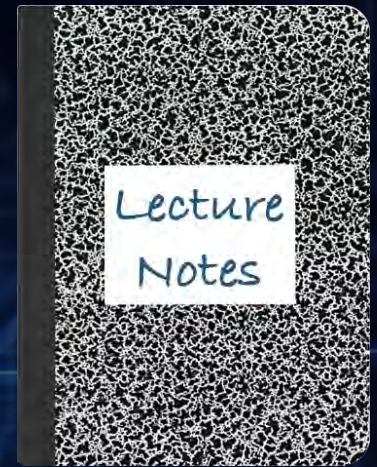


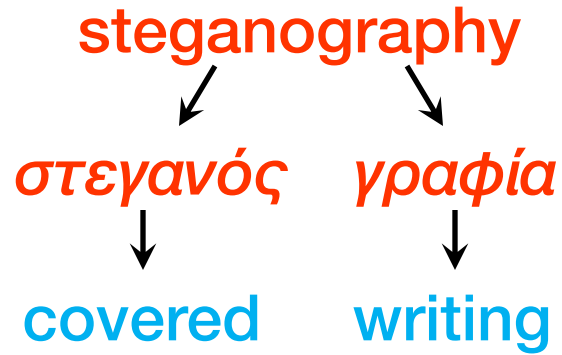
CS 419: Computer Security

Week 13: Hiding Communication Steganography

Paul Krzyzanowski



© 2025 Paul Krzyzanowski. No part of this content may be reproduced or reposted in whole or in part in any manner without the permission of the copyright owner.



The art of secret writing.

— *Oxford English Dictionary*

Steganography

Art and science of hiding information within another medium, such as an image, audio file, 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 is about adding distinct data to identify authorship (e.g., digital signatures)

Watermarking

- Encoding identifiable information (called a **watermark**), into content like images or video, to claim ownership or verify authenticity
- Applications
 - Copyright/creation affirmation
 - Embed information about owner
 - Label AI-created content
 - Copy protection rules
 - Embed rights management information
 - But you need a trusted player
 - Content authentication
 - Detect changes to the content

Watermarking vs. Steganography

Both techniques embed a message in data. Often used interchangeably

Goal of steganography: secrecy \Rightarrow conceal a message

- Intruder cannot detect there's a message in the content
- Primarily used for 1:1 communication

Goal of digital watermarking: identification \Rightarrow preserve a message

- Embed authorship or authenticity information into content
- Presence is intended: Doesn't have to be invisible
- The goal is to detect the watermark and preserve it
- Primarily used for 1:many communication

Classic techniques in watermarking

Change thickness of paper while wet via a pattern in the paper mold

- First used in 1282: identify paper maker or trade guild that made the paper
- The dry paper could be rolled again to create even thickness but varying density
- Later used in banknotes to enable detection of authentic banknotes
 - First used in 1661 issue of the Stockholms Banco



Watermarking: EURion constellation (Omron rings)

- Series of five small images are repeated throughout the banknote
- Software recognizes the pattern to prevent scanning
- Used by the Armenia, Australia, Canada, China, EU, India, Japan, Mexico, Switzerland, Thailand, UK, U.S, ... Zimbabwe



