Steganography and Steganalysis in digital age

Tomáš Pevný

Agent Technology Center, CTU

3rd December 2009

Outline

1. Introduction
   - What is steganography and steganalysis
   - Definition of security
   - Example of steganography
2. Detecting LSB matching
   - Subtractive Pixel Adjacency Matrix
   - Experimental verification
3. Future direction

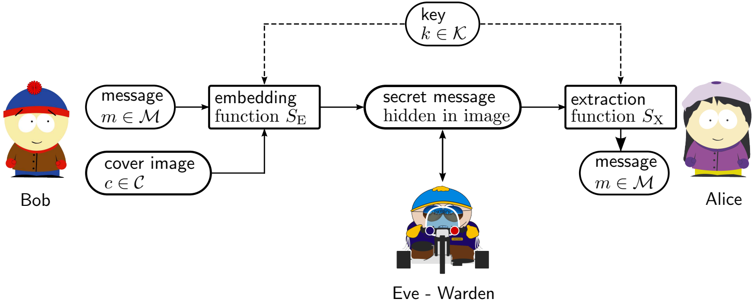
Outline

1 Introduction
- What is steganography and steganalysis
- Definition of security
- Example of steganography

2 Detecting LSB matching
- Subtractive Pixel Adjacency Matrix
- Experimental verification

3 Future direction

What is steganography?



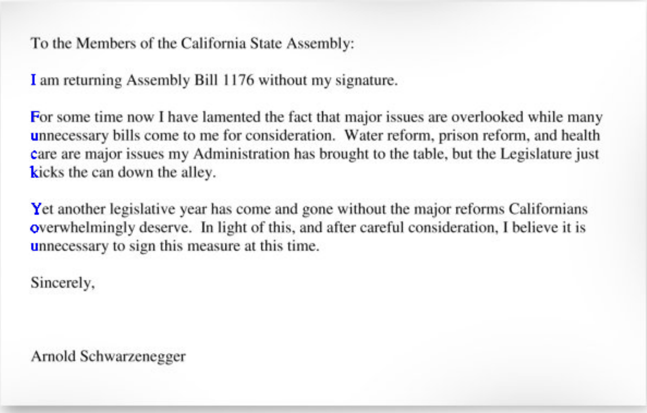
Steganography and Steganalysis

- *Steganography* is the art of undetectably communicating message in an innocuous looking object.
- *Steganos* (covered) + *graphia* (writing), J. Trithemius, 1499
- *Steganalysis* is an inverse topic.

Little history

- First written evidence comes from ancient Greece about 470BC (wax covered tablets, slave's scalp).
- Messages written on the back of postage stamps.
- Invisible ink (lemon juice, water, etc.).
- Microdots (Nazis, WWII).
- Transferred meanings of words (Japan, WWII).
- Com. J. Denton blinked by his eyes TORTURE in Morse code during propaganda filming in Vietnam prison.
- Steganography in its modern form is only approx. 17 years old.

Schwarzenegger’s letter



To the Members of the California State Assembly:

I am returning Assembly Bill 1176 without my signature.

For some time now I have lamented the fact that major issues are overlooked while many **u**nnecessary bills come to me for consideration. Water reform, prison reform, and health **c**are are major issues my Administration has brought to the table, but the Legislature just **k**icks the can down the alley.

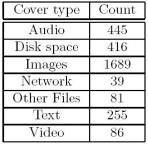
Yet another legislative year has come and gone without the major reforms Californians **o**verwhelmingly deserve. In light of this, and after careful consideration, I believe it is **u**nnecessary to sign this measure at this time.

Sincerely,

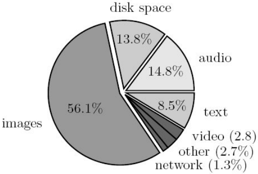
Arnold Schwarzenegger

A letter of gov. A. Schwarzenegger to T. Ammiano, S.F. Gate, October 28, 2009

Modern steganography



Cover type	Count
Audio	445
Disk space	416
Images	1689
Network	39
Other Files	81
Text	255
Video	86

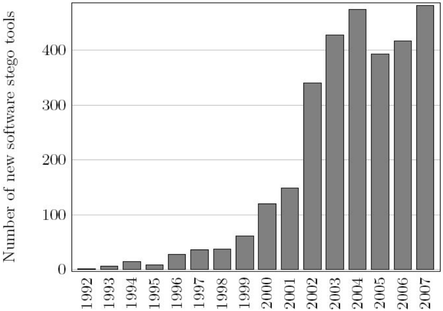


Steganographic software by type of hideout media.
(data provided courtesy of N. Johnson
figure provided courtesy of J. Fridrich)

Who uses steganography and why?

