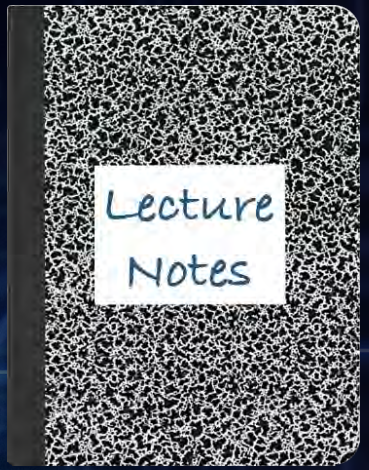
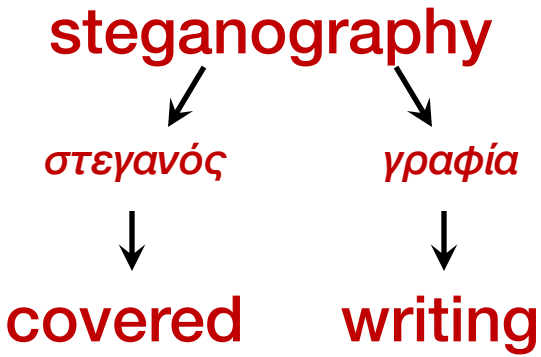
CS 419: Computer Security



Week 14: Hiding Communication
Part 1: Steganography

Paul Krzyzanowski



The art of secret (hidden) writing

Steganography

Art and science of communicating in a way that hides the existence of a message

Signal or pattern imposed on content

- Persistent under transmission
- Not encryption – original image/file is intact
- Not fingerprinting
  - Fingerprinting leaves separate file describing contents

Classic techniques

- Invisible ink (1st century AD - WW II)
- Tattooed message on head
- Overwrite select characters in printed type in pencil
- Pin punctures in type
- Microdots (early 20th century)
- Newspaper clippings, knitting instructions, XOXO signatures, report cards, …

Motivation

Steganography received little attention in computing until recently

- **Industry's desire to protect copyrighted digital work**
  - Detect counterfeit, unauthorized presentation, embed key, embed author ID
- **Covert way to distribute malware**
  - Embed in a JPEG file, which would raise no suspicion when downloaded
- **Covert way to exfiltrate data**
  - Upload harmless images with embedded data
  - **Network steganography**

Steganography ≠ Copy protection ≠ Cryptography

Code hidden in photo, files stolen: Upstate man stole GE technology to try to help China

syracuse.com

Anne Hayes • April 1, 2022

Schenectady, N.Y – A Schenectady County man who hid data in the code of a digital photograph of a sunset was convicted Thursday of conspiracy to commit economic espionage against General Electric in order to benefit the Chinese government.

Xiaoqing Zheng, 59, was originally accused of stealing GE trade secrets regarding turbine technology and planning to give the information to contacts in China, according to federal court documents.

Although the jury convicted Zheng, a U.S. citizen, of conspiracy to commit economic espionage, they could not reach a unanimous decision regarding the charge of economic espionage, according to a news release from the U.S. Attorney's Office of the Northern District of New York.

…

In 2018, Zheng used a means of hiding data within the code of another file to conceal 40 files in the code of a digital photograph of a sunset. He then emailed the photograph file to his personal email account, according to court documents.

URhttps://www.syracuse.com/crime/2022/04/code-hidden-in-photo-files-stolen-upstate-man-stole-ge-technology-to-try-to-help-china.htmlL

INDEPENDENT

Isis and al-Qaeda sending coded messages through eBay, pornography and Reddit

Kashmira Gander – Monday 2 March 2015 19:29 GMT

Isis and al-Qaeda members are communicating with each other via coded messages hidden on websites including eBay, Reddit, and inside pornographic photos, according to a new book.

Gordon Thomas, who has sources inside Israel's Mossad spy agency, has revealed that the organisation's cyber warfare department's most skilled cryptologists mastered a technique known as steganography, which is used to to conceal secret information within a digital file.
The spies found that al-Qaeda had used the technique to hide messages in goods offered for sale on eBay, according to extracts from *Gideon's Spies: The Secret History of the Mossad* published by *The New York Post*.

