# Steganography and steganalysis in criminology

Mp
magdap7@gazeta.pl

# Introduction

- Introduction to steganography
- Branches of criminology
- Steganographic programs
- Computer Forensics Tools
- Steganalytic tools
- Conclusions

# Branches of criminology

- Fingerprint identification
- Hand-writing recognition
- Speech recognition
- Audio authentication
- Computer forensic methods
- Many other concerning biology and chemistry beyond this presentation

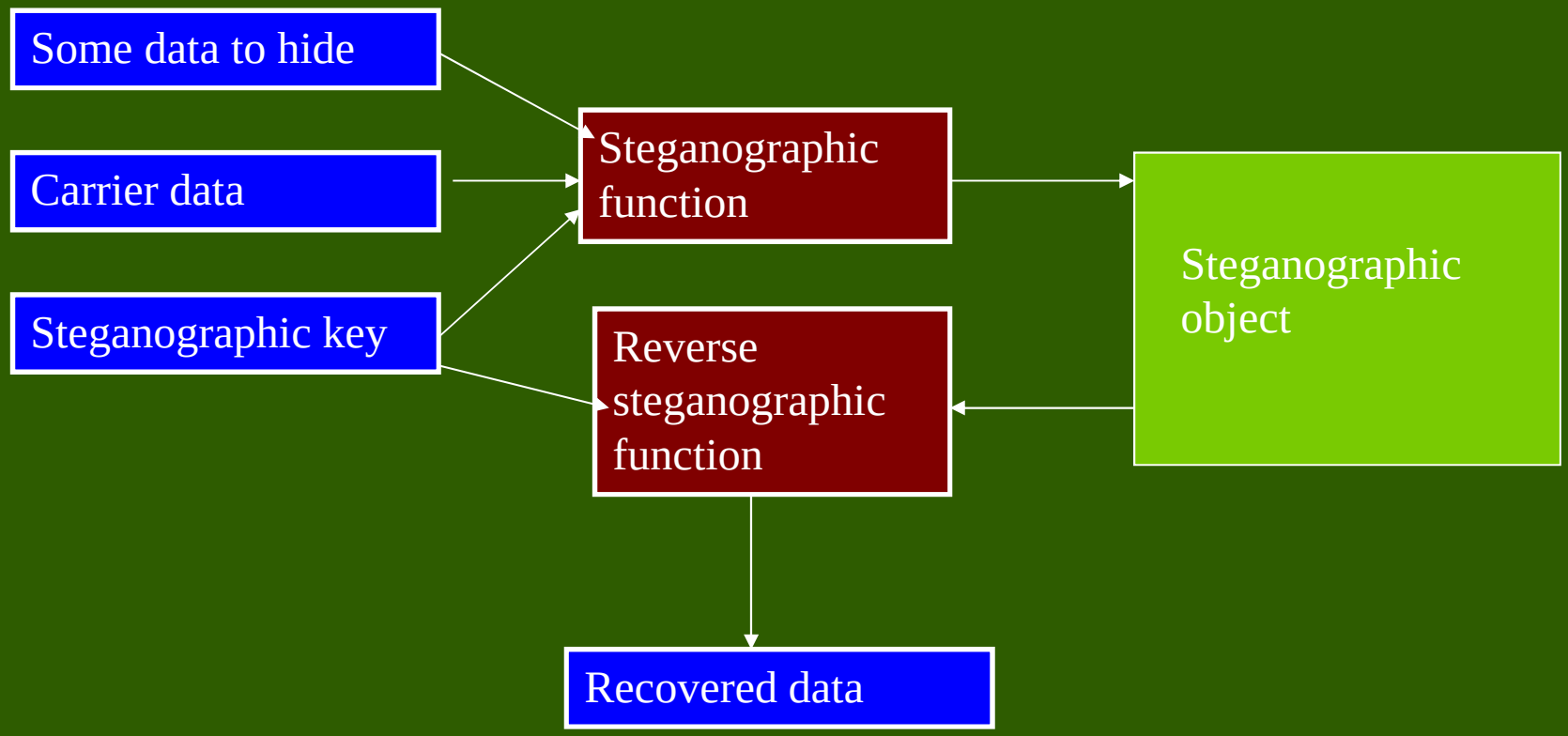
# Issues in computer forensic

- Forbidden data, photos, films
- Malicious scripts
- Illegal access and modification of data
- Violation of Intellectual property

# What steganography is

- The art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message
- The word *steganography* is of Greek origin and means "concealed writing" from the Greek words *steganos* (στεγανός) meaning "covered or protected", and *graphei* (γραφή) meaning "writing".