- In some countries the cryptography is prohibited (China, Belarus, Russia,. . .) or restricted (UK).
- Used by secret services (no information).
- Used by terrorists
  - Dhiren Barot, an Al Qaeda operative filmed reconnaissance video between Broadway and South Street and concealed it by splicing it into a copy of the Bruce Willis movie "Die Hard: With a Vengeance." Barot was sentenced to 40-to-life in Great Britain. *NY Times, 08/11/2006*
  - Technical Mujahid, a Training Manual for Jihadis contains chapter about steganography.
- Steganography program S-Tools was used to distribute child porn. This case occurred between 1998 and 2000.

Number of software titles by release date



Number of newly released steganographic software titles per year.
(data provided courtesy of N. Johnson
figure provided courtesy of J. Fridrich)

Interests from government and law enforcement

- Major US agencies funding research in steganography
  - US Air Force and AFOSR
  - National Institute of Justice (NIJ)
  - Office of Naval Research (ONR)
  - National Science Foundation (NSF)
  - Defense Advanced Research Project Agency (DARPA)
- Steganalysis is considered part of Computer Forensics.
- Steganalysis is important for protection against malware.
- Tools developed for steganalysis find applications in Digital Forensics in general (e.g., for detection of digital forgeries and integrity and origin verification).

Conferences

Major conferences

- SPIE Electronic Imaging, January, San Jose
- Information Hiding Workshop
- ACM Multimedia and Security Workshop
- IEEE Workshop on Information Forensics and Security
- IEEE International Conference on Image Processing

Research groups

- 5 university laboratories in U.S (Binghamton, Purdue,. . .)
- 7 research groups in Europe (Oxford, Dresden,. . .)
- other laboratories in China, Korea, India, Israel, etc.

Relation to other data hiding techniques

Steganography

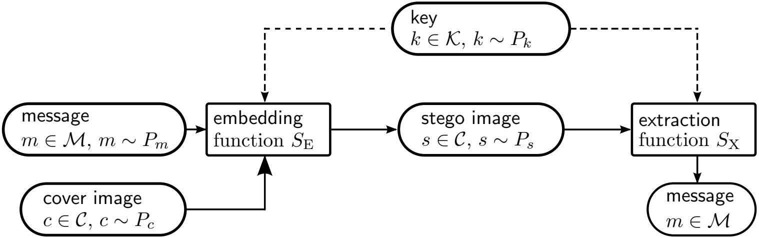
- It is fragile, as small change can make the message unreadable.
- It has to be undetectable.
- It should provide high capacity.

Watermarking

- *Watermarking* — robust against distortion / removal attacks.
- Its presence can be detected,
- It usually has low capacity.

Boundaries are blurred, other application exists (Secure Digital Camera).

Prisoner's problem


key
$k \in \mathcal{K}$, $k \sim P_k$
message
$m \in \mathcal{M}$, $m \sim P_m$
embedding function S_E
stego image
$s \in \mathcal{C}$, $s \sim P_s$
extraction function S_X
cover image
$c \in \mathcal{C}$, $c \sim P_c$
message
$m \in \mathcal{M}$

Steganographic algorithm

Steganographic algorithm is a tuple $(\mathscr{S}_\mathrm{E}, \mathscr{S}_\mathrm{X})$, where

- $\mathscr{S}_\mathrm{E} : \mathscr{C} \times \mathscr{M} \times \mathscr{K} \mapsto \mathscr{C}$ is an embedding function
- $\mathscr{S}_\mathrm{X} : \mathscr{C} \times \mathscr{K} \mapsto \mathscr{M}$ is an extraction function

Security of steganographic algorithms

Security of steganographic algorithm

Steganographic algorithm is ε-secure if KL-divergence

$$D_{\mathrm{KL}}(P_c \| P_s) = -\sum_{c \in \mathscr{C}} P_c(c) \log \frac{P_c(c)}{P_s(s)} < \varepsilon,$$

where P_c / P_s is pdf of cover / stego objects.

Practical issues

- Probability distribution of cover objects P_c is unknown.
- Space of all cover objects $\mathscr{C}$ is too large to sample P_c.
- We have to rely on simplified models (statistical / analytical).

