Cover channels, steganography and steganalysis

Tomáš Pevný

October 18, 2018

Outline

Covert channels

Side-channel attacks

Steganography

Steganalysis

Theory

Outline

Covert channels

Side-channel attacks

Steganography

Steganalysis

Theory

Covert channel

What is covert channel?

Covert channel

Covert channel is an information flow mechanism within a system that is based on the use of system resources not normally intended for communication between the users of the system.

Types of covert channel:

Trusted Computer Security Evaluation Criteria:

Storage channels use system variables and attributes (other than time) to signal information, e.g. — file status.

Timing channels vary the amount of time required to complete a particular task, e.g. — influencing the number of CPU cycles available to a given process in a given time frame.

Storage channels

1. The sending and receiving entities must have access to the same shared resource or attribute.
2. There must be some means by which the sending entity can force the shared resource or attribute to change.
3. There must be some means by which the receiving entity can detect the change.
4. There must be some mechanism for initiating the communication between the sending and receiving entities and for sequencing the events correctly. This mechanism could be a covert channel with a lower bandwidth.

If 1-3 are satisfied, one must find a scenario that satisfies 4. If such a scenario exists, a storage channel exists.

Timing channels

1. The sending and receiving entities must have access to the same shared resource or attribute.
2. The sending and receiving entities must have access to a time reference such as a real-time clock.
3. The sender must be capable of modulating the receiver's perception of time for detecting a change in the shared resource or attribute.
4. There must be some mechanism for initiating the channel and for sequencing the events transmitted over it.

Anytime a processor or memory is shared, there is a shared attribute. A change in response time is detected by the receiving process by means of monitoring the clock or its surrogate.

Harmful covert channels

- In innocuous cases covert channels exist between parties which are allowed do communicate.
- In dangerous cases covert channels exist between parties which are *not* allowed do communicate.

Identification of covert channels

- Identify all shared resources.
- Identify resources whose visibility is against policy.
- Estimate the bandwidth of the channel.

Outline

Covert channels

Side-channel attacks

Steganography

Steganalysis

Theory

General classes of side channel attack include:

- Cache attack
- Timing attack
- Power-monitoring attack
- Electromagnetic attack
- Acoustic cryptanalysis
- Differential fault analysis
- Software-initiated fault attack
- Data remanence
- Optical

Outline

Covert channels

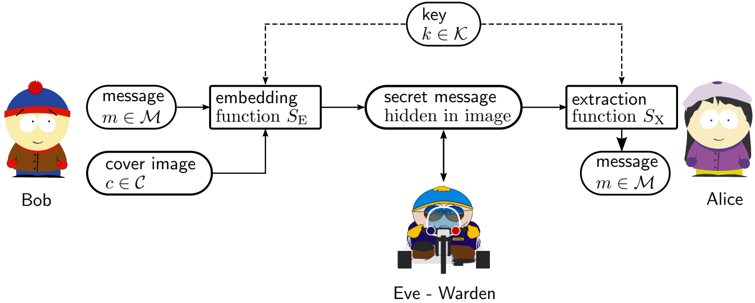
Side-channel attacks

Steganography

Steganalysis

Theory

What is steganography?



key $k \in \mathcal{K}$

Bob

message $m \in \mathcal{M}$

cover image $c \in \mathcal{C}$

embedding function S_{E}

secret message hidden in image

Eve - Warden

extraction function S_{X}

message $m \in \mathcal{M}$

Alice

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.

