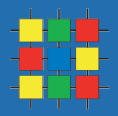
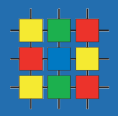


Steganography and Steganalysis in digital age

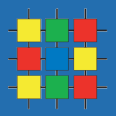
Tomáš Pevný

Agent Technology Center, CTU

23rd October 2013

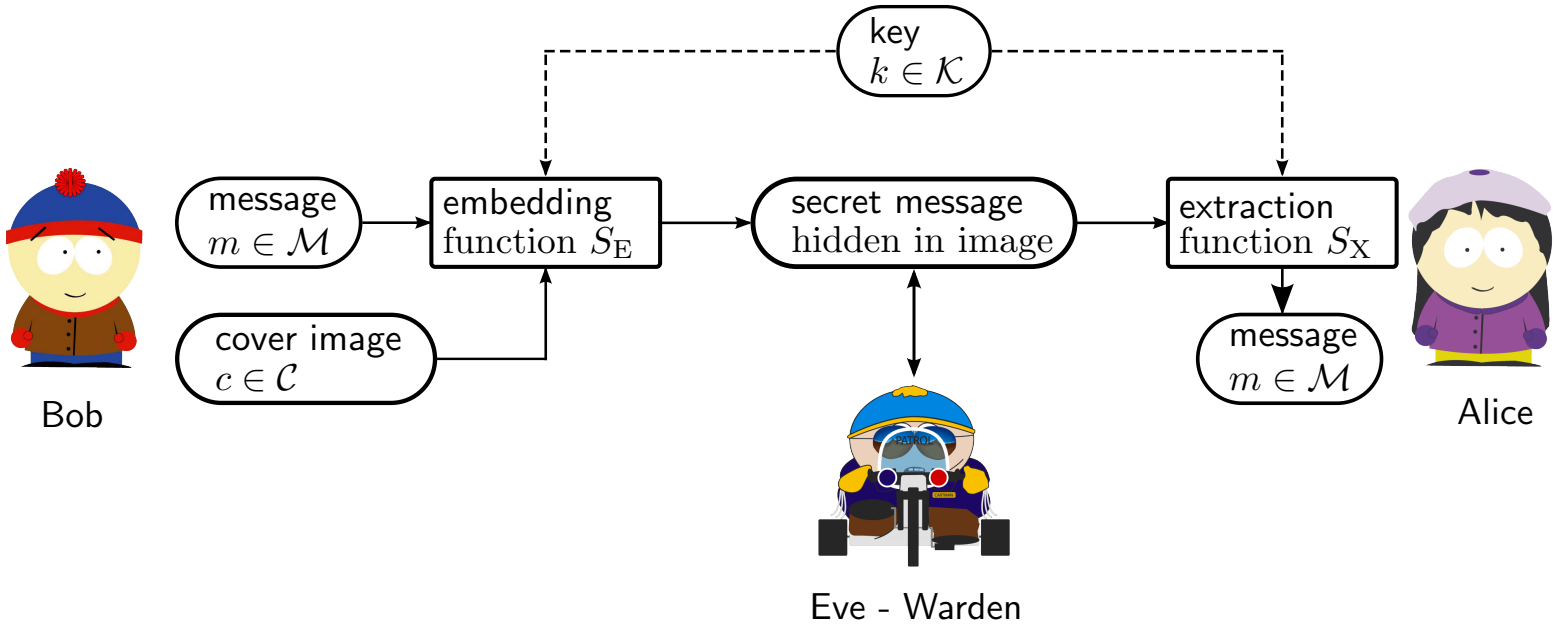


Outline



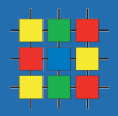
Outline

What is steganography?



Steganography

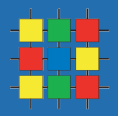
- *Steganography* is the art of undetectably communicating message in an innocuous looking object.
- *Steganos* (covered) + *graphia* (writing), J. Trithemius, 1499
- The most important property is undetectability.