Simple example — LSB replacement

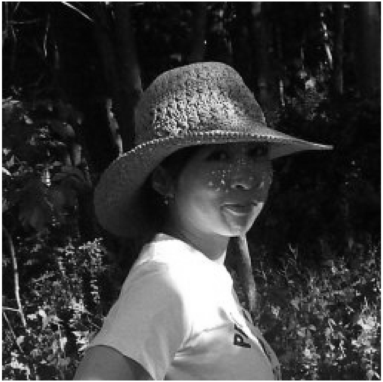


Image A

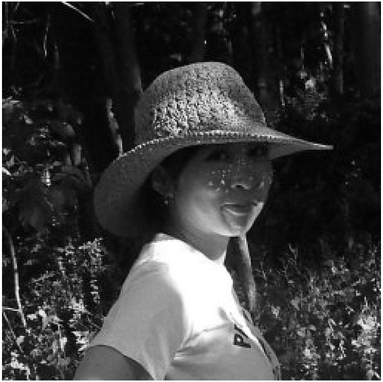
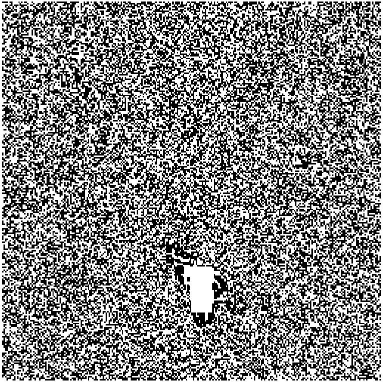
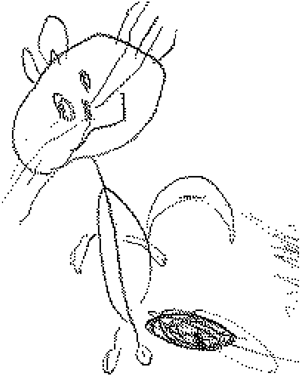


Image B

Simple example — LSB replacement



least significant bit of image A

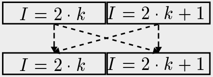


least significant bit of image B

LSB steganography in spatial domain

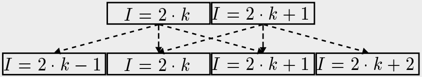
LSB Replacement

- replaces the least significant bit of the pixel with the message bit.
- is very detectable.
- It took about 5 years to be broken.


$I = 2 \cdot k$
$I = 2 \cdot k + 1$
$I = 2 \cdot k$
$I = 2 \cdot k + 1$

LSB Matching

- modulates the pixel value by adding ±1 to match the least significant bit with the message bit.
- very secure – hard to detect.
- has been broken in 2009.


$I = 2 \cdot k$
$I = 2 \cdot k + 1$
$I = 2 \cdot k - 1$
$I = 2 \cdot k$
$I = 2 \cdot k + 1$
$I = 2 \cdot k + 2$

Outline

1. Introduction
   - What is steganography and steganalysis
   - Definition of security
   - Example of steganography
2. Detecting LSB matching
   - Subtractive Pixel Adjacency Matrix
   - Experimental verification
3. Future direction

Different flavors of steganalysis

Heuristic steganalysis

100% relies on steganalyst detail knowledge of the algorithm.

Blind steganalysis

combines knowledge

- extracted from the training set
- from steganographic features.

Our approach to break LSB matching

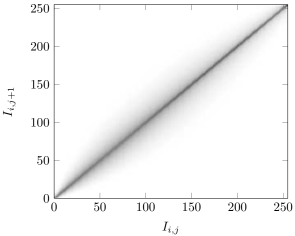
Motivation

- LSB Matching was very secure steganographic algorithm.
- We wanted to use very general, possibly high-dimensional image model and rely on robust machine learning algorithm.

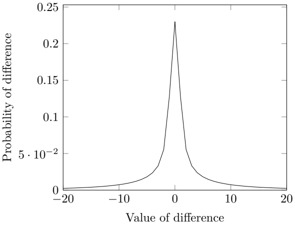
Approach in a nutsheel

- Natural noise in neighboring pixels is dependent due to image processing — defective pixel removal, demosaicing, noise reduction, etc.
- *The stego noise caused by LSB Matching is truly pixel to pixel independent* — it can be detected.

From image to noise model



Histogram of co-occurrences between adjacent pixels.



Histogram of differences between adjacent pixels.

- Detection of LSB Matching needs higher order statistics.
- Idea: instead of image, we model image noise from differences between adjacent pixels $D_{r,s} = I_{r+1,s} - I_{r,s}$

Noise model

- Differences are modeled by 2^{nd} order Markov model

  $$\mathbf{M}_{i,j,k} = P(D_{r+2,s} = i | D_{r+1,s} = j \wedge D_{r,s} = k),\ i,j,k \in \{-T,\ldots,T\}$$

  along 8 directions $\leftarrow, \rightarrow, \downarrow, \uparrow, \nwarrow, \searrow, \swarrow, \nearrow$
- The features $\mathbf{F}$ are formed from $\mathbf{M}$ by averaging

  $$\begin{aligned} \mathbf{F}^{\cdot}_{1,\ldots,k} &= \frac{1}{4}\left[\mathbf{M}^{\rightarrow}_{\cdot} + \mathbf{M}^{\leftarrow}_{\cdot} + \mathbf{M}^{\downarrow}_{\cdot} + \mathbf{M}^{\uparrow}_{\cdot}\right], \\ \mathbf{F}^{\cdot}_{k+1,\ldots,2k} &= \frac{1}{4}\left[\mathbf{M}^{\searrow}_{\cdot} + \mathbf{M}^{\nwarrow}_{\cdot} + \mathbf{M}^{\swarrow}_{\cdot} + \mathbf{M}^{\nearrow}_{\cdot}\right]. \end{aligned}$$

- The number of features depends on range of differences T and order of Markov chain (in our experiments, we used $T = 3$).

