Stefan Niculae
Alexandra Voda

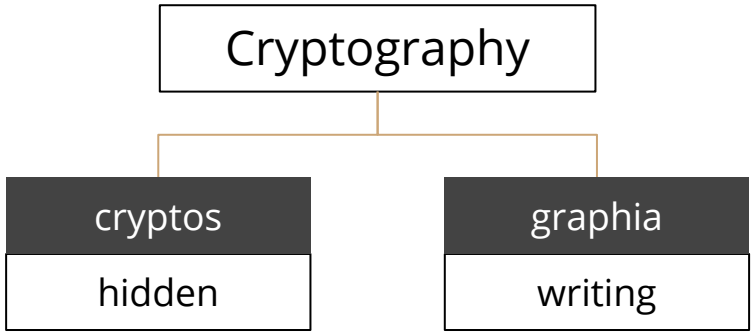
STEGANOGRAPHY

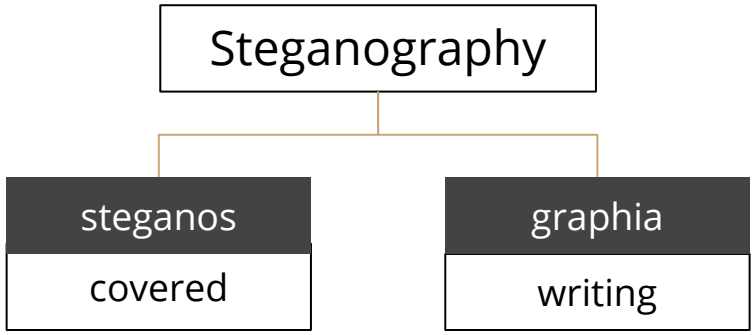
Hiding in plain sight

What is Stegano?

- Conceal a message within another
- Imperceptible, unless you know where to look
- Intuitively used since Antiquity, formalized in modern times

vs Crypto





- Writing in a secret code
- Alter message, without hiding
- Obfuscates data
- Defends

- Change message meaning
- Same message, hidden info
- Obfuscates the communication
- Doesn’t attract interest

vs Watermarking

Watermarking

- Unremovable message
- one:many communication

Steganography

- Undetectable message
- one:one communication

Since everyone can read, encoding text in neutral sentences is doubtfully effective

Since everyone can read, encoding text in neutral sentences is doubtfully effective

Secret inside

Secret
inside

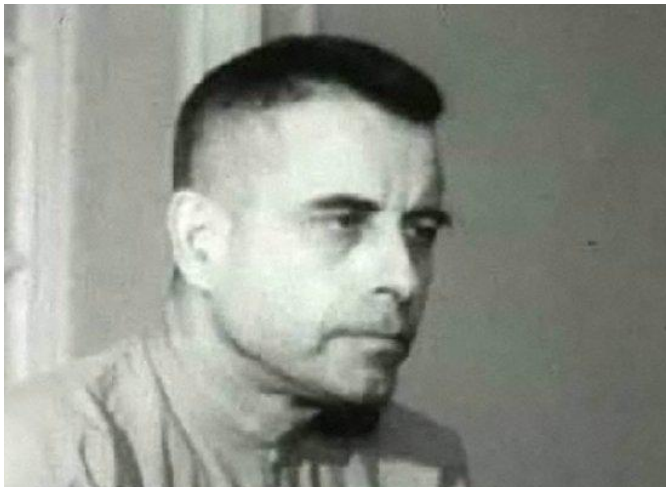
Historical Example

- Documented by Herodotus in 440BC:
  1. Shave slave head
  2. Mark Persian attack plans on scalp
  3. Wait for hair to regrow
  4. Send through enemy territory
  5. Greek receiver shaves head

- Drawbacks:
  - Waiting time for hair to grow
  - Limited message size

Recent Historical Examples

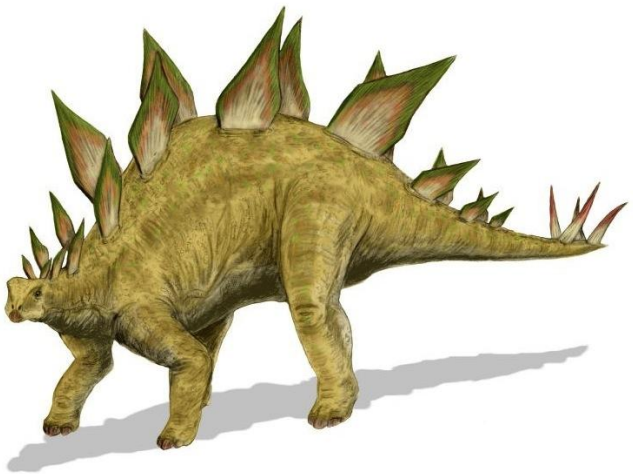
- 300 BC*:* Invisible ink, visible only when heated
- WW I*:* Message knitted into a piece of yarn worn by a courier
- WW II*:* Tiny “micro-dots” printed over newspapers
- 1966*:* Prisoner of war blinking “**TORTURE**” in morse code