Steganographic channel (1)

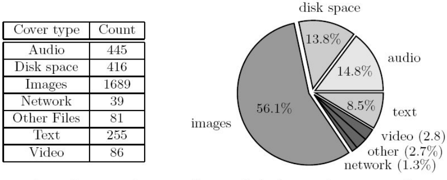
Steganographic channel

- 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)



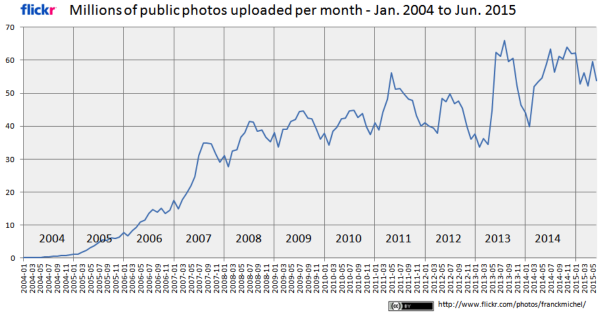
Cover type	Count
Audio	445
Disk space	416
Images	1689
Network	39
Other Files	81
Text	255
Video	86

Steganographic software by type of hideout media.

(data provided courtesy of N. Johnson

figure provided courtesy of J. Fridrich)

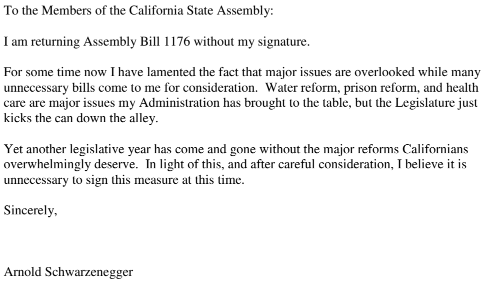
Why are image popular?



downloaded from:

https://www.flickr.com/photos/franckmichel/6855169886/in/photostream/

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

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

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

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`
- Steganography program S-Tools was used to distribute child porn. This case occurred between 1998 and 2000.

Used by dissidents and spies

- 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`

Used by malware

- 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
  - Vawtrack, March 2015
  - trend micro, October 2015
  - Stegano exploit kit, October 2015

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

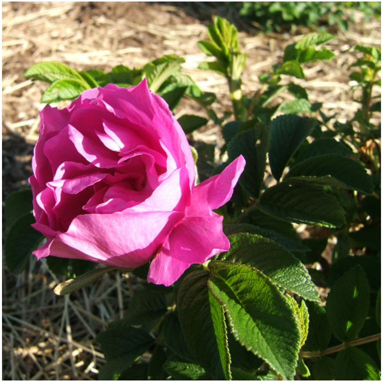
message

Example: LSB replacement (1)

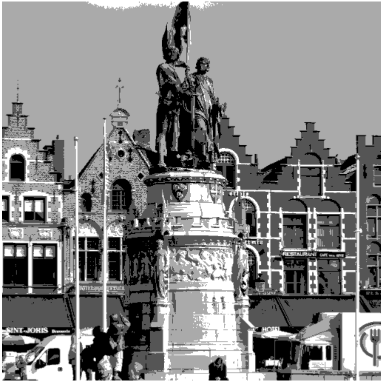
cover image

message

Example: LSB replacement (1)

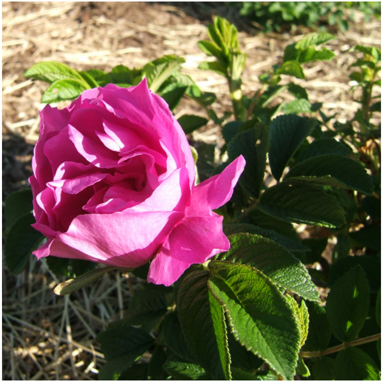


cover image

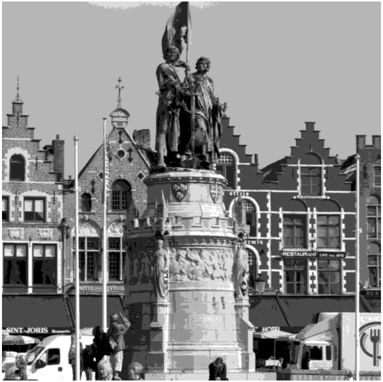


message

Example: LSB replacement (1)

