JPSEEK," 1999, http://linux01.gwdg.de/_alatham/stego.html.
[17] J. Barbier, E. Filiol, and K. Mayoura, "Universal JPEG steganalysis in the compressed frequency domain." in *Proc.*

Digital Watermarking, 5th International Workshop, IWDW 2006, ser. Lecture Notes in Computer Science, Y. Q. Shi and B. Jeon, Eds., vol. 4283. Jeju Island, Korea: Springer, November 8-10 2006, pp. 253–267.

[18] G. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, 1991.

[19] C. Brown and B. Shepherd, *Graphics File Formats, reference and guide*. Manning, 1995.

[20] I. Avicibas¸, N. Memon, and B. Sankur, "Steganalysis based on image quality metrics." in *Proc. SPIE, Security and Watermarking of Multimedia Contents III*, P. W. Wong and E. J. Delp, Eds., vol. 4314, 2001, pp. 523–531.

[21] H. Farid, "Detecting hidden messages using higher-order statistical models." in *ICIP (2)*, 2002, pp. 905–908.

[22] S. Lyu and H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines." in *Proc. Information Hiding, 5th International Workshop, IH 2002*, ser. Lecture Notes in Computer Science, vol. 2578. Noordwijkerhout, The Netherlands: Springer, October 7-9 2002, pp. 340–354, ISBN: 3-540-00421-1.

[23] ——, "Steganalysis using higher-order image statistics." *IEEE Transactions on Information Forensics and Security*, vol. 1, 2006.

[24] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems." in *Proc. Information Hiding, Third International Workshop, IH'99*, ser. Lecture Notes in Computer Science, A. Pfitzmann, Ed., vol. 1768. Dresden, Germany: Springer, September 29 - October 1 1999, pp. 61–76, ISBN: 3-540-67182-X.

[25] J. Fridrich, M. Goljan, and D. Hogea, "Steganalysis of JPEG images: breaking the f5 algorithm." in *Proc. Information Hiding, 5th International Workshop, IH 2002*, ser. Lecture Notes in Computer Science, vol. 2578. Noordwijkerhout, The Netherlands: Springer, October 7-9 2002, pp. 310–323, ISBN: 3-540-00421-1.

[26] "New methodology for breaking steganographic techniques for JPEGs." in *Proc. SPIE, Security and Watermarking of Multimedia Contents V*, Santa Clara, CA,USA, January 2003, pp. 143–155.

[27] Multi-Class Detector of Current Steganographic Methods for JPEG Format Tom´aˇs Pevn´y and Jessica Fridrich, IEEE member

[28] Lossless JPEG transcoding, Daniel Sanchez, JPEG transcoding ECE 533 Project Report Fall 2006

[29] Independent JPEG Group - http://www.ijg.org/: Their widely used software package includes an utility that performs lossless transcoding, *jpegtran.*

[30] A generalized Benford's law for JPEG coefficients and its applications in image forensics Dongdong Fu*a, Yun Q. Shi*a, Wei Sub aDept. of Electrical and Computer Engineering, New Jersey Institute of Technology Newark.

[31] Enhancement of JPEG-Compressed Images by Re-application of JPEG Aria Nosratinia (aria@utdallas.edu) Department of Electrical Engineering, University of Texas at Dallas, Richardson

[32] Image and the Benford's Law,J.M Jolion,Laboratoire Reconnaisance de Formes et Vision, INSA Lyon, France

[33] A Survey of Steganographic and Steganalytic Tools for the Digital Forensic Investigator Pedram Hayati, Vidyasagar Potdar, and Elizabeth Chang Institute for Advanced Studies in Basic Science of Zanjan, Iran

WWW home page: http:// http://www.iasbs.ac.ir/students/pedram,Digital Ecosystems and Business Intelligence Institute, Curtin Business School, Curtin University of Technology, Perth, Australia
[34] An Analysis of Terrorist Groups' Potential Use of Electronic Steganography, Stephen Lau, Sans InfoSec Reading Room, Feb 2003
[35] Analysis of LSB Based Image Steganography Techniques , R. Chandramouli, Nasir Memon, Proc. IEEE ICIP, 2001
[36] Applications for Data Hiding, W. Bender, IBM Systems Journal, 2000
[37] Detecting Covert Communications on the Internet: Some Challenges and Solutions, R. Chandramouli, Digital Forensic Research Workshop (DFRWS), 2003
[38] Detection of Hiding in the Least Significant Bit, O. Dabeer et al, Conference on Information Sciences and Systems, Mar 2003
[39] Detecting Steganographic Content On The Internet , Presentation, Niels Provos, Feb 2002