# Data embedding security scheme



# Domains of data hiding

- Physical {wax tablets, secret ink}
- Digital {text, image, audio, video}
- Network {packets}
- Printed {letter size, spacing}

# Domains of data embedding

- Unused areas of carrier objects
- Headers or tails of carrier objects
- Used but not significant areas
  - Time or spatial domain
  - Frequency domain
- Noisy parts of carrier objects
- Outside the audibility thresholds
- LSB method
- Indexed palette of colors
- Spread data evenly with a key
- Choose the best area for data hiding

# What steganography has to do in the art of criminology

- Data hiding and hidden data detection
- Watermarking
  - to protect property
  - to discover inconsistencies in audio files
- Pattern recognition
- Similar techniques of analysis

# Steganography and watermarking

- Inaudibility, as little as possible loss of audio quality
- Robustness, the algorithm should be robust against various attacks for malicious users
- Statistical invisibility, the algorithm should prevent unauthorized watermark detection/removal or alteration
- Similar compression characteristics with the original signal
- No original data is needed to recognize the watermark

# Approaches to fingerprint recognition

- Pattern-based (Image-based) algorithms
  - compare the basic fingerprint patterns (arch, whorl, and loop) between a previously stored template and a candidate fingerprint
- Minutia-based algorithms
  - compare several minutia points (ridge ending, bifurcation, and short ridge) extracted from the original image stored in a template with those extracted from a candidate fingerprint

# Selection of an optimization technique

- Exhaustive techniques (random walk, depth first, breadth first, enumerative)
- Calculus-based techniques (gradient methods, solving systems of equations)
- Partial knowledge techniques (hill climbing, beam search, best first, branch and bound, dynamic programming)
- Knowledge-based techniques (production rule systems, heuristic methods
- Hierarchical techniques: Generally, a coarse resolution employed to find a narrow range of the solution, then using a fine resolution in the narrow range search the optimal solution

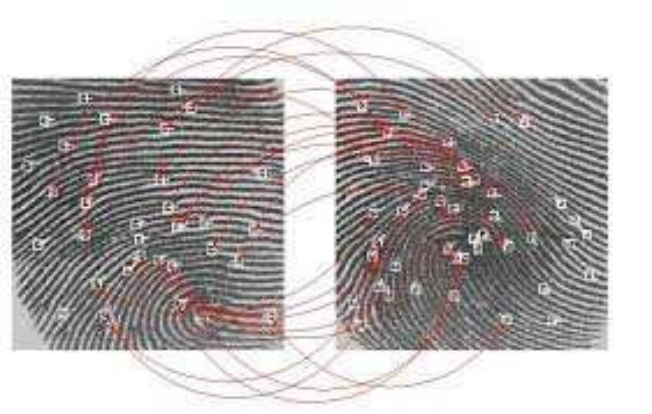
# Model for fingerprint comparison

- The optimized transformation
- The fitness function
- Genetic algorithm used to estimate the optimized transformation
- The phenomena – only small fraction of possibilities is calculated to prove the hypothesis

# Genetic approach to pattern recognition

- a vector values for optimization is represented as a chromosome (genotype)
- each chromosome consists of a set of genes (values bits or bytes), they are grouped into segments
- starting population is a subset of all possible genotypes - random choice of individuals
- in each iteration a new population is created by making small changes in the parent population

# Example



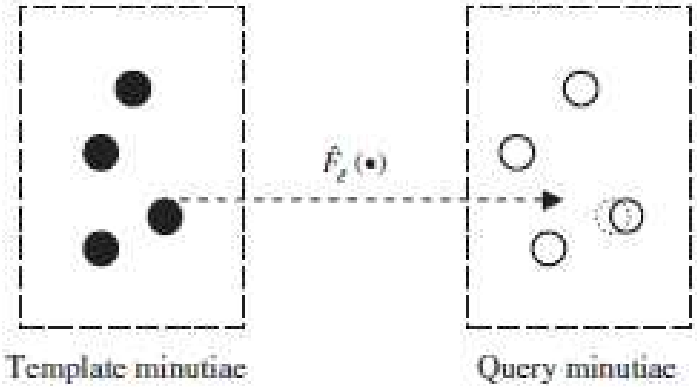



Fig. 3. Illustration of $\hat{F}_g(\bullet)$.