Experimental comparison — feature sets

Feature sets

- SPAM features with $T = 3$ (686 features).
- WAM features of Goljan et al., 2006 (81 features).
- ALE features of Cancelli et al., 2008 (10 features).

Classifiers

All classifiers were implemented by Support Vector Machines with Gaussian kernel. The error was measured by

$$P_{\mathrm{Err}} = \frac{1}{2}\left(P_{\mathrm{Fp}} + P_{\mathrm{Fn}}\right).$$

Practical issues with test images

- Images needed to evaluate performance of newly proposed steganographic and steganalytic methods have to be
  - clean (no hidden data)
  - not compressed by lossy compression (JPEG).
- We cannot use publicly available images (flicker, Picassa, etc.) — we do not know their history.
- Ideal images are stored in camera (raw) format.
- Most researchers rely on private sources / databases.

Used image databases

1. **CAMERA** contains ≈ 9200 images captured by 23 different digital cameras in the raw format and converted to grayscale.
2. **BOWS2** contains ≈ 10800 grayscale images with fixed size 512×512 used in the BOWS2 contest.
3. **NRCS** consists of 1576 raw scans converted to grayscale.
4. **JPEG85** contains 9200 images from **CAMERA** database compressed by JPEG with quality factor 85.
5. **JOINT** contains images from all four databases above, ≈ 30800 images.

By LSB matching, we created 2 sets of stego images with payloads 0.25 and 0.5 bits per pixel.

Comparison to prior art

database	bpp	SPAM	WAM	ALE
CAMERA	0.25	**0.057**	0.185	0.337
BOWS2	0.25	**0.054**	0.170	0.313
NRCS	0.25	**0.167**	0.293	0.319
JPEG85	0.25	**0.008**	0.018	0.257
JOINT	0.25	**0.074**	0.206	0.376
CAMERA	0.50	**0.026**	0.090	0.231
BOWS2	0.50	**0.024**	0.074	0.181
NRCS	0.50	**0.068**	0.157	0.259
JPEG85	0.50	**0.002**	0.003	0.155
JOINT	0.50	**0.037**	0.117	0.268

Tab:Error P_{Err} of SVM classifiers using different feature sets.

ROC curves on Joint database

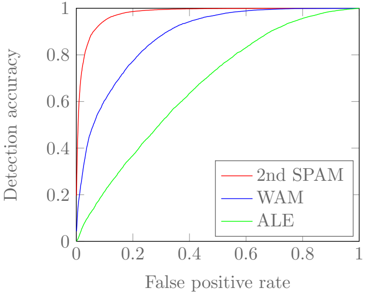


Fig: Payload 0.25 bits per pixel

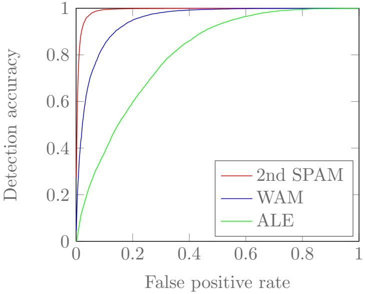


Fig: Payload 0.5 bits per pixel

Outline

1. Introduction
   - What is steganography and steganalysis
   - Definition of security
   - Example of steganography

2. Detecting LSB matching
   - Subtractive Pixel Adjacency Matrix
   - Experimental verification

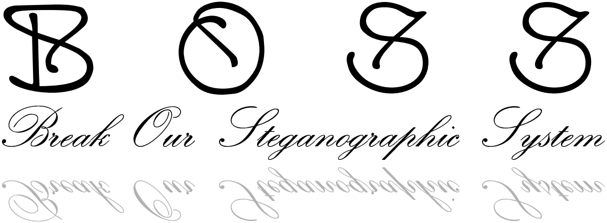
3. Future direction

Conclusion & future directions

Future directions in steganalysis

- Discrepancy between theory and practice — absent knowledge of attacked algorithm.
- Make it more robust against variations in macroscopic properties of images.
- Estimate confidence of performed steganalysis.
- Pooled steganalysis.

Do you want to join the game?



Steganalytic challenge is coming up in 2010!
1000 images, 500 with a hidden message
Guess which ones!

`http://boss.gipsa-lab.grenoble-inp.fr`