Difference to cryptography

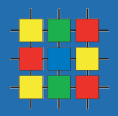
Difference to cryptography

- Crypto makes the message unintelligible, but the existence of secret message is obvious (overt).
- Stego conceals the very presence of message (covert), the communicated object is just a decoy.
- Cryptography provides privacy.
- Steganography provides secrecy.



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 utilizing digital media 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.

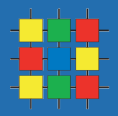
For some time now I have lamented the fact that major issues are overlooked while many unnecessary bills come to me for consideration. Water reform, prison reform, and health care are major issues my Administration has brought to the table, but the Legislature just kicks the can down the alley.

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

Sincerely,

Arnold Schwarzenegger

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



Schwarzenegger's letter

To the Members of the California State Assembly:

I am returning Assembly Bill 1176 without my signature.

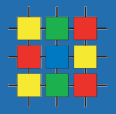
For some time now I have lamented the fact that major issues are overlooked while many unnecessary bills come to me for consideration. Water reform, prison reform, and health care are major issues my Administration has brought to the table, but the Legislature just kicks the can down the alley.

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

Sincerely,

Arnold Schwarzenegger

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



Steganographic channel (1)

Steganographic channel

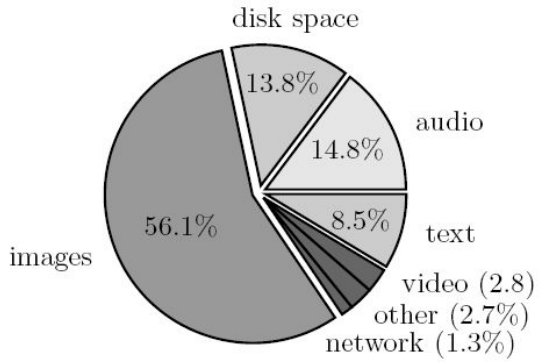
- Enables the exchange of the "innocuous" messages.
- Any periodically visited site with medias is good.

Examples

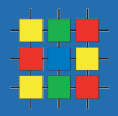
- Media sharing sites: flicker, youtube, picasa, e-bay etc.
- voice-over-IP (skype), timing of IP packets
- Yogurt story

Steganographic channel (2)

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

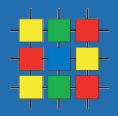


Steganographic software by type of hideout media.
(data provided courtesy of N. Johnson
figure provided courtesy of J. Fridrich)



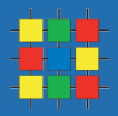
Who uses steganography and why?

- In some countries the cryptography is prohibited (China, Belarus, Russia,...) or restricted (UK).
- Used by secret services (no information).
  - June 2010, russian spies in US alleged to use steganography. `http://www.darkreading.com/security/news/225701866`
- Steganography program S-Tools was used to distribute child porn. This case occurred between 1998 and 2000.
- Malware for concealing Command and Control channel.



Used by terrorist

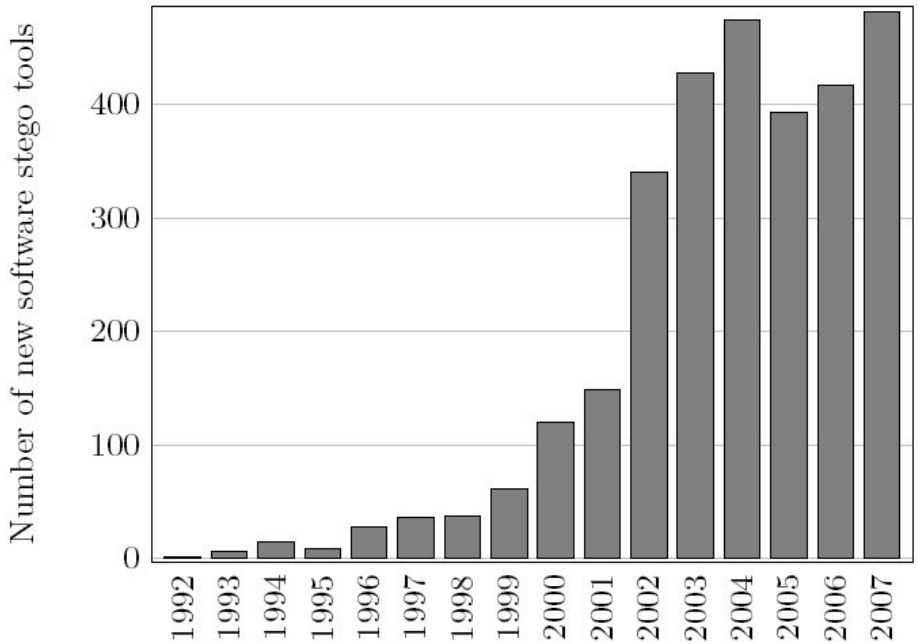
- Technical Mujahid, a Training Manual for Jihadis contains chapter about steganography.
- Dhiren Barot, an Al Qaeda operative filmed reconnaissance video between Broadway and South Street and concealed it by splicing it into a copy of the Bruce Willis movie "Die Hard: With a Vengeance."
  Barot was sentenced to 40-to-life in Great Britain. *NY Times, 08/11/2006*
- 1st May, 2012 CNN reported Al Qaeda courier was caught in Germany.
  `http://edition.cnn.com/2012/04/30/world/al-qaeda-documents-future/index.html`



