Digital Forensics 2: Discovery, Recovery, and Analysis

FORS 201

Chris Edwards

School of Computing

Semester One 2024

Section 1

Data Preservation

Physical Preservation

Storage media taken as evidence must be protected against a variety of physical threats:

- Unauthorised access
- Physical shock
- Heat and fire
- Extreme cold
- Moisture (humidity, condensation, flooding)
- Smoke, dust, harmful chemicals, mould spores
- Electromagnetic radiation
- Static electricity

Conditions for Physical Preservation

What is optimal depends on the medium:

- Paper self-ignites around 233°C
- Electronic components often have an 85°C or 105°C limit
- High humidity can promote mould growth on magnetic media
- ...and accelerate corrosion
- Extremely low humidity can also cause damage, increase static risk
- Magnetic media should be kept below 52°C and below 85% relative humidity

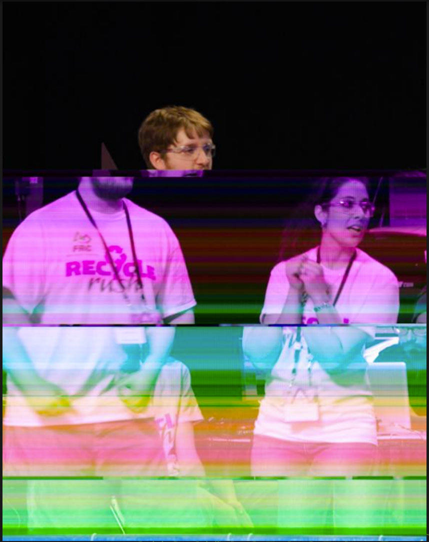
In general: stable, cool, dry conditions are best. May need air conditioning, filtration, etc.

Data Storage Cabinets

- Lockable protective cabinets for physical media
- Portable (for on-site storage and safe transport) or stationary
- Various sizes and degrees of protection
- Protection against intrusion, heat, water, etc.

"Bit Rot" (Data Degradation)

- Gradual decay or corruption of stored digital information
- Due to physical changes in the storage medium
- Backups and checksumming/parity/hashing can help detect and even correct/recover data
- Tools such as **`ddrescue`** can retry persistently to read/recover data
- Data scrubbing periodically rewrites data to "refresh" it



Redundancy Is Your Friend

- Digital data can be cloned cheaply and without loss
- RAID (Redundant Array of Independent Disks) safeguards against downtime
- Use parity files or resilient archive formats such as PAR2
- Take backups, use version control
- Lost Of Copies Keeps Stuff Safe

KEEP
BACKUPS
AND
CARRY
ON

Section 2

Discovery, Recovery and Analysis

Dealing with the Data

Searching and analysing an evidentiary image can be a Herculean task. Consider that the contents of a 4 TB drive, printed as text, would form a stack of paper around 100 km high!

```
A4page = 4 kbyte
4 terabyte / A4page * 0.1 mm -> km
100.0
```

(Frink syntax)

Need for Teamwork

The ability to clone and verify copies of digital data makes it practical to use teams to “divide and conquer” the task of analysis.

However, like imaging, cloning can be time-consuming, and trying to find relevant data by hand is generally impractical, even using a team of analysts.

Need for Automation

Therefore, automated tools must be used for searching for data that may be relevant to the case (names, e-mail addresses, bank account numbers, software or other data). There are many such tools available, conveniently available on bootable ("live") forensic disks such as CAINE.

Tools are available that can reconstruct timelines of activities for ease of analysis and reporting.

Information of Interest

Some specific types of information worth examining:

- Installed software (what, when, by whom?)
- Configuration files (how was the software being used?)
- Cache files
- System and application log files (e.g. system uptime, reboot timestamps, logins, errors)
- User documents, photos, video, etc.
- Downloaded files

Information of Interest (2)

- Browser history
- Browser cookies
- Contacts database
- Messages (e-mail, SMS, etc.)
- Call logs (cellular, VOIP)
- System software updates
- *Metadata* of all sorts...

Metadata

Metadata are data about data.

- Filesystem metadata such as creation/modification/access timestamps
- Generally per file, folder
- File comments, origin URLs, etc.
- Security metadata such as ownership, read/write permissions, shared access
- Office documents: author, institution, creation date, editing time, revision history
- Digital photographs (EXIF standard): camera make, model, serial number, timestamp, GPS location
- Tools: **`file`**, **`extract`**, **`exiftool`**, native application

Dennis Rader (the BTK Killer)

- American serial killer
- Killed ten people between 1974 and 1991 in Kansas but evaded arrest
- Sent letters to police and news media on his killings
- 2005: floppy disk sent to KSAS-TV, searched by police
- Metadata in a deleted Word document pointed to "Dennis" and "Christ Lutheran Church", at which Rader was council president
- This, combined with DNA and other evidence, lead to his arrest and prosecution

Anti-Forensic Software

Naturally, there are countermeasures against forensic analysis:

- `shred` command or system "secure delete" operation (multiple overwrites of file and metadata with random data)
- Filesystem defragmentation following deletion (not very reliable)
- Metasploit has some anti-forensic capabilities

Of course, these may have legitimate uses as well (personal/commercial privacy).

Forensic Analysis Software

There are many software tools and suites (both free and proprietary) for conducting the broader process of forensic analysis. These typically permit cases to be filed and managed within databases, associating case notes and metadata with the actual evidence, and can assist in producing reports for use in court.

Particular data of interest may include user files, system logs, browser history and cache, and browser cookies. Many computer forensic analysis tools are able to generate timelines of activity, making patterns of user behaviour clearer, and helping link to other evidence about a suspect’s activity, actions and whereabouts.