Modern Examples

- Large cover (relative to secret) means easier to hide
- Media files are ideal
  - Text, audio, images, video
- Unicode characters that look like standard ones
- Set every 100th pixel of an image to an ASCII code

- Ignored sections of the file
- Delays in network packets sent



Steganosaurus: a covered lizard

Properties

- Imperceptibility
  - A measure of the amount of distortion to the cover
  - Stego-medium indistinguishable from stego-cover
- Embedding capacity
  - Amount of data that can be hidden in a cover, compared to the size of the cover
- Undetectability
  - Maintain statistical properties of the cover file
- Robustness
  - Retain the hidden data even after the cover has been subjected to various changes
  - Should be difficult to destroy the secret information without destroying the cover
- Tamper resistance
  - Resistance to the attempts of altering the hidden data

Image Stegano

- Most popular type of Stegano
- Large embedding capacity
- Methods focus on noise manipulation
  - Hard to detect by the human eye
  - Circumvent statistical methods masquarading as randomness

Text Stegano

- Most difficult type of stegano
- Low embedding capacity
- Dependent on the used language
- Requirements:
  - Letter frequencies resembling a natural language
  - Most words should be found in a good dictionary
  - Syntactically correct sentence

Text Methods

- Format Based
  - Uses:
    - Word shift coding: shift horizontal location of word in text
    - Line shift
    - Unicode characters
  - Vulnerable to OCR or retyping

- Semantic Based
  - Based on linguistic transformations
    - Word synonyms
    - Word deletion
  - Can alter sentence meaning

The idea is a **powerful** one → The idea is a **potent** one
This computer is **powerful** → This computer is **potent**

Word	Synonym
big	large
find	observe
familiar	popular
chilly	cool

Mimic Functions

start → **adj noun tense verb**
adj → the **size** | a **size**
size → tiny | small | large | big
noun → saw | ladder | truth | boy
tense → is | was
verb → waiting | standing

cover: The large ladder was waiting
hidden:

Mimic Functions

start → **adj noun tense verb**
adj → the **size** | a **size** – first production
size → tiny | small | large | big
noun → saw | ladder | truth | boy
tense → is | was
verb → waiting | standing

cover: The large ladder was waiting
hidden: 0

Mimic Functions

start → **adj noun tense verb**
adj → the **size** | a **size**
size → tiny | small | large | big – third production
noun → saw | ladder | truth | boy
tense → is | was
verb → waiting | standing

cover: The large ladder was waiting
hidden: `010`

Mimic Functions

start → **adj noun tense verb**
adj → the **size** | a **size**
size → tiny | small | large | big
noun → saw | ladder | truth | boy – second production
tense → is | was
verb → waiting | standing

cover: The large ladder was waiting
hidden: `01001`

Mimic Functions

start → **adj noun tense verb**
adj → the **size** | a **size**
size → tiny | small | large | big
noun → saw | ladder | truth | boy
tense → is | was – second production
verb → waiting | standing

cover: The large ladder was waiting
hidden: `010011`

Mimic Functions

start → **adj noun tense verb**
adj → the **size** | a **size**
size → tiny | small | large | big
noun → saw | ladder | truth | boy
tense → is | was
verb → waiting | standing – first production