Machine ID codes in laser printers

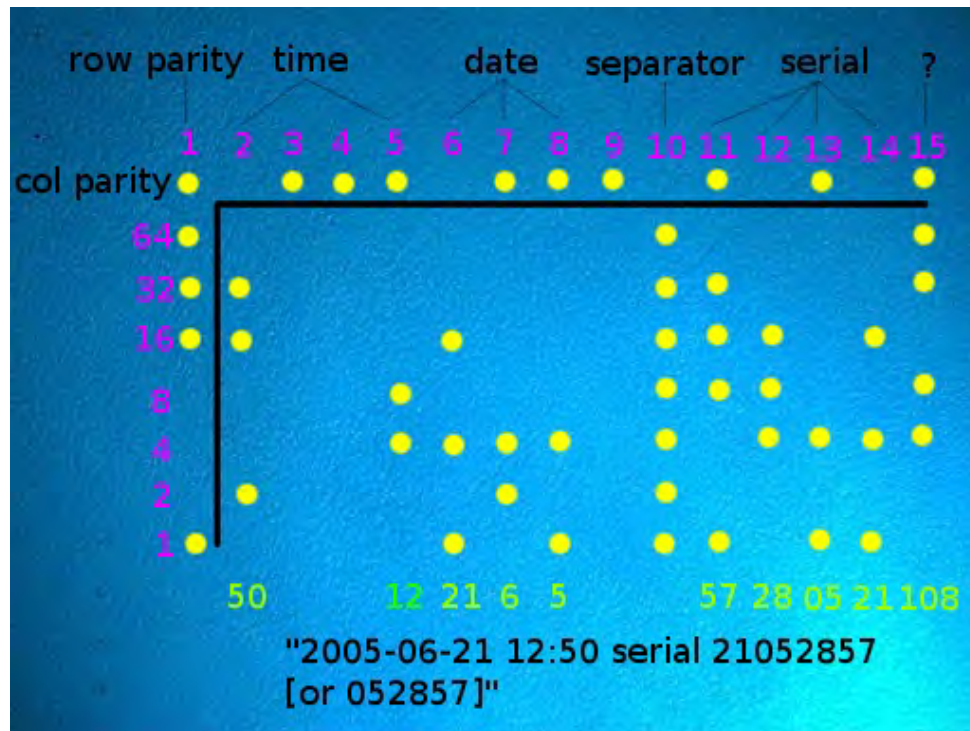


See <http://www.eff.org/Privacy/printers/>

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

Fragile vs. Robust Watermarking

Fragile watermarks: designed to break if the content is modified

- This is good for authentication and tamper detection
- Examples: currency, passports, entry tickets
 - These should no longer be valid if tampered

Robust watermarks: designed to survive transformation

- This is good for tracking content authorship and ownership
- Examples: photos, videos, audio, documents

Classic techniques in steganography

- **Invisible ink (1st century AD - WW II)**
- **Tattooed message on head**
- **Wax covered clay tablets**
 - Ancient Greece: engrave message in wood; cover with wax
- **Overwrite select characters in printed type in pencil**
- **Pin punctures in type**
- **Microdots (early 20th century)**
- **Newspaper clippings, knitting instructions, XOXO signatures, report cards, ...**

Null Cipher (concealment cipher)

- **Hide message in a large amount of irrelevant data**
- **Agreed technique for extracting content**
 - First letter of each word, Nth 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 1

(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

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.

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
 - This is mostly *watermarking* (more on that later)
- **Covert way to distribute malware – bypass detection**
 - E.g., 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 \neq Copy protection \neq Cryptography

Exfiltration via steganography

Trick content-inspecting firewalls by disguising malware and/or data

- **Data Exfiltration**

- Russian hackers hid malware in a legitimate update from SolarWinds, a popular IT management platform
 - Successfully breached Cisco, Intel, Microsoft, and U.S. government agencies
 - Used steganography to disguise stolen information as XML files

- **Malware Infiltration**

- 2019: A nation-state actor also used steganography to hide Windows DLLs (dynamic linked libraries) inside of WAV files to install a cryptomining app
- 2020: Attackers embedded skimming malware in SVG graphics in an attack on Dutch eCommerce platform Sansec

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.