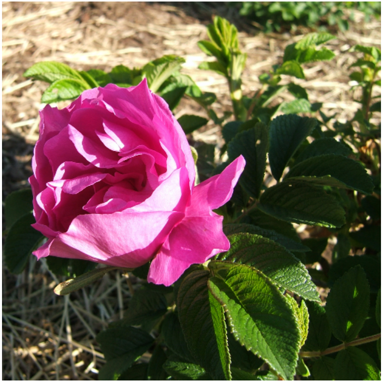


cover image

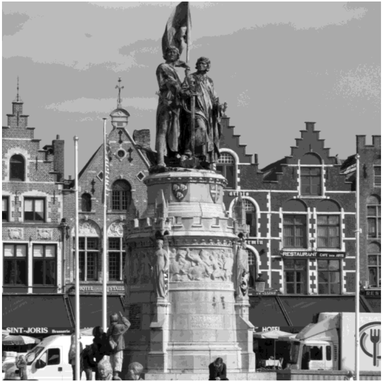


message

Example: LSB replacement (1)

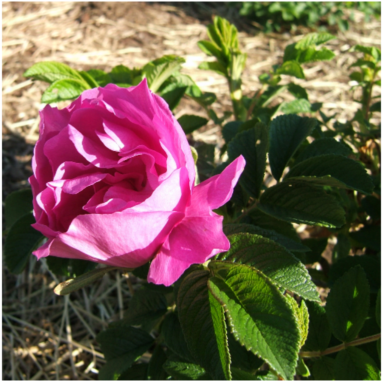


cover image

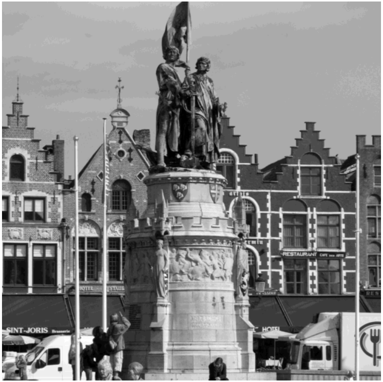


message

Example: LSB replacement (1)



cover image



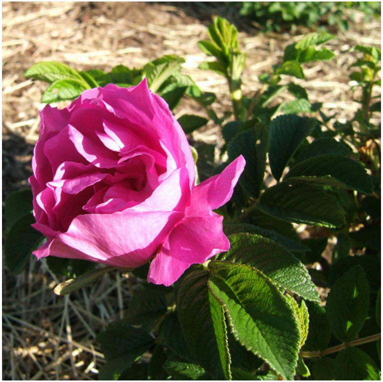
message

Example: LSB replacement (1)

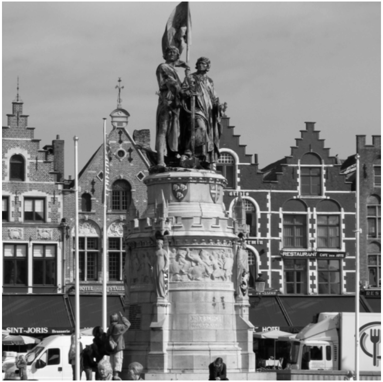
cover image

message

Example: LSB replacement (1)



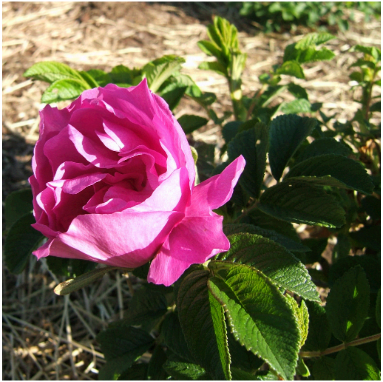
cover image



message

Example: LSB replacement (1)