cover: The large ladder was waiting
hidden: `0100111`

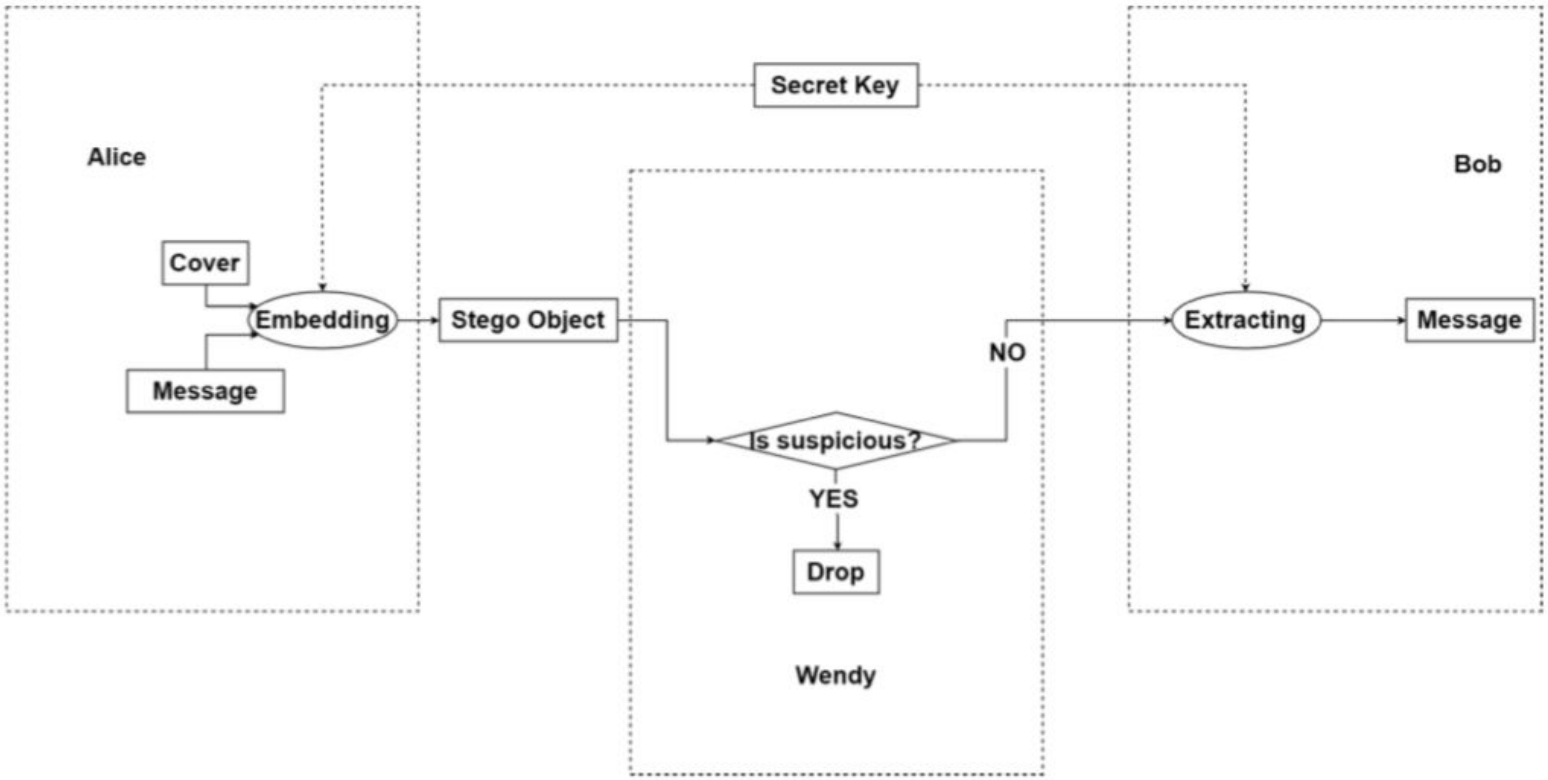
Implemented Techniques

Methods Categories

- Pure (no key)
- Secret key
- Public key

Stego-medium = cover + secret + key

Block Scheme



Text Demo

Zero-width unicode characters:

Imperceptibly hiding a text in a regular message

Unicode Steganography with Zero-Width Characters

Cover Text: Clear (length: 29)

This is an innocuous message.

Hidden Text: Clear (length: 15)

Secret message!

Encode »

Steganography Text: Clear (length: 0)

Unicode Steganography with Zero-Width Characters

Cover Text: Clear (length: 29)

This is an innocuous message.

Hidden Text: Clear (length: 0)

« Decode

Steganography Text: Clear (length: 149)

This is an innocuous message.

Unicode Steganography with Zero-Width Characters

Cover Text: Clear (length: 29)

This is an innocuous message.

Hidden Text: Clear (length: 15)

Secret message!

« Decode

Steganography Text: Clear (length: 149)

This is an innocuous message.

Remarks

Pros:

- High embedding capacity
- High imperceptibility
- Easy to implement

Cons:

- Very vulnerable to programs that remove blank spaces in text
- Vulnerable to retyping
- Increases the length of the cover image