Code hidden in photo, files stolen: Upstate man stole GE technology to try to help China

syracuse.com

Anne Hayes • April 1, 2022

Schenectady, N.Y – A Schenectady County man who hid data in the code of a digital photograph of a sunset was convicted Thursday of conspiracy to commit economic espionage against General Electric in order to benefit the Chinese government.

Xiaoqing Zheng, 59, was originally accused of stealing GE trade secrets regarding turbine technology and planning to give the information to contacts in China, according to federal court documents.

Although the jury convicted Zheng, a U.S. citizen, of conspiracy to commit economic espionage, they could not reach a unanimous decision regarding the charge of economic espionage, according to a news release from the U.S. Attorney's Office of the Northern District of New York.

…

In 2018, Zheng used a means of hiding data within the code of another file to conceal 40 files in the code of a digital photograph of a sunset. He then emailed the photograph file to his personal email account, according to court documents.

URhttps://www.syracuse.com/crime/2022/04/code-hidden-in-photo-files-stolen-upstate-man-stole-ge-technology-to-try-to-help-china.htmlL

BLEEPINGCOMPUTER

New SteganoAmor attacks use steganography to target 320 orgs globally

By Bill Toulas

April 15, 2024 04:31 PM 0

A new campaign conducted by the TA558 hacking group is concealing malicious code inside images using steganography to deliver various malware tools onto targeted systems.

Steganography is the technique of hiding data inside seemingly innocuous files to make them undetectable by users and security products.

TA558 is a threat actor that has been active since 2018, known for targeting hospitality and tourism organizations worldwide, focusing on Latin America.

The group's latest campaign, dubbed "SteganoAmor" due to the extensive use of steganography, was uncovered by Positive Technologies. The researchers identified over 320 attacks in this campaign that affected various sectors and countries.

Null Cipher (concealment cipher)

- **Hide message in a large amount of irrelevant data**
- **Agreed technique for extracting content**
  - First letter of each word, N^{th} letter of each word
  - Some specific pattern to define which words or letters are significant (e.g., 4-5-5-4 words)

Null Cipher (concealment cipher)

Sent by a German spy in WWI:

APPARENTLY NEUTRAL'S PROTEST IS THOROUGHLY DISCOUNTED AND IGNORED. ISMAN HARD HIT. BLOCKADE ISSUE AFFECTS PRETEXT FOR EMBARGO ON BYPRODUCTS, EJECTING SUETS AND VEGETABLE OILS.

Reference: David Kahn, *The Codebreakers*, p. 521

Null Cipher (concealment cipher)

The 2nd letter of each word contains the message

APPARENTLY NEUTRAL'S PROTEST IS THOROUGHLY DISCOUNTED
AND IGNORED. ISMAN HARD HIT. BLOCKADE ISSUE AFFECTS
PRETEXT FOR EMBARGO ON BYPRODUCTS, EJECTING SUETS AND
VEGETABLE OILS.

PERSHING SAILS FROM NY JUNE I

(BTW, the intelligence was inaccurate: Pershing sailed May 28)

By WWII, not used by spies but by regular people trying to beat the censor.

Reference: David Kahn, *The Codebreakers*, p. 521

BBC News 27 April 2006

Judge creates own Da Vinci code

The judge who presided over the failed Da Vinci Code plagiarism case at London's High Court hid his own secret code in his written judgement.

Seemingly random italicised letters were included in the 71-page judgement given by Mr Justice Peter Smith, which apparently spell out a message.

Mr Justice Smith said he would confirm the code if someone broke it.