cover image

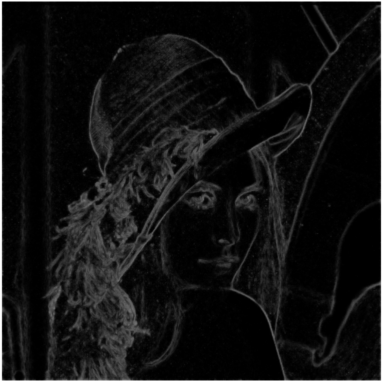


stego image with greyscale image inside

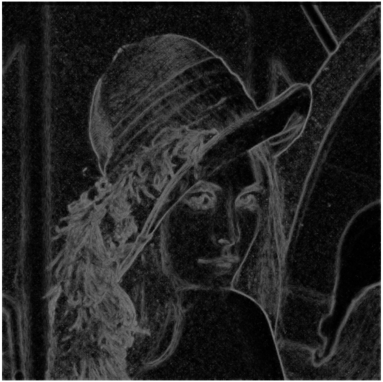
Current approaches

- 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

Covert channels

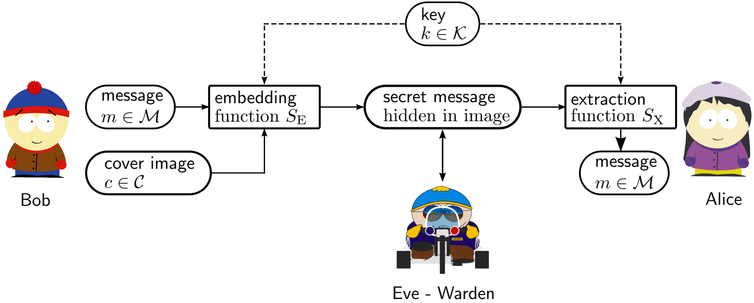
Side-channel attacks

Steganography

Steganalysis

Theory

What is steganalysis?



Steganalysis

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

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

Steganalysis as a statistical hypothesis test

Statistical testing

Hypothesis $\mathbf{H}_0$ states that scrutinized images is cover, $\mathbf{H}_1$ states that scrutinized images is stego. If detection statistic $T(x) > \tau$, the test output image is stego, otherwise cover.

Errors:

1. Type I occurs if test returns $\mathbf{H}_1$ when $\mathbf{H}_0$ is true (false positive), α.
2. Type II occurs if test returns $\mathbf{H}_0$ when $\mathbf{H}_1$ is true (false negative), β.

Different flavors of steganalysis

Visual steganalysis

- human intensive.
- rarely used in practice.

Heuristic steganalysis

100% relies on steganalyst detail knowledge of the algorithm.

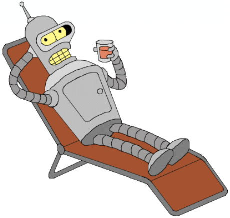
Blind steganalysis

combines knowledge

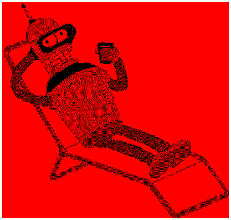
- extracted from the training set
- from steganographic features.

Visual steganalysis

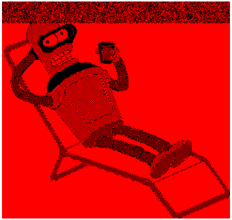
Invisible changes may become visible after appropriate processing.



Stego image



LSB of red channel of cover image

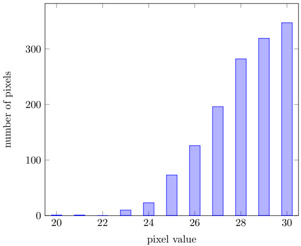


LSB of red channel of stego image

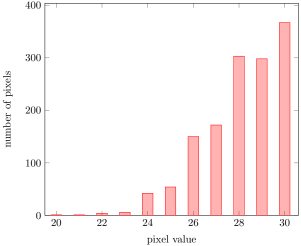
(source: A. Westfeld and A. Pfitzmann, "Attacks on Steganographic Systems," Lecture Notes in Computer Science, vol.1768, Springer-Verlag, Berlin, 2000, pp. 61–75)