Λογισμικό

[40] JP Hide and Seek, http://linux01.gwdg.de/~alatham/stego.html
[41] Jsteg Jpeg, http://www.nic.funet.fi/pub/crypt/steganography/
[42] OutGuess, http://www.outguess.org/d ownload.php
[43] PGM Stealth, ftp://ftp.funet.fi/pub/crypt/steganography/
[44] Steghide, http://steghide.sourceforge.net/
[45] Crypto123, http://www.kellysoftware.com/software/Crypto123.asp
[46] Hermetic Stego, http ://www.hermetic.ch/hst/hst.htm
[47] IBM DLS, http://www.research.ibm.com/image_apps/commerce.html
[48] Invisible Secrets, http://www.neo-bytesolutions.com/
[49] Info Stego, http://www.antiy.net/infostego/
[50] Syscop, http://www.mediasec.com/html/en/products_services/syscop.htm
[51] http://www.datamark-tech.com
[52] Cloak, http://www.softslist.com/download -9-5-16609.html
[53] Contraband Hell, http://jthz.com/puter/
[54] Dound, http://evidence-eliminators.co.uk/dou nd.htm
[55] Camouflage, http://camouflage.unfiction.com/
[56] Hide and Seek, ftp://ftp.funet.fi/pub/crypt/steganography/hdsk41.zip
[57] InThePicture, http://www.guillermito2.net/stegano/inthepicture/index.html
[58] S-Tools, ftp://ftp.fun et.fi/pub/crypt/mirrors/idea.sec.dsi.unimi.it/code/
[59] Jpegx, http://www.leetupload.com/dbindex2/index.php?dir=Win32/
[60] Steganos, http://www.steganography.com/
[61] BMP Secrets, http://www.pworlds.com/products/bmp-secrets.phtml
[62] StegoTif, http://www.geocities.com/SiliconValley/9210
[63] ScramDisk, http://www.scramdisk.clara.net/
[64] StegoWav, http://www.geocities.com/SiliconValley/9210/
[65] Hide4PGP, http://www.heinz-repp.onlinehome.de/Hide4PGP.htm
[66] Paranoid, ftp://ftp .hacktic.nl/pub/crypto/macintosh/
[67] PGPn123, http://www.securityfocus.com/tools/1435
[68] Nicetext, http://www.nicetext.com/
[69] Snow, http://www.darkside.com.au/snow/
[70] Tex to, ftp://ftp.funet.fi/pub/crypt/steganography
[71] Sam's Big Play Maker, http://www.scramdisk.clara.net/play/playmaker.html
[72] Steganosaurus, http://www.fourmilab.ch/stego /
[73] FFEncode, http://www.burks.de/stegano/ffencode.html
[74] Mimic, http://www.nic.funet.fi/pub/crypt/old/mimic/Mimic-Manual.txt
[75] wbStego http://members.xo om.com/wbailer/wbstego/ind ex.htm

[76] Spam Mimic, http://www.spammimic.com/
[77] Secret Space, http://www.soft14.com/Utilities_and_Hardware/Security_and_Encryption/SecretSpace_1388_Review.html
[78] MergeStreams, http://www.ntkernel.co m/w&p.p hp?id=23
[79] Easy File & Folder Protector, http://www.softstack.com/fileprotpro.html
[80] Invisible Files, 2000 http://www.downloadstock.info/Invisible-Files-2000-Pro-60_de10671.html
[82] Magic Folders, h ttp://www.pc-magic.com/
[83] Dark Files, http://www.redsofts.com/soft/494/36867/Dark_Files.html
[84] bProtected, 2000 http://www.clasys.com/b_protected.html
[85] BuryBury, h ttp://www.winsite.com/bin/Info?1000000 034 624
[86] StegFS, http://www.mcdonald.org.uk/StegFS/
[87] Folder Guard Jr, http://www.winability.com/folderguard/
[88] Dmagic, ftp://ftp.elf.stuba.sk/pub/pc/security
[89] Hide Folders ,http://www.fspro.net/
[90] Paranoid(File System Steganography), http://sac-ftp.externet.hu/security10.html
[91] GZSteg, http://www.funet.fi/pub/crypt/steganography/
[92] KPK file, http://www.kpkfile.com
[93] Hiderman, http://www.alnini.com/Hiderman/dt-8782.html
[94] Hydan, http://www.crazyboy.com/hydan/
[95] Covert.tcp, http://www.firstmonday.org/issues/issue2_5/rowland/
[96] 2Mosaic, http ://www.petitcolas.net/fabien/watermarking/2mosaic/index.html
[97] StirMark Benchmark, http://www.petitcolas.net/fabien/watermarking/stirmark/index.html
[98] Phototile, http://www.pcadvisor.co.uk/download s/index.cfm?categoryID=1490&itemID=7895
[99] Steganography Analyzer Real-Time Scanner, http://www.sarc-wv.com/stegalyzerrts.aspx
[100] StegBreak, http://www.outguess.org/download.php
[101] StegDetect, http://www.outguess.org/detection.php
[102] StegSpy, http://www.spy-hunter.com/
[103] Stego-Suite, http://www.000.shoppingcartsplus.com/catalog/item/4170630/4050552.htm
[104] Windows File Analyzer, http://mitec.cz/wfa.html
[105] FileAlyzer, http://www.safer-networking.org/en/filealyzer/index.html
[106] BlindSide, http://www.cs.bath.ac.uk/~jpc/blindside
[107] Camera Shy, http://hack tivismo.co m/projects/index.php
[108] dc-Steganograph, http://members.tripod.com/~Nikola_Injac/stegano/
[109] F5, http://wwwrn.inf.tu-dresden.de/~westfeld/f5.html
[110] Gif Shuffle, http://www.darkside.com.au/gifshuffle
[111] Hide4PGP, http://www.heinz-repp.onlinehome.de/Hide4PGP.htm

Υπερσύνδεσμοι

[112] Computer Forensics, http://en.wikipedia.org/wiki/Computer_forensics
[113] Steganography, http://en.wikipedia.org/wiki/Steganography
[114] Steganalysis,http://en.wikipedia.org/wiki/Steganalysis
[115] Benford's Law,http://en.wikipedia.org/wiki/Benford%27s_law
[116] Hello kitty Incident report, http://afp.google.com/article/ALeqM5ieuIvbrvmfofmOt8o0YfXzbysVuQ
[117] Electronic Crime Statistics, http://www.securitystats.com/infosec.html
[118] Digital Watermarking, http://en.wikipedia.org/wiki/Digital_watermarking
[119] Jpegsnoop, http://www.impulseadventure.com/photo/jpeg-snoop.html