Steganography

COMP620

“A picture can hide as much as it reveals.”

Alexandra Petri

Steganography

- **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
- A form of security through obscurity

Kerckhoffs's principle

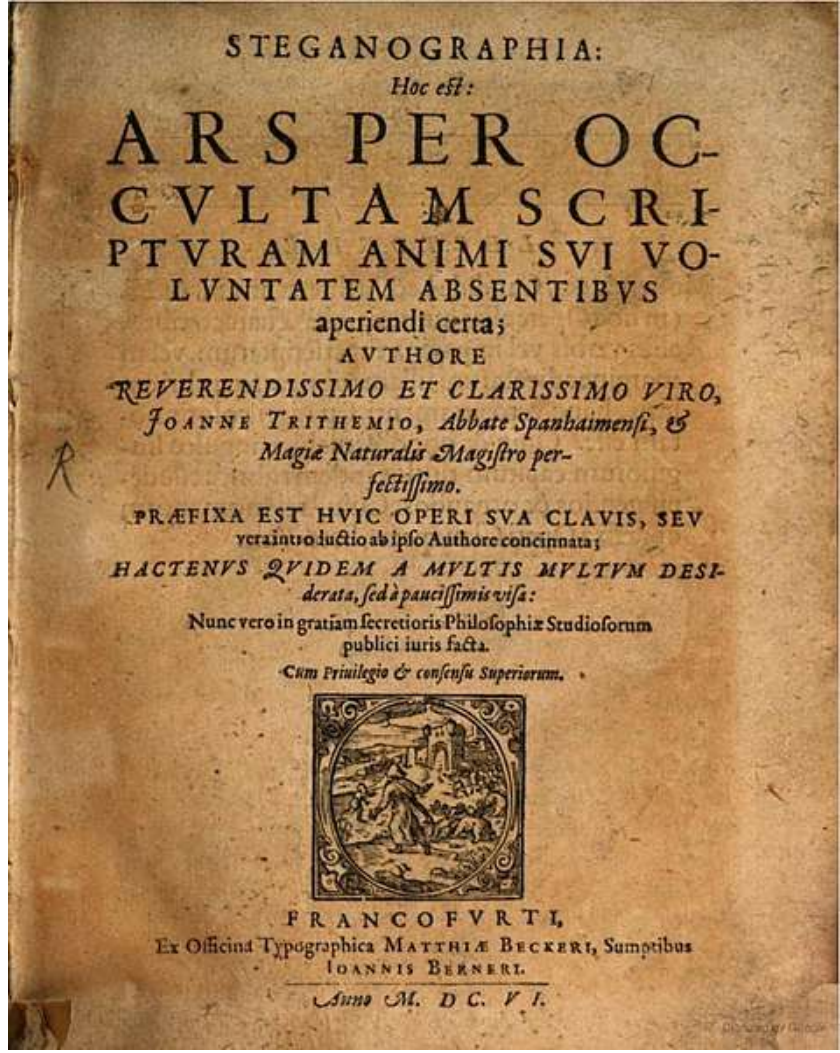
- Stated in 1883 Auguste Kerckhoffs
- A cryptosystem should be secure even if everything about the system, except the key, is public knowledge
- Claude Shannon repeated the principle as “the enemy knows the system”

- Steganography is often based on “security through obscurity”

Wikipedia

Origin of Steganography

- The first recorded use of the term was in 1499 by Johannes Trithemius in his Steganographia, a treatise on cryptography and steganography, disguised as a book on magic
- Published in 1606 and placed on the Index Librorum Prohibitorum in 1609
- With the encryption key, the books are actually concerned with cryptography and steganography


STEGANOGRAPHIA:
Hoc eſt:
ARS PER OC-
CVLTAM SCRI-
PTVRAM ANIMI SVI VO-
LVNTATEM ABSENTIBVS
aperiendi certa;
AVTHORE
REVERENDISSIMO ET CLARISSIMO VIRO,
IOANNE TRITHEMIO, Abbate Spanhaimenſi, &
Magiæ Naturalis Magiſtro per-
fectiſſimo.
PRÆFIXA EST HVIC OPERI SVA CLAVIS, SEV
vera intuitio ductio ab ipſo Authore concinnata;
HACTENVS QVIDEM A MVLTIS MVLTVM DESI-
derata, ſed à pauciſſimis viſa:
Nunc vero in gratiam ſecretioris Philoſophiæ Studioſorum
publici iuris facta.
Cum Priuilegio & conſenſu Superiorum.
FRANCOFVRTI,
Ex Officina Typographica MATTHIÆ BECKERI, Sumptibus
IOANNIS BERNERI.
Anno M. DC. VI.