Heuristic steganalysis

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



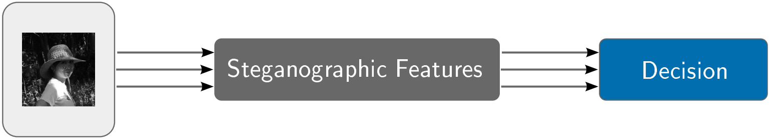
cover image



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

Covert channels

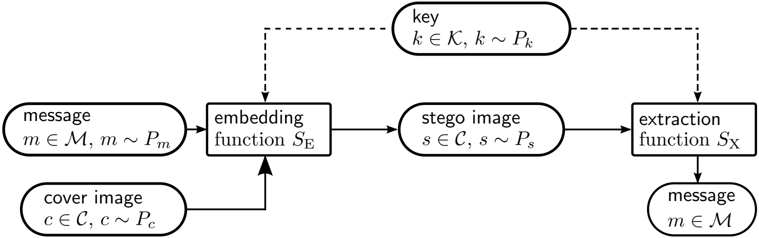
Side-channel attacks

Steganography

Steganalysis

Theory

Problem setting


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

Cachin's defition of steganographic security

- 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.

Why KL-divergence is important?

Provides bounds on best achievable detector.

Steganalysis as a statistical hypothesis test

Statistical testing

Hypothesis $\mathbf{H}_0$ states that scrutinized images is cover, $\mathbf{H}_1$ states that scrutinized images is stego. If detection statistic $f(x) > \tau$, the test output image is stego, otherwise cover.
Errors:

1. Type I occurs if test returns $\mathbf{H}_1$ when $\mathbf{H}_0$ is true (false positive), α.
2. Type II occurs if test returns $\mathbf{H}_0$ when $\mathbf{H}_1$ is true (false negative), β.

Question

Is the theory ($D_{\mathrm{KL}}(P_c \| P_s) < \varepsilon$) useful in practice?

Question

How does secure message length grows with size of cover media?

Square-root law for independent covers

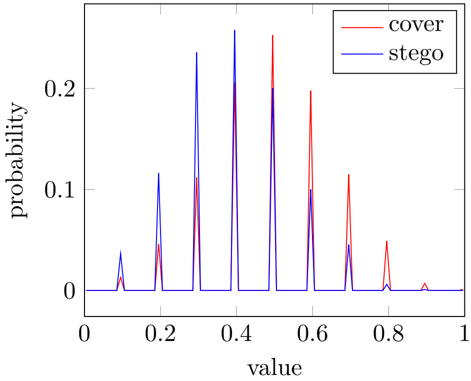
Assumptions:

1. Cover consists of n iid pixels $(x_1, \ldots, x_n) \sim p^n(x)$.
2. Payload of size m causes each pixel to be replaced, independently of each other and the cover, with probability $\lambda = \frac{m}{n}$, and that replaced pixels are independent each with mass function $q(x)$.
3. Suppose that for all x, $p(x) \neq 0$ and $q(x) \neq 0$, and there exists y such that $p(y) \neq q(y)$.

Statement:

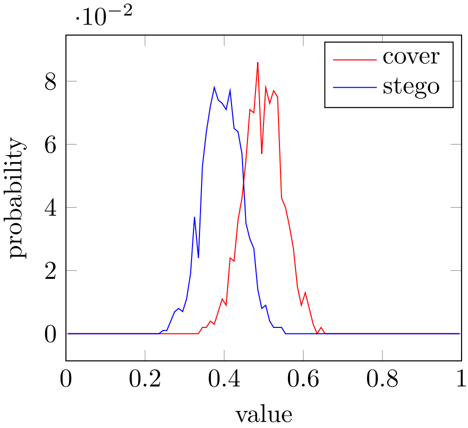
1. If $\frac{m}{\sqrt{n}} \mapsto \infty$ then, for sufficiently large n, covers and stego objects can be distinguished with arbitrarily low error rate.
2. If $\frac{m}{\sqrt{n}} \mapsto 0$ then, for sufficiently large n, any detector must have arbitrarily high error rate.