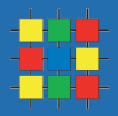
Used by terrorist

- Malware embeds payload into meta data in image containers.
  - blog.sucuri.net, July 2013
  - Fireeye report, page 15
- Replacing portion of images with the payload
  - blog.malwarebytes.org, February 2014
- "Decent" steganography
  - Lurk, February 2014

Number of software titles by release date

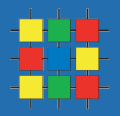


Number of newly released steganographic software titles per year.
(data provided courtesy of N. Johnson
figure provided courtesy of J. Fridrich)



Publicly available software

- http://www.jjtc.com/Steganography/toolmatrix.htm
- https://chrome.google.com/webstore/detail/secretbook/plglafijddgpenmohgiemalpcfgjjbph



Relation to other data hiding techniques

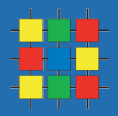
Steganography

- 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 (robust steganography, fragile watermarking), other application exists (Secure Digital Camera).



Example: LSB replacement (1)

cover image

message

Example: LSB replacement (1)

cover image

message

Example: LSB replacement (1)

cover image

message

Example: LSB replacement (1)

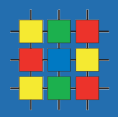
cover image

message

Example: LSB replacement (1)

cover image

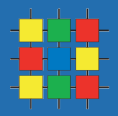
message



Example: LSB replacement (1)

cover image

message



Example: LSB replacement (1)

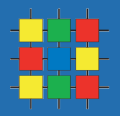
cover image

message

Example: LSB replacement (1)

cover image

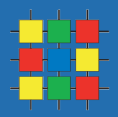
message



Example: LSB replacement (1)

cover image

stego image with hidden 256 color image



Example: LSB replacement (2)

cover image

Let's try to hide as much data, as possible.

Example: LSB replacement (2)

cover image

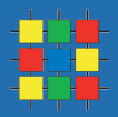
- 1st LSB bitplane overwritten.
- approximately 98Kb hidden.
- filesize 784Kb.

Example: LSB replacement (2)

- 1st and 2nd LSB bitplanes overwritten.
- approximately 196Kb hidden.
- filesize 784Kb.

Example: LSB replacement (2)

- 1st, 2nd, and 3rd LSB bitplanes overwritten.
- approximately 294Kb hidden.
- filesize 784Kb.

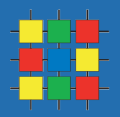


Example: LSB replacement (2)

- 1–4 LSB bitplanes overwritten.
- approximately 392Kb hidden.
- filesize 784Kb.

Example: LSB replacement (2)

- 1–5 LSB bitplanes overwritten.
- approximately 490Kb hidden.
- filesize 784Kb.