Wikipedia

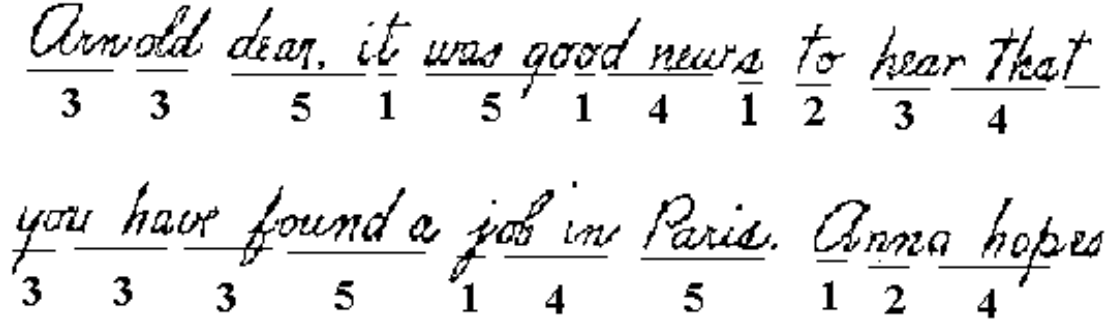
Physical Steganography

- Messages written in Morse code on yarn and then knitted into a piece of clothing worn by a courier
- Messages tattooed on the head of a courier where the hair will grow back and cover it
- Hidden messages on paper written in secret inks, under other messages or on the blank parts of other messages
- Messages written on envelopes in the area covered by postage stamps

Wikipedia

Coded Message

By counting the number of letters between those letters whose "tails" point upwards, we get the following sequence of numbers.



33 51 51 41 23 43 33 51 45 12 43 24 11 34 34 11 34 34 42 33 11 44 42 43 33

Now use the following table to decrypt this message:

	1	2	3	4	5
1	A	F	L	Q	V
2	B	G	M	R	W
3	C	H	N	S	X
4	D	I=J	O	T	Y
5	E	K	P	U	Z

To get: NEEDMONEYFORASSASSINATION

Steganography and Encryption

- You can use both steganography and encryption to send a hidden encoded message
- A simple method is to encrypt the message and then hide it

Text Steganography

- Adding unnecessary but not displayed mark up in HTML

- Using non-printing Unicode characters Zero-Width Joiner (ZWJ) and Zero-Width Non-Joiner (ZWNJ). They do not effect Roman letters and are not displayed

Hidden Information in Files

- PNG files are a series of “chunks” with a specified chunk type
- There is a list of defined chunk types
- A new chunk type is ignored providing upward compatibility
- You can add a new chunk to hide data in the file
- It does not impact the display of the file

Steganography in Images

- Removing all but the last 2 bits of each color from the tree picture component produces an almost completely black image. Making that image 85 times brighter produces the cat image.

Wikipedia

Review of Graphic File Formats

- **Vector** – Pictures are drawn as a series of lines. The hardware draws a line from point A to point B.
- **Raster** – Pictures are represented as a matrix of picture elements or pixels

Which format is more popular?

A. Raster
B. Vector
C. About the same

Picture Size and Resolution

- Screen resolution is measured in horizontal and vertical pixels per inch.
- A good monitor can display about 90 pixels per inch.
- A good printer can print about 600 to 1200 pixels per inch.
- True size of a picture depends on the number of pixels and the device.

Pixel

- Each pixel has a color.
- Monochrome pictures have only two colors.
- The color is usually represented as three numbers that are the intensity of the three primary colors, Red, Green and Blue (RGB)
- The pixel represents both the color and the intensity or darkness of the pixel.

Bits / Pixel

- The number of bits it takes to represent a pixel depends on the number of possible colors.
- Black and white requires only 1 bits per pixel.
- Full color pictures require 24 bits per pixel
- Simple graphics can use 4 or 8 bits per pixel

Graphics Formats

- There are many formats for graphical data
  - BMP
  - JPEG
  - GIF
  - TIFF
  - PNG
  - etc.

Bit Mapped File

- BMP is a very simple format for graphics
- It was frequently by Microsoft systems before PNG
- Each pixel is stored as a number. The number of bits per pixel is determined by the number of different colors

BMP format

BMP File Header	Stores general information about the BMP file.
Bitmap Information (DIB header)	Stores detailed information about the bitmap image.
Color Palette	Stores the definition of the colors being used for indexed color bitmaps.
Bitmap Data	Stores the actual image, pixel by pixel.