Image Demo

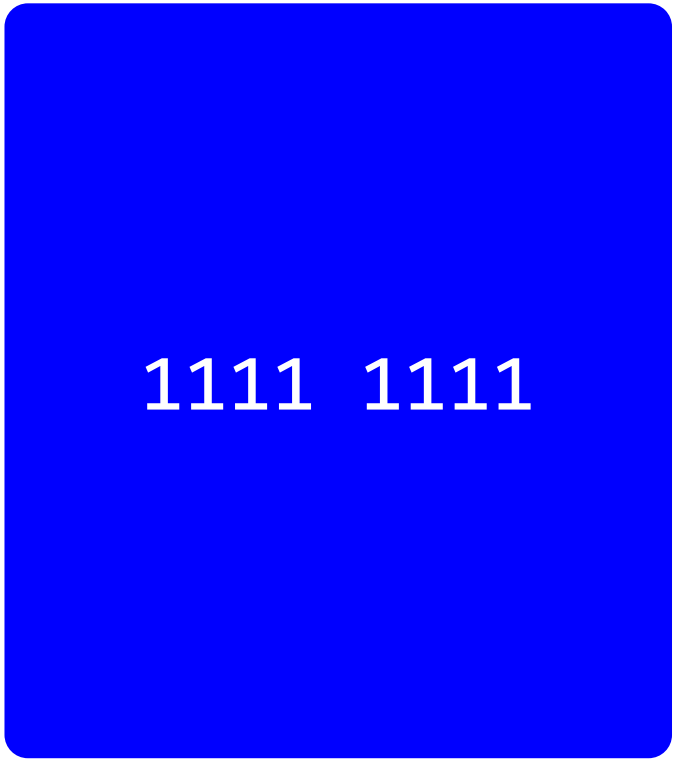
LSB insertion:

Hiding an image in the Least Significant Bits of another

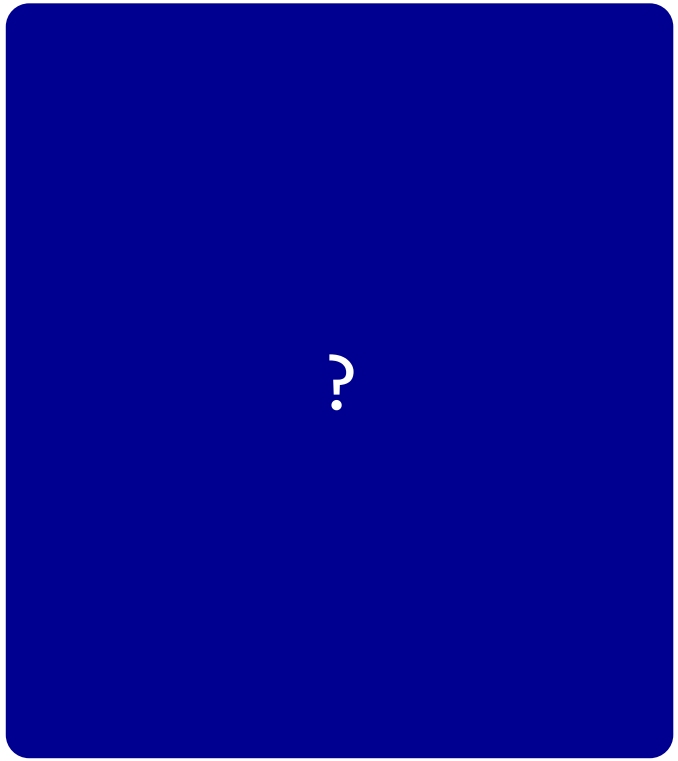
LSB Insertion

- Objective: changes to carrier (injecting payload) to be visually (and statistically) negligible
- Images are a good carrier
  - Analog signal digitization
  - Lossy compression error
- Bitmap (`.bmp`) uses 8 bits for each RGB channel
  - Slight changes undetectable by human eye (256 levels)
- Best in noisy areas
  - Allows payload to blend in with natural color variation
  - Wide, solid areas magnify any added noise