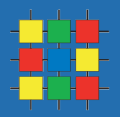


Example: LSB replacement (2)

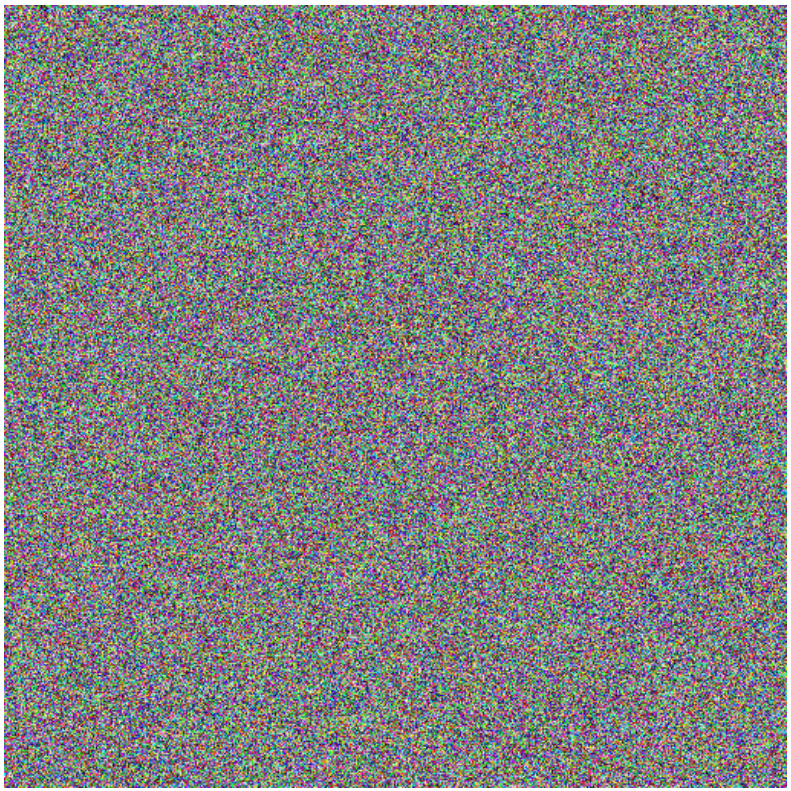
- 1–6 LSB bitplanes overwritten.
- approximately 588Kb hidden.
- filesize 784Kb.

Example: LSB replacement (2)

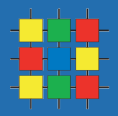
- 1–7 LSB bitplanes overwritten.
- approximately 686Kb hidden.
- filesize 784Kb.



Example: LSB replacement (2)

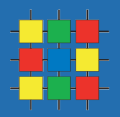


- 1–8 LSB bitplanes overwritten.
- approximately 784Kb hidden.
- filesize 784Kb.



Current approaches

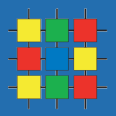
- Uses coding (syndrome trellis codes) to increase embedding efficiency.
- The location of embedding changes depends on the image content.