Intuition of the proof



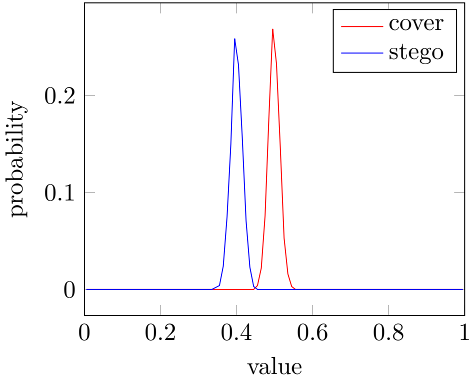
Distribution of estimates from 10 pixels

Intuition of the proof



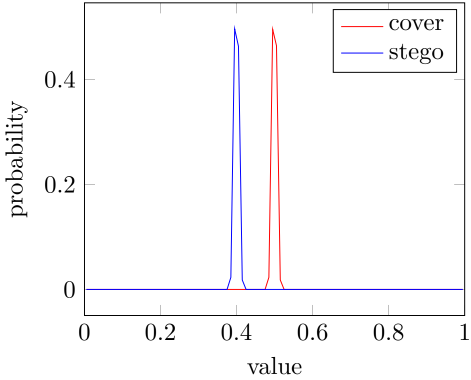
Distribution of estimates from 100 pixels

Intuition of the proof



Distribution of estimates from 1000 pixels

Intuition of the proof



Distribution of estimates from 10000 pixels

Formal proof of the second part

If $\frac{m}{\sqrt{n}} \mapsto \infty$ then, for sufficiently large n, covers and stego objects can be distinguished with arbitrarily low error rate.

1. cover pixels are iid with $(x_1, \ldots, x_n) \sim \mathrm{Ber}(p)$,
   stego pixels are iid with $(x_1, \ldots, x_n) \sim \mathrm{Ber}(q)$, $q > p$.
2. embedding operation changes $m = \lambda n$ pixels.
3. decision statistic $T = \sum_{i=1}^{n} x_i$,
4. stego iff $T > p + c\sqrt{n}$.

Hoeffding's inequality

Let X be a sum of n independent, not necessarily identically distributed, random variables each bounded in $[0,1]$, then for $t > 0$

$$Pr[X \geq E[X] + nt] \leq \exp(-2nt^2),$$

and

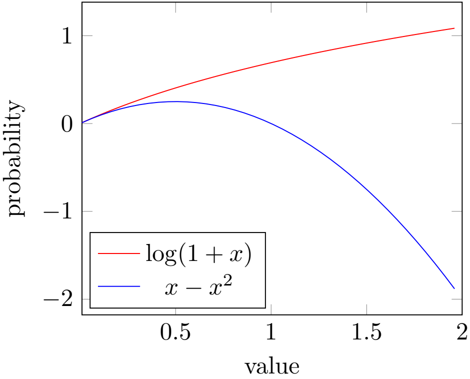
$$Pr[X \leq E[X] - nt] \leq \exp(-2nt^2).$$

Formal proof of the second part

If $\frac{m}{\sqrt{n}} \mapsto 0$ then, for sufficiently large n, any detector must have arbitrarily high error rate.

Need to show that $D_{\mathrm{KL}}(P_c \| P_s) \to 0$.

Formal proof of the second part



lower bound on $log(1+x) > x - x^2$

Square root law requires linear key

Assumptions:

1. Cover consists of n pixels $(x_1, \ldots, x_n)$ independent and identically distributed each with mass function$p(x)$.
2. Payload of size m which causes exactly m pixels to be replaced with mass function $q(x)$.
3. Sender, recipient, and attacker share knowledge of a set K of secret keys, each of which generates a path of length m determining the payload locations, but only the sender and recipient know which key is used.
4. Exists y such that $p(y) \neq q(y)$.

Statement:
If $\frac{\log |K|}{m} \mapsto 0$, as $m \mapsto +\infty$ and $m \mapsto +\infty$ as $n \mapsto +\infty$, then, for sufficiently large n, covers and stego objects can be distinguished with arbitrarily low error rate.

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.