Human Eye Test

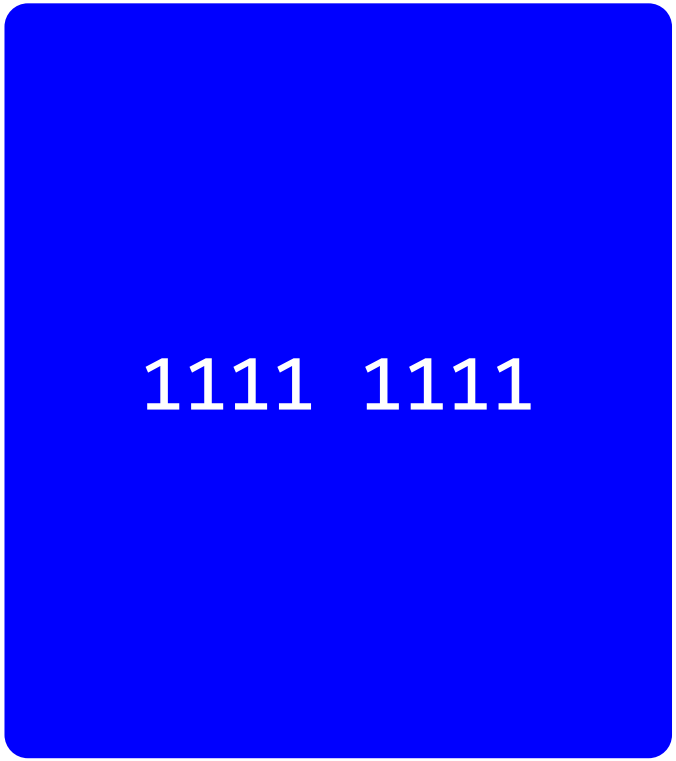


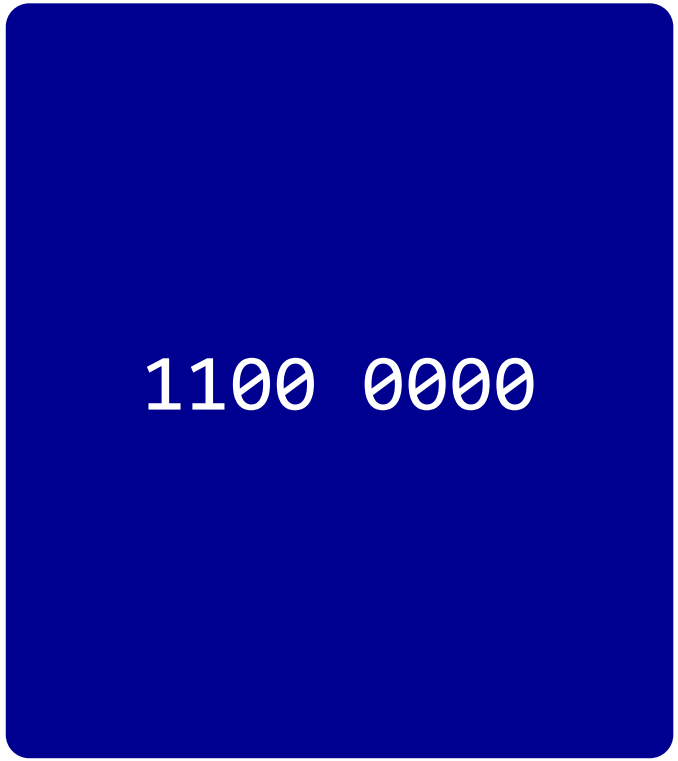
Pure blue



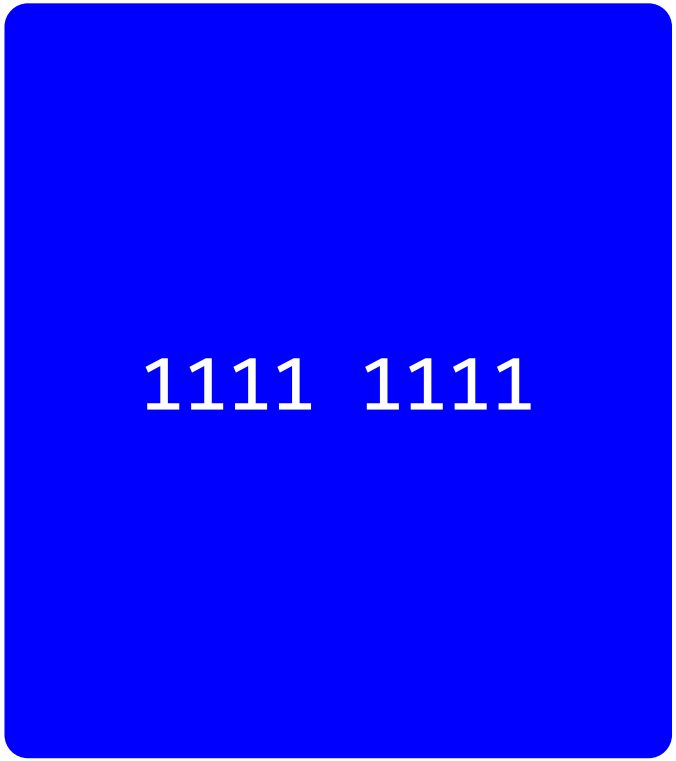
How many bits differ?