Hugo — content adaptive steganography

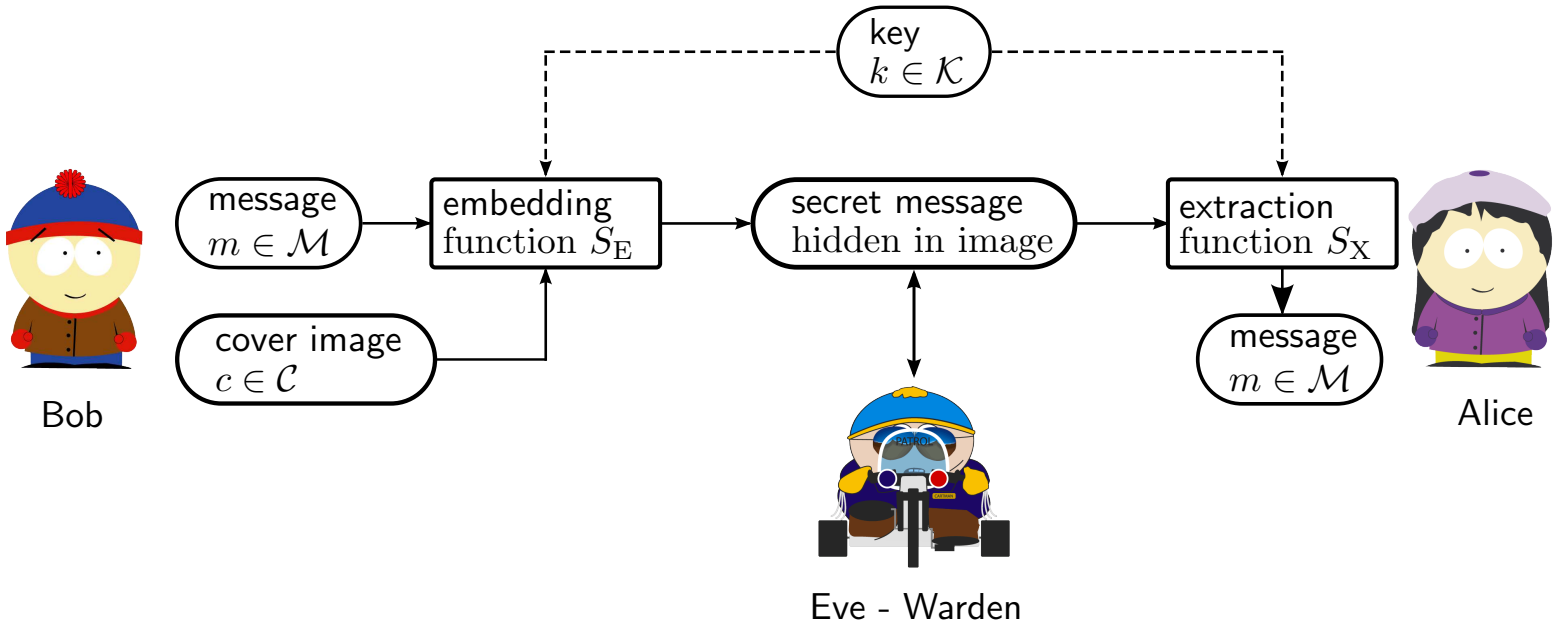
0.25 bits per pixel

0.5 bits per pixel



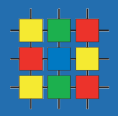
Outline

What is steganalysis?



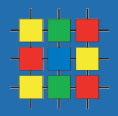
Steganalysis

- *Steganalysis* aims to detect the presence of secret message.



Who is interested in steganalysis?

- Interests from government and law enforcement.
- 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).
- 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 in a wide sense

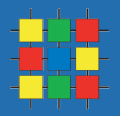
Traditional steganalysis

- Traditional *steganalysis* detects the mere presence of secret message.

Forensic steganalysis

Detection is not sufficient, we want to know more:

- identification of the embedding algorithm (LSB,± 1 ,. . .)
- the stego software used (F5 , Outguess, Steganos, . . .)
- the stego key
- the hidden bit-stream
- the decrypted message



Different flavors of steganalysis

Visual steganalysis

- human intensive.
- rarely used in practice.

Heuristic steganalysis

100% relies on steganalyst detail knowledge of the algorithm.

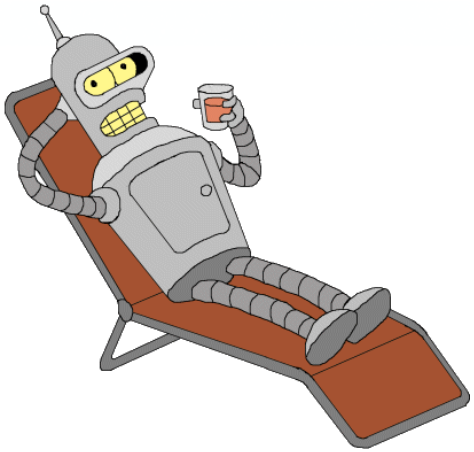
Blind steganalysis

combines knowledge

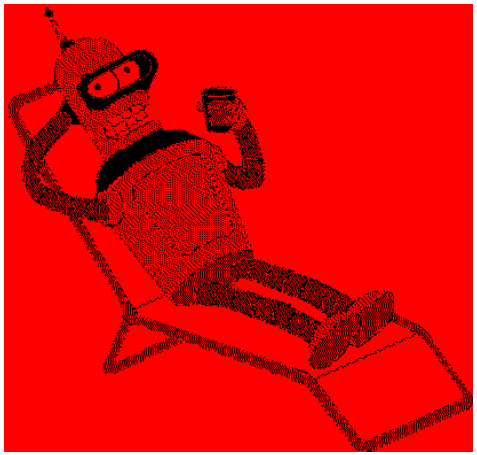
- extracted from the training set
- from steganographic features.