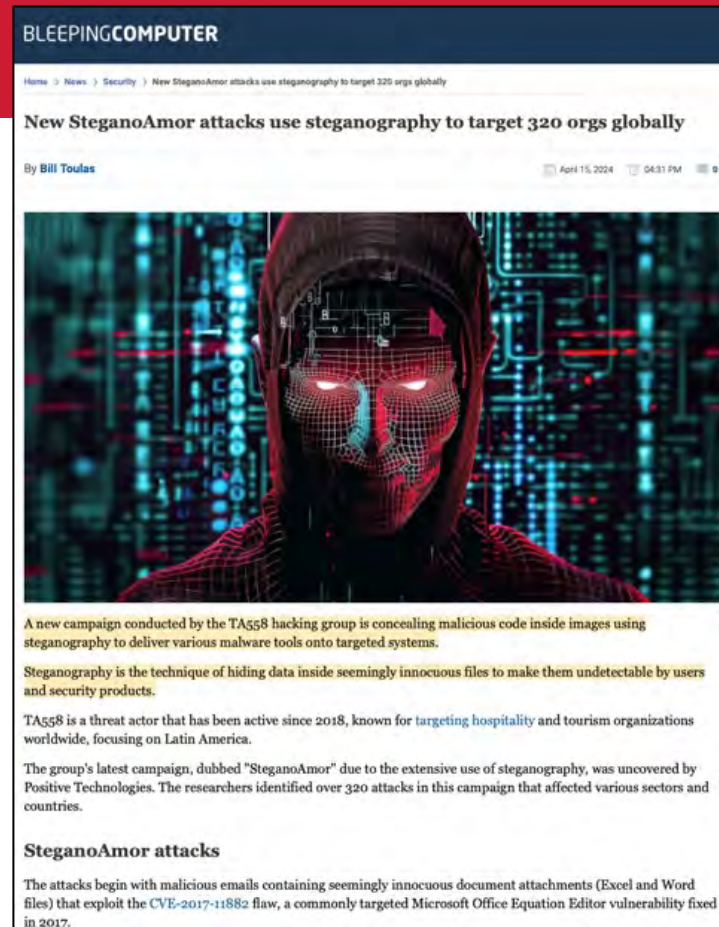
...

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.

UR<https://www.syracuse.com/crime/2022/04/code-hidden-in-photo-files-stolen-upstate-man-stole-ge-technology-to-try-to-help-china.html>

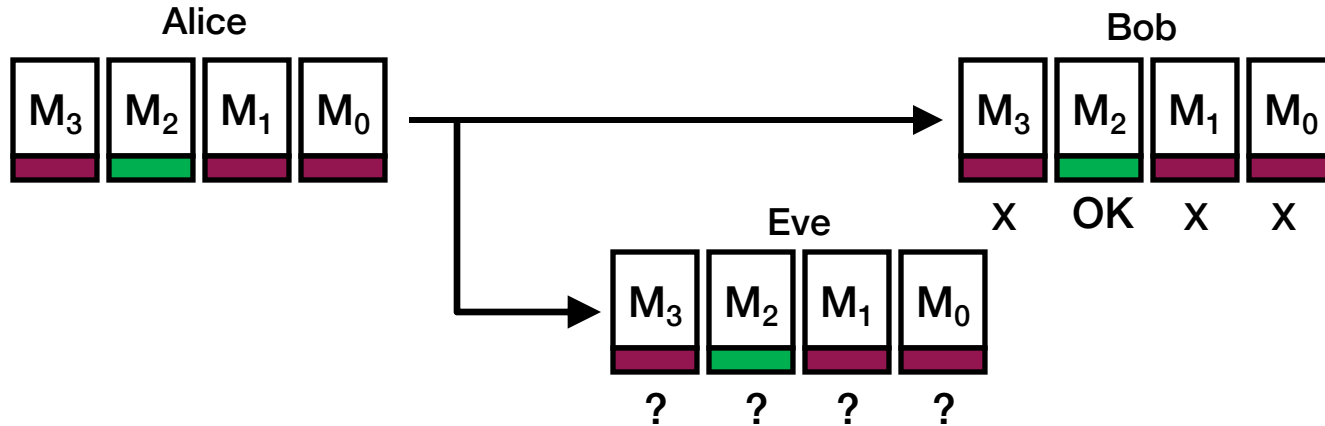
2024 attack campaign

- **Attacker sends emails with "harmless" content exploiting a vulnerability in the Microsoft Office Equation Editor**
 - Memory corruption vulnerability: stack buffer overflow overwrites the return address to execute the attacker's code
 - The victim has to open the file to activate this
- **When the attachment is opened, it sends a request to a URL to download an RTF document. When opened, it runs a VBS script that fetches payload embedded in images in URLs**