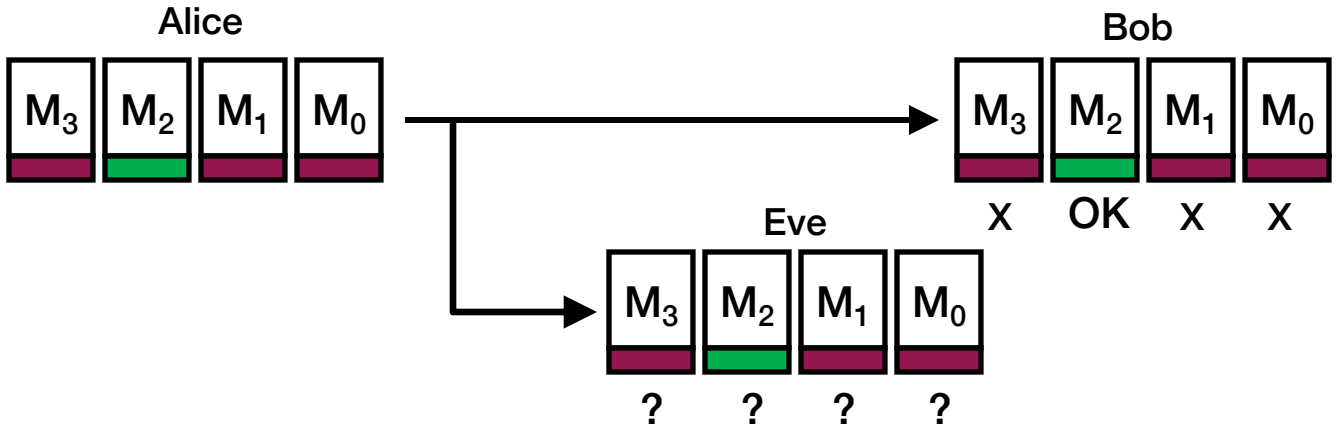
"I can't discuss the judgement, but I don't see why a judgement should not be a matter of fun," he said.

Italicised letters in the first few pages spell out **"Smithy Code"**, while the following pages also contain marked out letters.

http://news.bbc.co.uk/go/pr/fr/-/1/hi/entertainment/4949488.stm

Chaffing & Winnowing

- **Separate good messages from the bad ones**
  - Easy for someone who has the key, difficult for someone who does not
- **Stream of un-encoded messages with signatures or MACs**
  - Some signatures are bogus
  - Need to have the key to test



Steganography in images

Spatial domain

- Bit setting (LSB image steganography)
- Color separation

Frequency domain

- Apply FFT/DCT transform first
- Embed signal in select frequency bands
- Alter the least perceptible bits to avoid detection
  - But watch out: these are the same bits targeted by lossy image compression software (such as jpeg)

Metadata

- Add information the end of a PNG image's metadata or EXIF header

Just the picture

With the U.S. Declaration of Independence embedded

New Steganography Breakthrough Enables “Perfectly Secure” Digital Communications

SciTechDaily

University of Oxford • March 7, 2023

A group of researchers has achieved a breakthrough in secure communications by developing an algorithm that conceals sensitive information so effectively that it is impossible to detect that anything has been hidden.

The team, led by the University of Oxford in close collaboration with Carnegie Mellon University, envisages that this method may soon be used widely in digital human communications, including social media and private messaging. In particular, the ability to send perfectly secure information may empower vulnerable groups, such as dissidents, investigative journalists, and humanitarian aid workers.

…

Despite having been studied for more than 25 years, existing steganography approaches generally have imperfect security, meaning that individuals who use these methods risk being detected. This is because previous steganography algorithms would subtly change the distribution of the innocuous content.