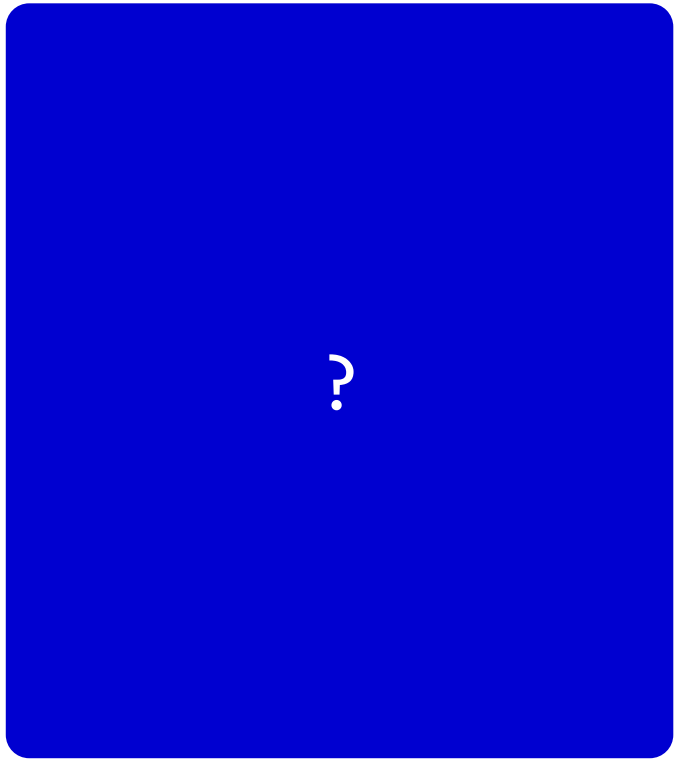
Human Eye Test





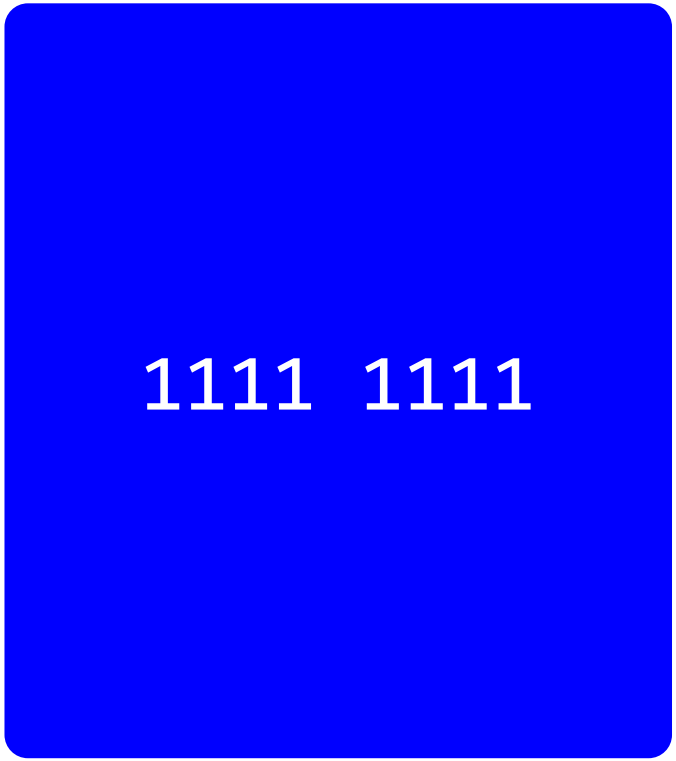
Human Eye Test

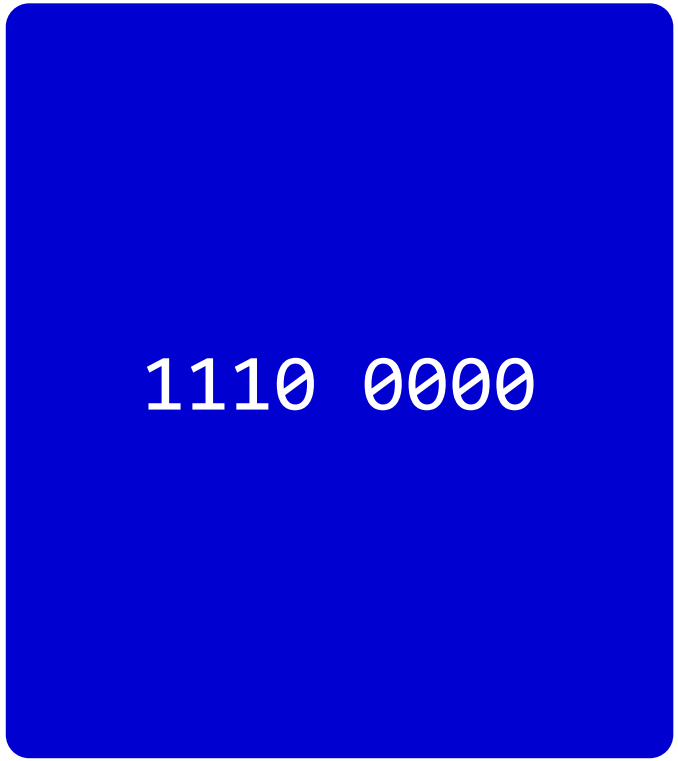




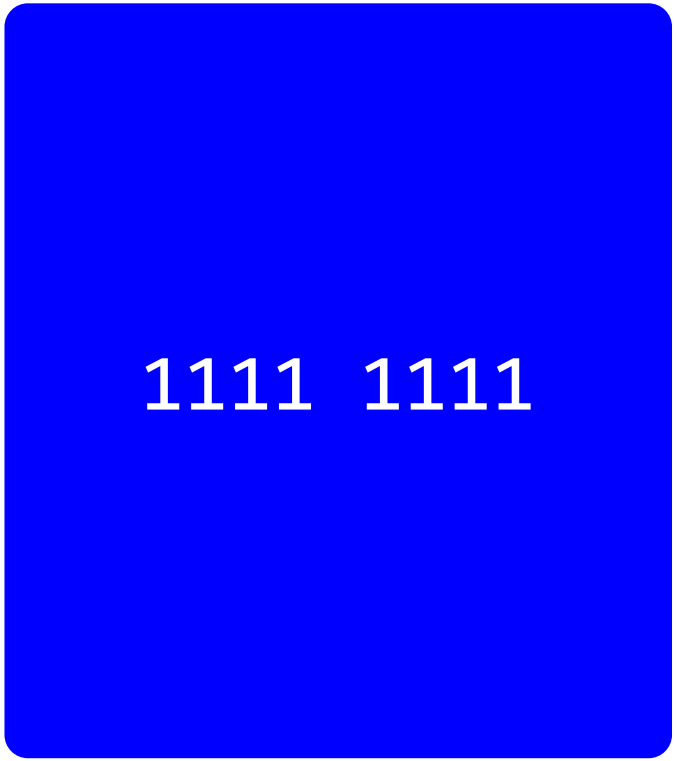
Same color?