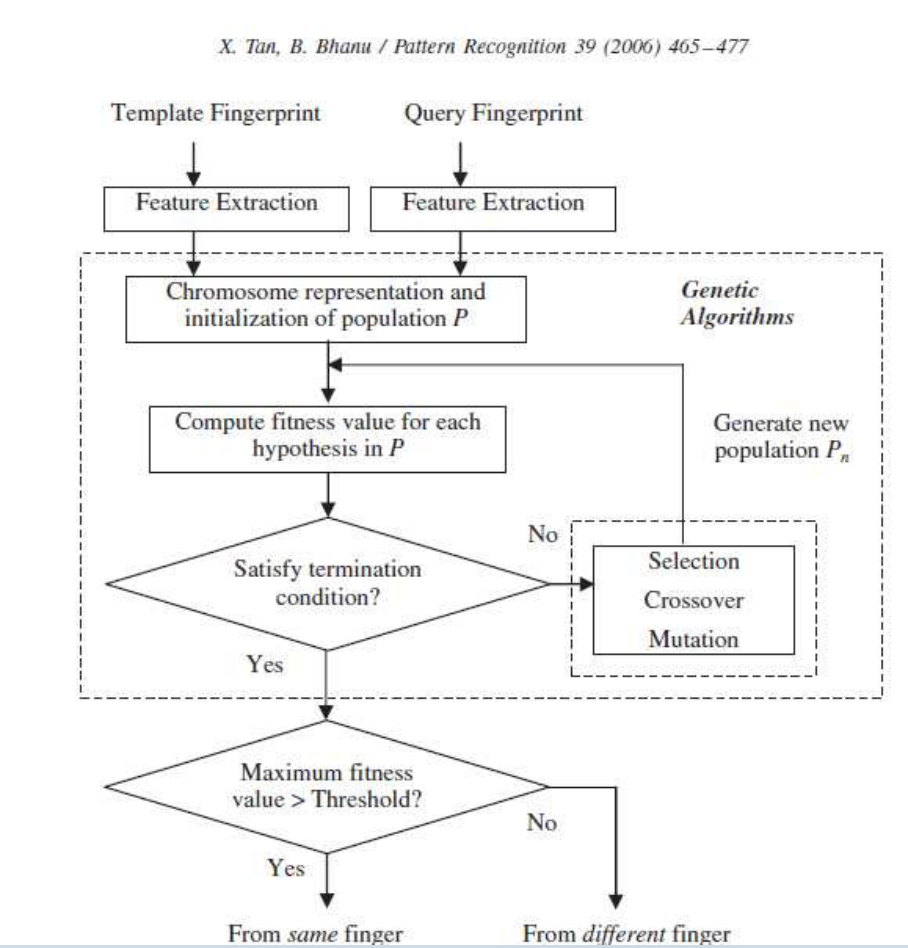
$j = 1, 2, 3, \ldots, N$. Let

$$d_j = \min_k \left\{ \left| \hat{F}\left( \begin{bmatrix} x_{j,1} \\ x_{j,2} \end{bmatrix} \right) - \begin{bmatrix} y_{k,1} \\ y_{k,2} \end{bmatrix} \right| \right\}$$

- Fitness function to be optimized

# Generating new population

- crossing over segments from parents
- random permutation of segment
- excluding chromosomes, which are
  - already created (repeating)
  - have the fitness value under the desired threshold (outside the desired range)
- sorting the population in decreasing order by the fitness values
- selection of the next generation of descendants



# Approaches to watermark creation

- Any bit-sequence may be seen under two different views:
  - 1) *Syntactically*, i.e. how it looks like as a sequence of 0's and 1's. Then the sequence's characteristics and properties are determined simply by the pattern of 0's and 1's.
  - 2) *Semantically*, i.e. whether in fact, it represents *by design* another entity/object converted into the bit-sequence under the action of a suitable encoding. This time the sequence, in addition to its syntactic characteristics, may also be seen as possessing characteristics and properties *inherited* from the entity/object from which it resulted.