BMP Efficiency

- Each pixel in a BMP file is represented by a number.
- There is no compression.
- It does not matter how “complex” the image, the file size is determined by the number of pixels or the size of the picture.

JPEG files

- The JPEG format was created by the Joint Photographic Experts Group
- JPEG is a commonly used method of compression for photographic images
- JPEG uses lossy compression. Some image quality is lost
- You can control the level of compression and therefore the image quality
- JPEG uses a Discrete cosine transform for compression

JPEG Example

- Q = 100
- 81.3 KB
- 219,726 bytes if BMP
- 37% of BMP

JPEG Example

- Q = 50
- 14.7 KB
- 18% of Q100
- 6.7% of BMP

JPEG Example

- Q = 25
- 9.32 KB
- 11.5% of Q100
- 4.2% of BMP

JPEG Example

- Q = 10
- 4.67 KB
- 5.7% of Q100
- 2.1% of BMP

JPEG Example

- Q = 1
- 1.48 KB
- 1.8% of Q100
- 0.67% of BMP

JPEG Artifacts

Appropriate Use

- BMP files are not distorted in any way, but they are large
- JPEG works well for photographs
- PNG works well for diagrams with no data compression loss
- GIF works well for diagrams or charts

Hiding Data in Images

- Data can easily be hidden in the least significant bits of the pixels of a complex graphics file
- One bit of text can replace the lowest bit of one (or all) of the colors
- Changing only the least significant bit is very hard to detect visually
- Statistically the text actually changes the bit only 50% of the time