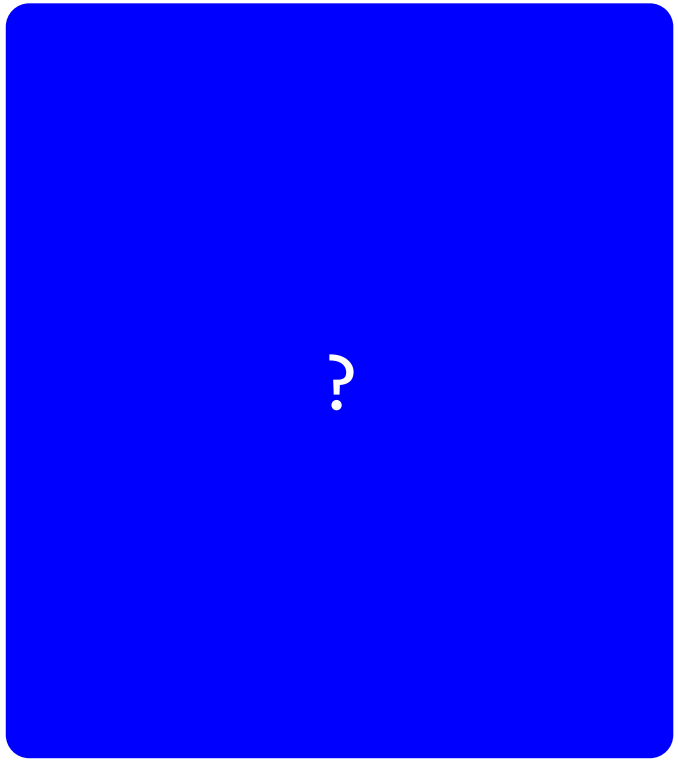
Human Eye Test





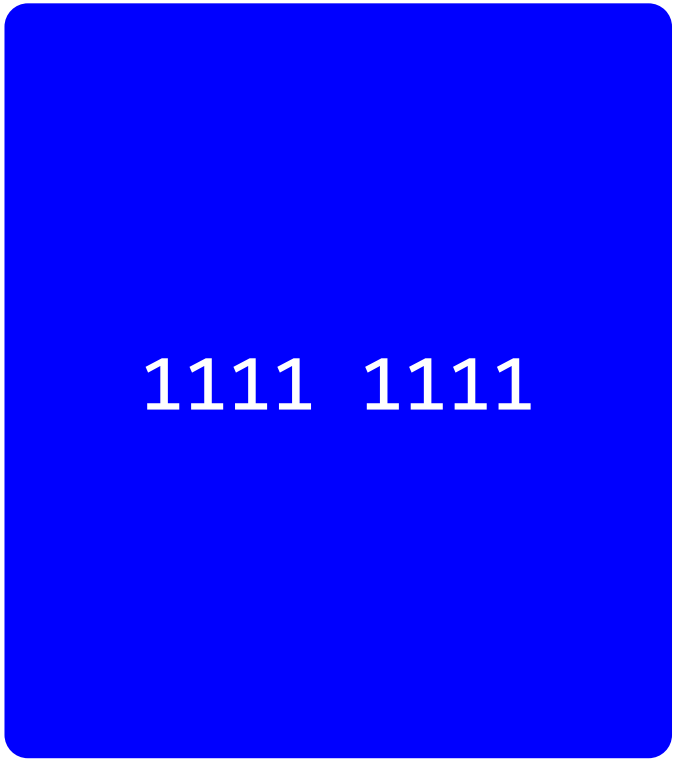
Human Eye Test

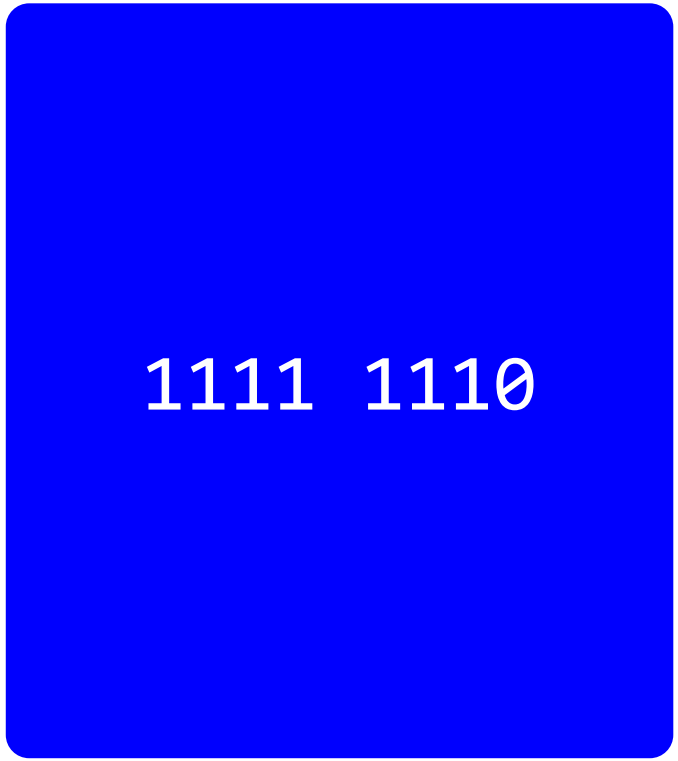




What about now?

Human Eye Test





Cover

Cover

encode

Secret

Message

Message

decode

Recovered

Cover

encode

Secret

Message

decode

Recovered

Robustness

- Very vulnerable to transformations
  - Geometrical
  - Blurring
- Compression destroys it
  - LSB insertion exploits what lossy compression algorithms (JPEG) rely on: low human eye sensitivity to added noise
- Useless for watermarking
  - Doesn’t withstand destruction attempts
  - Doesn’t translate well to print
- Better suited for stegano
  - Robustness not so important
  - High data rate

Solution

- Hinder malicious attempts at reading secret
- Randomize the placement of bits
  - Using a cryptographical random function (scattering)
- Can’t decode without seed

Steganalysis

- Statistical analysis to detect hidden messages:
- Split image into blocks
- Compute average value of LSBs in each block
- Random data should have around 0.5

Closing Words

Recent Advances

- ML-based attacks
  - Embedding location finder for image steganography
  - Employ transfer learning for CNNs
- ML-based detection
  - Continuation of statistical steganalysis methods
  - Alleviates the need for feature engineering
- Domain at the intersection of:
  - Information Theory