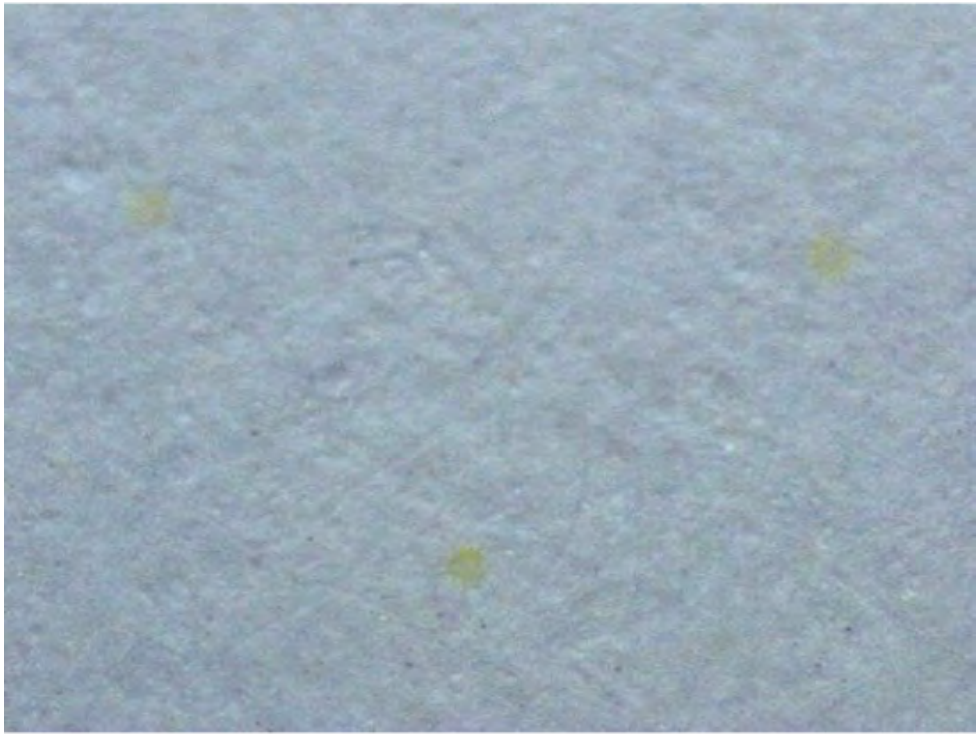
To overcome this, the research team used recent breakthroughs in information theory, specifically minimum entropy coupling, which allows one to join two distributions of data together such that their mutual information is maximized, but the individual distributions are preserved.

https://scitechdaily.com/new-steganography-breakthrough-enables-perfectly-secure-digital-communications/