# Embedding data in mp3 files

- Audio file is divided into frames
- In each frame the Scalefactors are the values of sound amplitude for a given frequency
- Differences between adjacent scalefacors are calculated
- The watermark pattern is embedded in these areas where the changes would cause the least loss of quality

# Steganography tools

| | |
|---|---|
| Contraband | BMP |
| F5.jar | JPEG |
| MP3Stego | MP3 |
| OpenStego | BMP,PNG |
| Invisible Secrets | BMP,JPEG |
| S-Tools | BMP,GIF,WAV |
| VSL | BMP,PNG,JPG,TIFF |

# The types of attack on steganographic algorithms

- Known program and unknown algorithm
- Chosen message and series of chosen carrier files
- Unknown steganographic program and a single suspected file

# Simple methods to detect steganography

- File signature
- Steganographic fingerprint
- Statistical anomaly
- Brute force attack (time and space consuming)

# Hexadecimal Preview

- HEX Viewer
- Jhead.exe

NIJ-ddtooltest.pdf

| Offset | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 00000000 | 25 | 50 | 44 | 46 | 2D | 31 | 2E | 33 | 0D | 25 | E2 | E3 | CF | D3 | 0D | 0A | %PDF-1.3.%âãÏÓ.. |
| 00000010 | 32 | 35 | 32 | 20 | 30 | 20 | 6F | 62 | 6A | 0D | 3C | 3C | 20 | 0D | 2F | 4C | 252 0 obj.<< ./L |

hackingcat.jpg

| Offset | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 00000000 | FF | D8 | FF | E0 | 00 | 10 | 4A | 46 | 49 | 46 | 00 | 01 | 01 | 00 | 00 | 01 | ÿØÿà..JFIF..... |
| 00000010 | 00 | 01 | 00 | 00 | FF | DB | 00 | 43 | 00 | 06 | 04 | 05 | 06 | 05 | 04 | 06 | ....ÿÛ.C..... |

FBI.CYBERFORENSICS.PROPOSAL.doc

| Offset | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 00000000 | D0 | CF | 11 | E0 | A1 | B1 | 1A | E1 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | ÐÏ.à¡±.á..... |
| 00000010 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 00 | 3E | 00 | 03 | 00 | FE | FF | 09 | 00 | .....>...þÿ.. |

# File Signature Analysis

- 25h 50h 44h 46h — PDF
- FFh D8h FFh — JPEG
- D0h CFh 11h E0h A1h B1h 1Ah E1h — MS Office

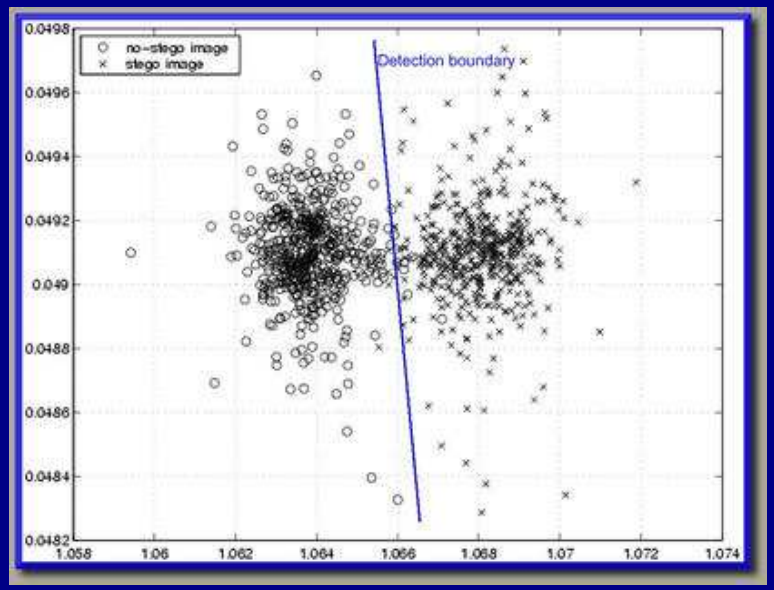
By comparing it with file extensions we discover hidden files (for instance naked.doc)

# Tools for Steganalysis

- Stegdetect is an automated tool for detecting steganographic content in JPEG images
- Stegbreak is used to launch dictionary attacks against JSteg-Shell, JPHide and OutGuess 0.13b.