Visual steganalysis

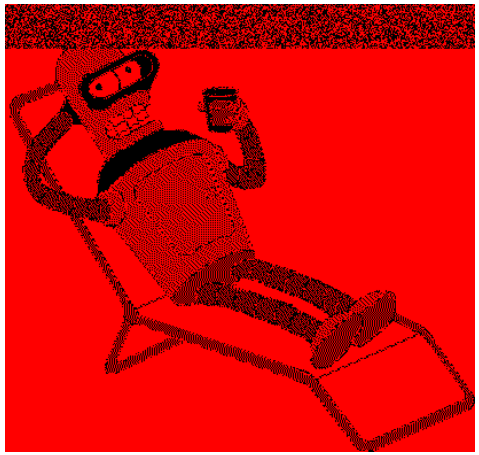
Invisible changes may become visible after appropriate processing.



Stego image



LSB of red channel of cover image

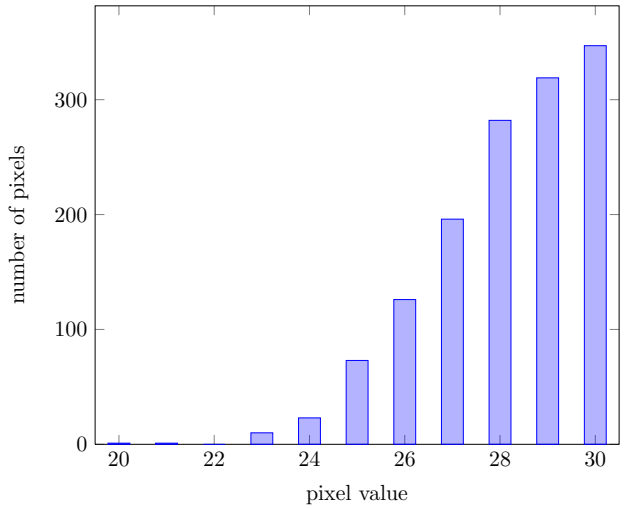


LSB of red channel of stego image

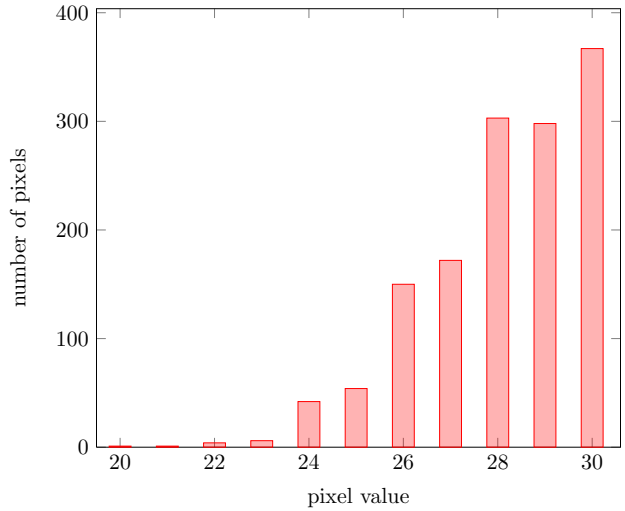
(source: A. Westfeld and A. Pfitzmann, "Attacks on Steganographic Systems," Lecture Notes in Computer Science, vol.1768, Springer-Verlag, Berlin, 2000, pp. 61–75)

Heuristic steganalysis

- amounts to find quantity predictably changing with the length of hidden message.



cover image



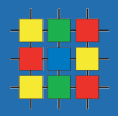
stego image

Histograms of pixel values.