  - Data compression
  - Cryptography

Applications

- *Domain still young*
- Military
  - Exchange attack plans without the enemy catching on
- Illegal trades
  - Use code words in written communication to obfuscate from authorities
- Piracy
  - Hide activation keys in pictures of game covers
- IP Protection
  - Hide fictional places in proprietary maps to identify duplication attempts
- Anti-counterfeit
  - Hidden information impossible to reproduce when home printing

Thank you!

References

1. C.P. Sumathi et al, **A Study of Various Steganographic Techniques Used for Information Hiding**, International Journal of Computer Science & Engineering Survey, 2013
2. A. Odeh & K. Elleithy, **Steganography in Text by Merge ZWC and Space Character**, 2013
3. R. Kumar et al, **An efficient text steganography scheme using Unicode Space Characters**, 2015
4. D. Salomon, **Data Hiding in Text. In: Data Privacy and Security**, Springer, 2003
5. H.A. Atee et al, **Extreme learning machine based optimal embedding location finder for image steganography**, 2017
6. D. Dumitrescu et al, **Steganography techniques**, 2017
7. S.M. Thampi, **Information Hiding Techniques: A Tutorial Review**, LBS College of Engineering, 2008
8. K. Thangadurai & G.S. Devi, **An analysis of LSB based image steganography techniques**, International Conference on Computer Communication and Informatics, 2014