# Tools for Steganalysis

- StegDetect
  - Uses Linear discriminant analysis computes a dividing hyperplane that separates the no-stego images from the stego images.
- Virtual Steganographic Laboratory
  - Set of steganographic modules
  - Set of tools for statistical analysis
  - Possibility to write own add-ins
- Matlab Tool
  - Visual preview of multi-dimensional data
  - Built-in Statistical tools
  - Operations on vectors and matrixes of values



# Hash analysis

- specify a list of MD5s
- finding several known bad files,
  - files from a rootkit
  - illegal images
- the MD5s that compose file (KNOWN.BAD) of already known files such as
  - f53ce230616c1f6aafedf546a7cc0f0f Trojan ps
  - bbf3aeb654477c4733bddf9a6360d2c5 Illegal Image
- run md5deep against all of the files in the directory
- It compares the file's hashes with the contents of the list of known hashes. If a match occurs, it lists it on standard out

# Finding Files by Type and Keyword Searches

- # find / -type f \( -name '*.gif' –or –name '*.jpg' –or –name '*.bmp' –or –name '*.png' \)

- # grep –i –r –f keywords /image/* > /evidence/grep.results
  - i - case insensitive search, thus 'cocaine,' 'COCAINE.' and 'CoCainE'
  - r - a recursive search, i.e., traverse all of the subdirectories beneath the current directory.
  - f - the next parameter is the file containing our keywords.

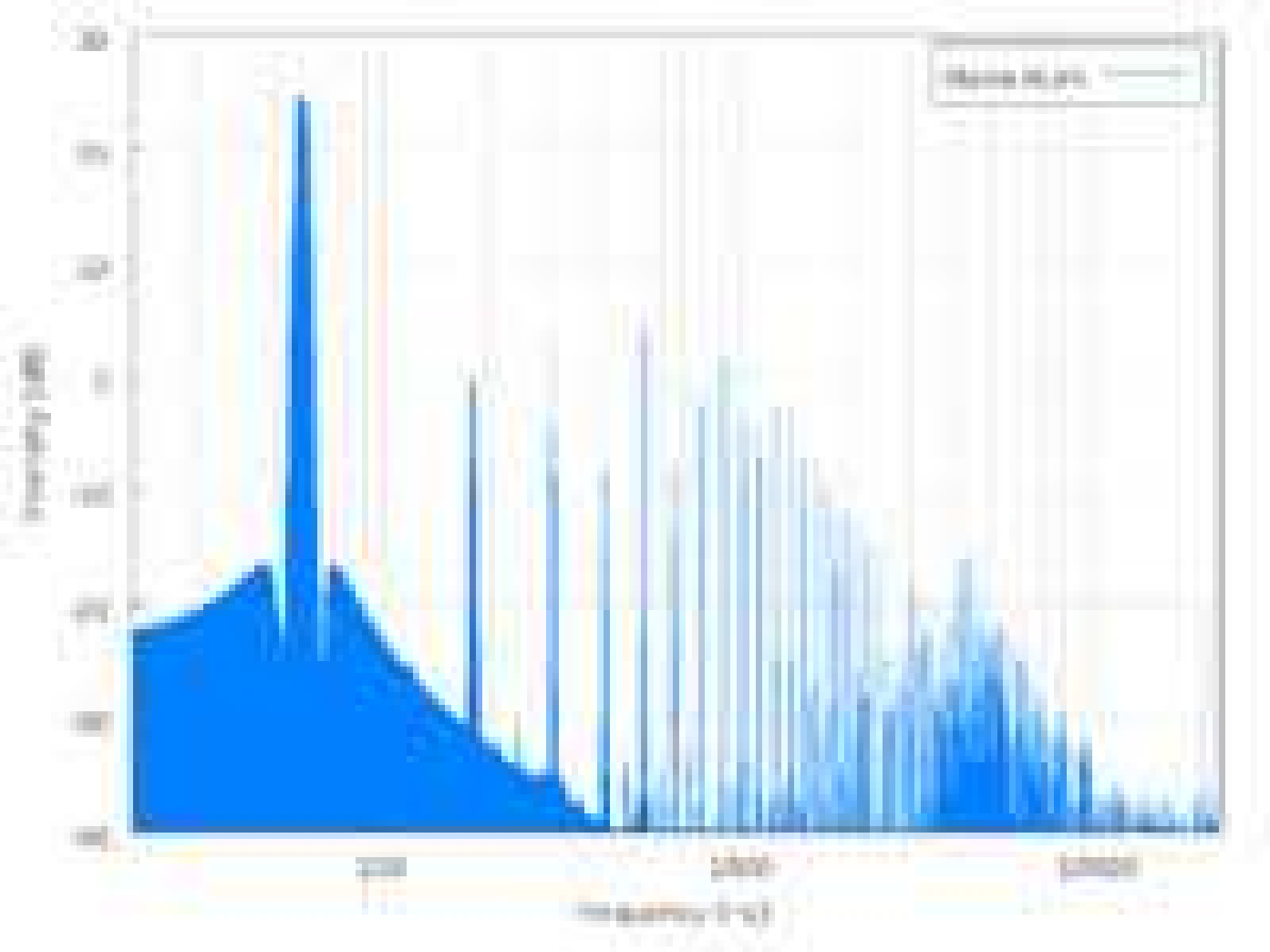
# Hierarchic file search

- To find all graphical files regardless of extension – 3 steps
  - 1: find command to find all regular files on the hard drive
  - Pipe the results to the next step
  - 2: file command, which returns the type of file using header information.
  - Pipe the results to the next step
  - 3: grep command to search for graphical-related keywords.