Blind steganalysis

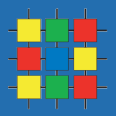
Blind steganalysis

- uses features to provide low-dimensional model of natural images (more later).
- pattern recognition algorithms are used to learn differences between cover and stego images.
- state of the art in steganalysis.



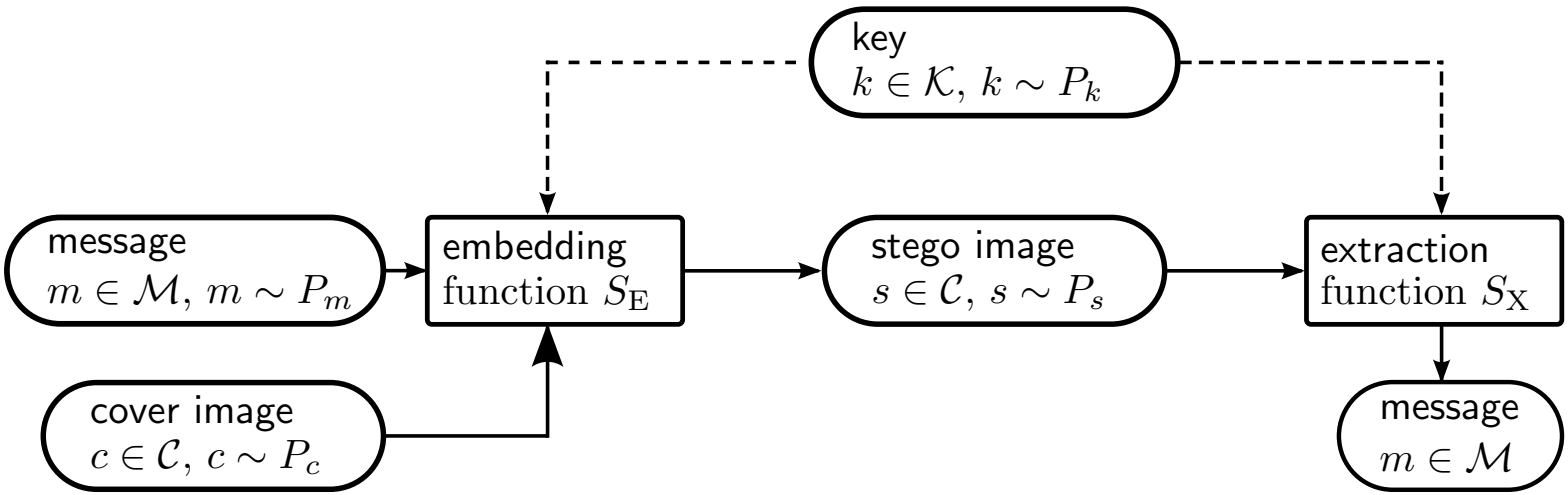
Current approaches

- Image is described by a large number of features (up to 50 000) sensitive to noise.
- Machine learning algorithms learns the difference between cover and stego.
- Problem with over-fitting / cover source mismatch.



Outline

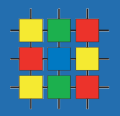
Steganographic algorithm



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

Goal

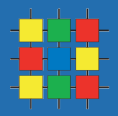
The most important property is undetectability.

Security of steganographic algorithm

- Kerckhoffs' principle
- For perfect steganographic algorithm holds $P_c = P_s$.
- Cachin's definition: steganographic algorithm is ε-secure iff 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.



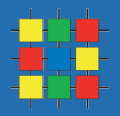
Practical Issues

Practical issues

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

Balance

- unrealistic conclusions (steganography).
- curse of dimensionality (steganalysis).
- cat and mouse game.



Current trends and open problems

Steganography

- Design of distortion functions, content adaptive steganography.
- Embedding in stream of images
- Steganography for color images / video / (timing channels).

Steganalysis

- High dimensional models, learning models.
- Learning from large number of images.
- Pooled steganalysis
- Evidence in front of the court.
- Cover-source mismatch / overfitting.