Example Images 1 blue bit / pixel

Original image

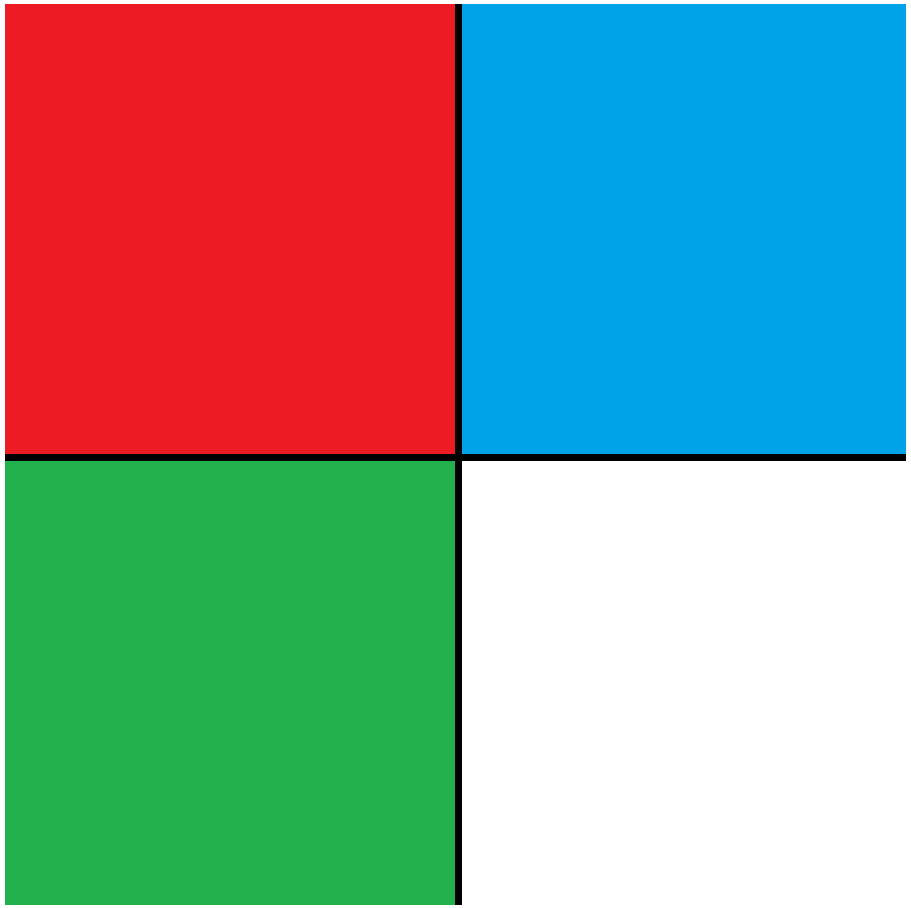
with hidden text

Example Images 8 bit / pixel

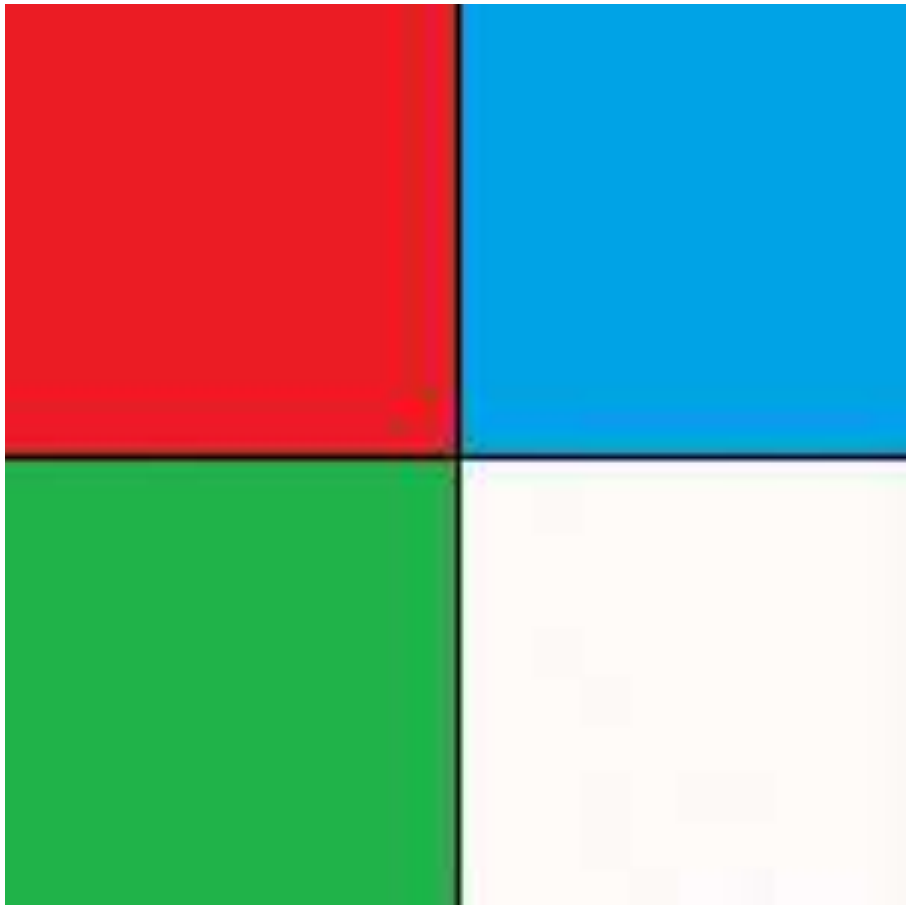
Original image

with hidden text

Example Images 8 bit / pixel



Original image



with hidden text

Which file format would be the worst for hiding text in the LSB of the image?

A. BMP
B. GIF
C. JPEG
D. PNG

Legitimate Use

- Some brands of color printers add tiny yellow dots giving the printer serial number plus the time and date
- Digital watermarking can be used to identify ownership
  - May be visible or invisible
  - Robust watermarking can survive certain transformations to the file

Chaffing and Winnowing

- Chaffing and winnowing is a cryptographic technique to achieve confidentiality without using encryption when sending data over an insecure channel
- The technique was conceived by Ron Rivest
- Different from encryption or steganography
- Appears to be useful in places that do not allow encryption

Devise a plan in a group of 3-4 students

- Phred the Terrorist wants to send messages to his comrades on a daily basis
- He manages a web site
- The NSA, CIA, FBI and XYZ suspect him and look at his website frequently
- Devise a plan by which he can send messages to his comrades without the intelligence agencies knowing what he is sending

Steganalysis

- Intentionally, steganography can be hard to recognize
- If you have a copy of the original image, a simple comparison will indicate if the file has been modified
- Steganography can change the frequency of words, bytes or bits
  - If a message is hidden in the least significant bits of an image, these bits may be more random than other bits

Compression and Steganography

- Lossy compression algorithms lose some of the information
- Since a steganographic message is usually hidden in the least significant portion of a file, it is most likely to be lost in compression

Steganography tools

- Wikipedia lists 23 steganography tools
- Experiments showed you can hide a significant amount of data in a JPEG image that is visually undetectable

original

With first page of syllabus in LSB

Countermeasures

- Content Threat Removal (CTR) replaces original messages with manufactured equivalents
- Placed between the sender and receiver, the CTR system might send the receiver a message with the same text, but not the original file