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 steganography** – embed messages in least-significant bits
- 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



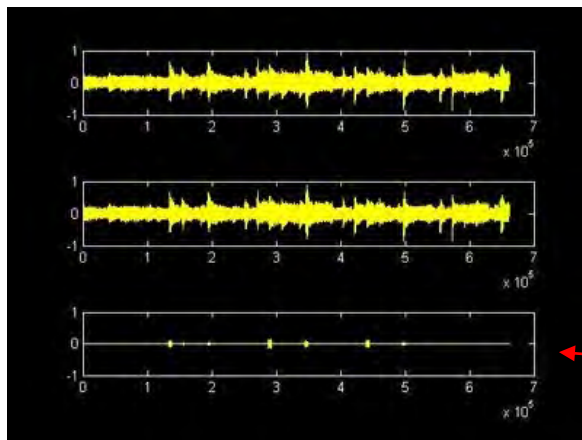
There are differences – but you don't notice them in the photo

The choice of *cover medium* (the content the casual user sees) is crucial:

Media with high noise (photos, music) works better because small changes are less noticable.

Perceptual coding

- Inject signal into areas that will not be detected by humans
- LSB steganography can also be used
- May be obliterated by compression






Amazon MP3 audio

Identifies where the song was
purchased, not the user

Difference





Amazon used inaudible signaling to prevent Echo devices from activating during its ads

THE VERGE TECH SCIENCE CULTURE CARS REVIEWS LONGFORM VIDEO MORE   

ENTERTAINMENT TECH AMAZON

Amazon has a clever trick to make sure your Echo doesn't activate during its Alexa Super Bowl ad


By Chaim Gartenberg | @cgartenberg | Feb 2, 2018, 3:31pm EST

ZDNET tomorrow belongs to those who embrace it today    

Home / Tech / Security

WAV audio files are now being used to hide malicious code

Steganography malware trend moving from PNG and JPG to WAV files.



Written by **Catalin Cimpanu**, Contributor
Oct. 16, 2019 at 9:00 a.m. PT

- **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!

Text

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



more
more

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

Text-based steganography

“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

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

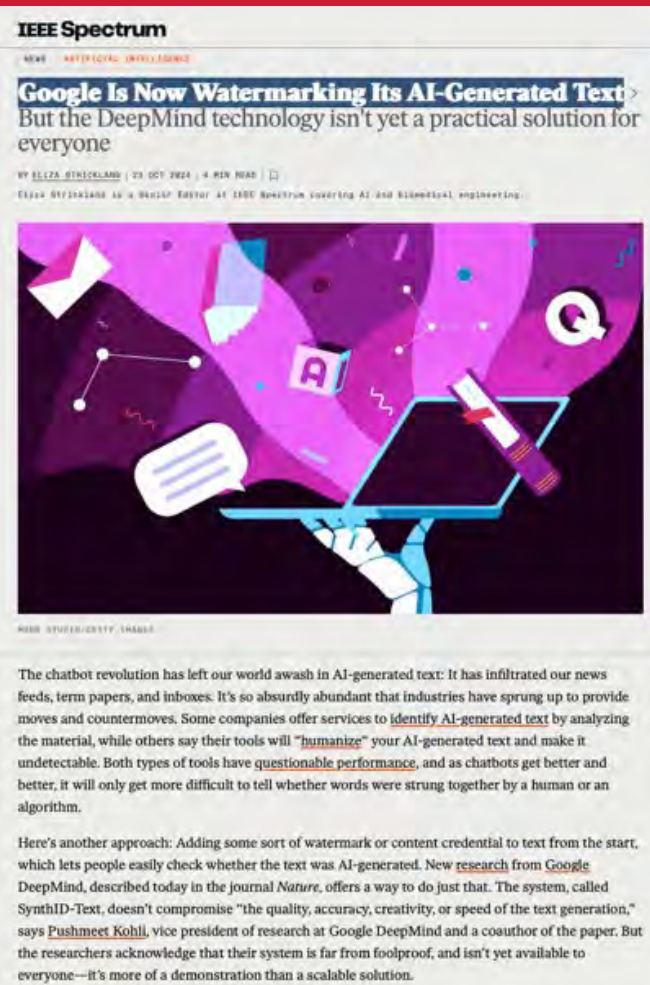
Watermarking AI-Generated Text

Google SynthID

Goal: Identify LLM-generated text, realizing that text can be easily altered to remove any watermarks