# Timestamp in electric hum

- European electricity network
- local fluctuations in the current frequency (around expected value of 50Hz)
- the frequency is the function of time and equal on a given area
- a correlation is observed in the fluctuations in all union area (Union for the Coordination of the Transmission of Electricity)
- one can estimate the time of audio recording, because an additional signal is added due to the frequency fluctuations
- PSE-operators store values of frequency in data bases since 1997
- Thus an investigator can detect if a given evidence audio has been manipulated, reassembled or partially replaced

# Example spectrum



# Forensic Tools

- Pasco (www.foundstone.com) - parses the contents of index.dat files, and outputs the results into a tab delimited file
- rifuiti (www.foundstone.com) - interprets the binary contents of the INFO2 file.

- EnCase (Forensic or Enterprise Editions www.guidancesoftware.com)
- Accessdata's Forensic Toolkit (part of the Ultimate Toolkit: www.accessdata.com).
  - imaging;
  - reading multiple file systems;
  - reading
  - multiple image formats;
  - file viewing; advanced string searches;
  - graphical/gallery views;
  - email analysis;
  - compressed file analysis;
  - known file filters/hash analysis;
  - bad file extension determination

# Commercial Forensic Tools

- • ARS Data's SMART (runs under Linux): http://www.asrdata.com/tools/
- • ILook Investigator (law enforcement only): http://www.ilook-forensics.org/
- • Maresware Forensic Tools: http://www.dmares.com/maresware
- • New Technologies Forensic Suite: http://www.forensics-intl.com/tools.html
- • Paraben Forensic Toolks: http://www.paraben-forensics.com/

# Conclusions

- People DO use steganography, while governments ban cryptography
- Steganography CAN be useful due to watermarking
- Steganography is difficult to be detected – it requires brute force attacks and huge size of computer resources

# Do you have questions?

- search of an automated tool which will automatically RUN, provide with PASSWORDS, and retrieve RESULTS form various steganographic programs
- needed help from computers on the Internet to run simultaneously (sth similar to the BOINC project

# Links

- http://en.wikipedia.org/wiki/Steganography
- http://www.outguess.org/detection.php
- http://www.zvetcobiometrics.com/Support/security_technology/algorithms.php

# Bibliography - 1


- „Information Hiding Using Audio Steganography - A Survey”, Jayaram P, Ranganatha H R, Anupama H S, Department of Computer Science and Engineering, R V College of Engineering, Bangalore, INDIA, The International Journal of Multimedia & Its Applications (IJMA) Vol.3, No.3, August 2011
- „A Watermarking Scheme for MP3 Audio Files” - Dimitrios Koukopoulos, Yiannis Stamatiou,International Journal of Information and Communication Engineering, 2006
- „Hide and Seek: An Introduction to Stegangography” - Niels Provos and Peter Honeyman, IEEE Security & Privacy Magazine, May/June 2003.
- „Fingerprint matching by genetic algorithms” - Xuejun Tan, Bir Bhanu, Center for Research in Intelligent System, University of California, Riverside, CA 92521, USA
- „Analiza wahań częstotliwości prądu sieciowego w badaniach autentyczności nagrań cyfrowych” - Iwona Biernacka, Rafał Korycki, Jacek Rzeszotarski, Przegląd Bezpieczeństwa Wewnętrznego, ABW
- „Analiza działania wybranych aplikacji steganograficznych” - Marta Walenczykowska, Przegląd Bezpieczeństwa Wewnętrznego, ABW

# Software downloads

- http://www.jjtc.com/
- http://www.outguess.org/
- http://sourceforge.net/projects/vsl/
- http://www.invisiblesecrets.com/
- https://www.mathworks.com/products/matlab/trial.html
- http://md5deep.sourceforge.net/
- https:// www.foundstone.com
- https:// www.guidancesoftware.com
- https:// www.accessdata.com