Differences

There are differences – but you don’t notice them in the photo

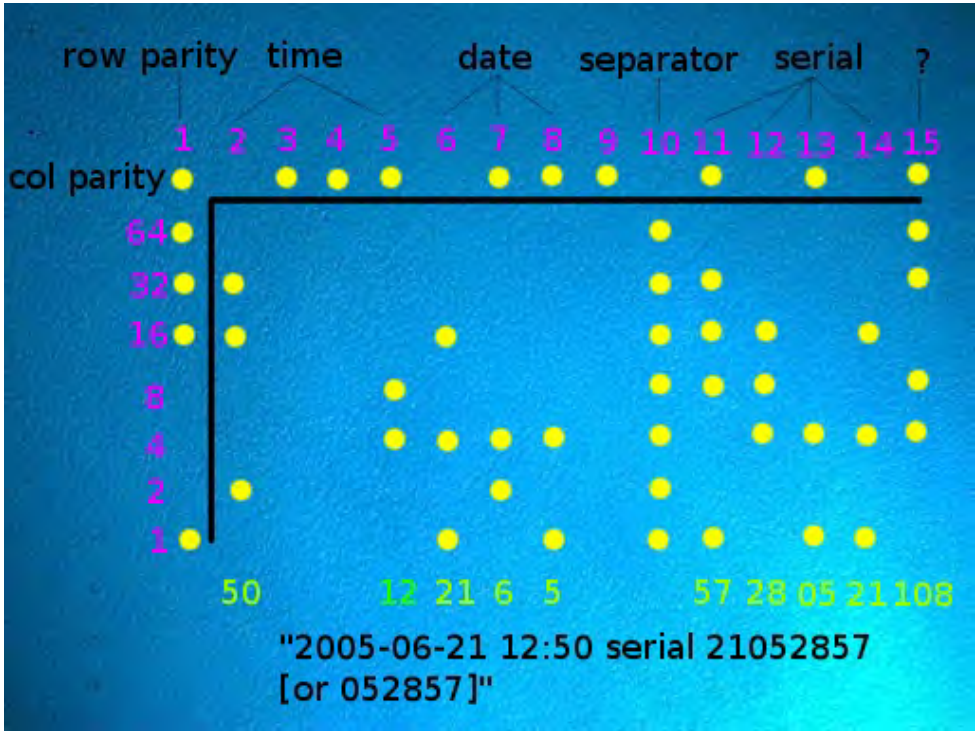
Machine ID codes in laser printers



See http://www.eff.org/Privacy/printers/

Machine ID codes in laser printers

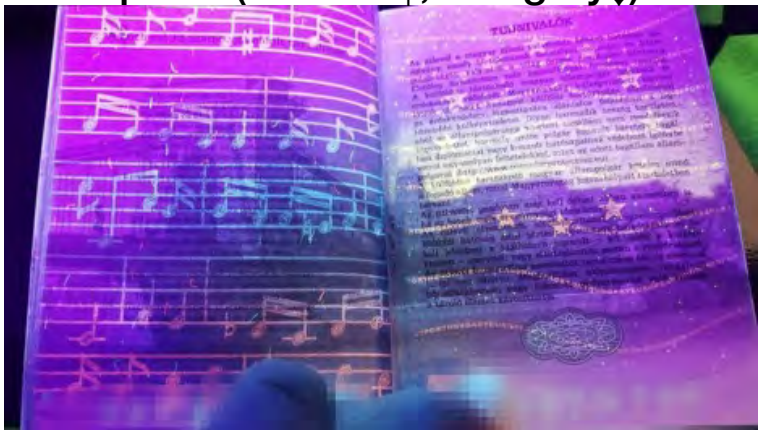
Machine ID codes in laser printers



Designed by Xerox to identify counterfeit currency and help track down counterfeiters

UV Watermarking

Passports (Canada↑, Hungary↓)



Also, currency, hand stamps for amusement park/club re-entry

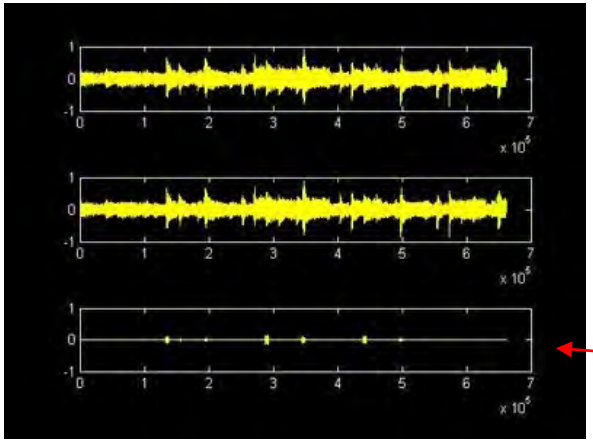
Video

- **Coding still frames - spatial or frequency**
- **Modify motion vectors**
- **Caption/subtitle data**
- **Audio channel**
- **Visible watermarking**
  - used by most networks (logo at bottom-right)
    This isn't steganography!

Audio

Perceptual coding

- Inject signal into areas that will not be detected by humans
- May be obliterated by compression



Amazon MP3 audio

Identifies where the song was purchased, not the user

Difference

Text

- Text lines shifted up/down
  (40 lines text $\Rightarrow 2^{40}$ codes)
- Word space coding
- Character encoding — minor changes to shapes of characters

Text

- Text lines shifted up/down
  (40 lines text $\Rightarrow 2^{40}$ codes)
- Word space coding
- Character encoding — minor changes to shapes of characters

Works only on "images" of text e.g., PDF, postscript

Text – non-visual

- **Embed zero-width non-printing characters**
  - Zero-width space or zero-width non-joiner (used to prevent ligature use)
- **White text on white background**
- **Overlapping objects**
- **PDF hidden pages**
- **HTML – invisible text designed to be picked up by search engines**
  - Non-rendered text (CSS element to not display text)
  - White on white
  - Obscured by other objects

Text-based stegranography

"Apparently, during the 1980's, British Prime Minister Margaret Thatcher became so irritated at press leaks of cabinet documents that she had the word processors programmed to encode their identity in the word spacing of documents, so that disloyal ministers could be traced."

– *Ross Anderson*
Stretching the Limits of Steganography

Watermarking vs. Steganography

Both techniques embed a message in data

Goal of steganography

- Intruder cannot detect the message
- Primarily 1:1 communication

Goal of watermarking

- Intruder cannot remove or replace the message (robustness is important)
- Doesn't have to be invisible
- Primarily 1:many communication

Watermarking applications

- **Copyright affirmation**
  - Embed information about owner
- **Copy protection rules**
  - Embed rights management information
  - But you need a trusted player
- **Content authentication**
  - Detect changes to the content

The End