- Alters some words that a chatbot outputs to “introduce a statistical signature into the generated text”
 - SynthID-Text randomly assigns scores to candidate words that may be generated by the LLM and has the LLM output words with higher scores
 - A detector can then calculate the over
- Acknowledged to be “far from foolproof”
 - Users can make significant edits or ask another chatbot to summarize the text

<https://spectrum.ieee.org/watermark>



C2PA Standard (c2pa.org)

- **Coalition for Content Provenance and Authenticity**
 - Alliance between Adobe, Arm, Intel, Microsoft and Truepic
 - Standards for certifying the source and history of media content
- **Requires a C2PA-enabled capture device**
 - Content is hashed & signed to create a tamper-evident Content Credentials record
 - Can include attribution information
 - Any changes (crops, additions, removals, AI mods) are recorded
- **Can be easily removed – but goal is to have it to show provenance**



Leica M11-D
Leica M11-P



Canon EOS R1



Canon EOS
R5 Mark II



Fujifilm X-T50



Nikon Z6 III



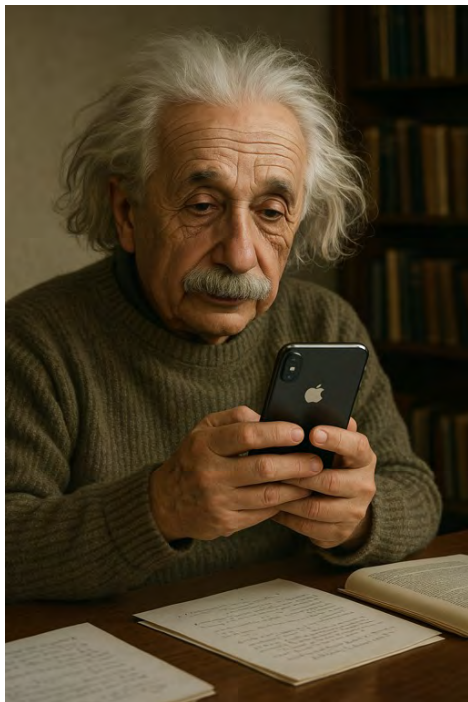
Fujifilm GFX100S II



Sony a1
Sony a7s III
Sony a7 IV
Sony a9 III



ChatGPT
Bing



content credentials

Select another file from your device or drag and drop anywhere

image.png
Issued by OpenAI

image.png
Issued by OpenAI

image.png
Issued by OpenAI

Process

The app or device used to produce this content recorded the following info:

App or device used
ChatGPT

Ingredients

image.png
Issued by OpenAI

About this Content Credential

Issued by
OpenAI

Change language

Fit

Compare

Statistical Steganalysis

Can we identify if content contains steganography?

Various attacks:

- **Histogram** (for LSB methods)
 - In natural images, histograms usually show smooth, continuous curves
 - Steganography (especially LSB embedding) can create unnatural patterns
- **Chi-square test**
 - LSB bits are normally not perfectly random
 - When secret data is embedded into LSBs (especially if it's random-looking encrypted data), it tends to randomize the LSB distribution
 - The chi-square test measures how far the observed distribution is from the expected one
- **Machine learning**: learn patterns of clean images to distinguish steganography

Payload capacity trade-offs

- The more data you hide, the greater the risk of introducing detectable artifacts

New Steganography Breakthrough Enables “Perfectly Secure” Digital Communications

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/>

The End