Section 3

Hidden Data

When you go looking for anything at all, your chances of finding it are very good. —Darryl Zero

Hidden Data

A challenge for any digital forensic investigation is the possibility of hidden data.

- Deliberate or incidental
- Highly dependent on technical expertise of criminal
- Deleted files may be recoverable
- Encryption and steganography present significant challenges
- Even some encrypted files may be recoverable
- Know where (and how) to look

Data Camouflage and “Ghosting”

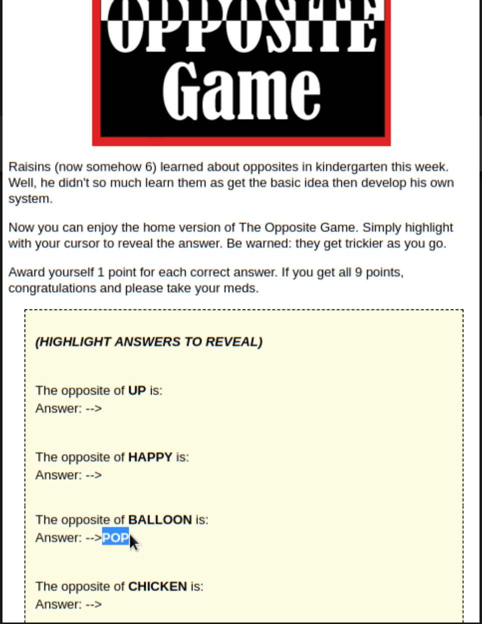
Some rudimentary concealment methods:

- Rename file with diversionary suffix/extension, e.g. `.DLL`
- Set text to same colour as background
- Shrink text or graphic items
- Layer sensitive items underneath others

Source: https://www.tapeplanet.com/

“Ghosting” Limitations

- Only superficially invisible
- Searching the underlying data will reveal all
- A recurring problem with digital documents with redactions...



OPPOSITE Game

Raisins (now somehow 6) learned about opposites in kindergarten this week. Well, he didn't so much learn them as get the basic idea then develop his own system.

Now you can enjoy the home version of The Opposite Game. Simply highlight with your cursor to reveal the answer. Be warned: they get trickier as you go.

Award yourself 1 point for each correct answer. If you get all 9 points, congratulations and please take your meds.

(HIGHLIGHT ANSWERS TO REVEAL)

The opposite of **UP** is:
Answer: -->

The opposite of **HAPPY** is:
Answer: -->

The opposite of **BALLOON** is:
Answer: -->POP

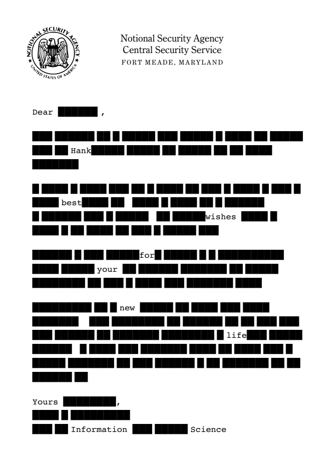
The opposite of **CHICKEN** is:
Answer: -->

Adapted from http://www.thesneeze.com/2009/the-opposite-game.php

Improper Redaction in Digital Documents

Naïvely overlaying black rectangles on sensitive text

- All the data are still there
- Hidden text easily recoverable from digital file
- Might be acceptable if only published in printed form
- There are many, many, many, many, many, many, many, many examples of this!



Notional Security Agency
Central Security Service
FORT MEADE, MARYLAND

Dear [redacted],

[redacted] Hank [redacted]

[redacted] best [redacted] wishes [redacted]

[redacted] for [redacted] your [redacted]

[redacted] new [redacted] life [redacted]

Yours [redacted],
[redacted]
[redacted] Information [redacted] Science

Redacting Properly

- Ensure that you are removing the sensitive content, not just overlaying it
- Use proper redaction tools (e.g. Adobe Acrobat Pro)
- If using a word processor, replace sensitive text with "[REDACTED]" (gives no clue as to length of redaction—esp. important if document uses fixed-width type)
- Don't neglect metadata
- Verify that the output file does not contain the sensitive data
- See also `https://lawyerist.com/how-to-redact-a-pdf/`
- See also `https://fas.org/sgp/othergov/dod/nsa-redact.pdf`

More Hiding Places

Text data in particular can be concealed in many unexpected locations:

- File comments (filesystem metadata)
- Document properties
- Code comments (e.g. MS VBA)
- ...

File “Deletion”

- Computer filesystems generally perform “lazy” deletion
- i.e., unlink the file from the filesystem and mark the sectors as available
- Generally does not actually erase the contents
- True whether you use the Trash/Recycle Bin or not
- Data may remain intact on disk for a long time
- (Cloud storage is also generally versioned and replicated, permitting recovery)

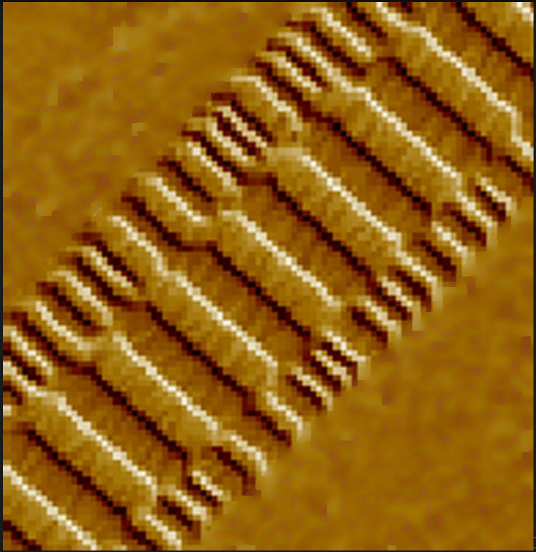
Recovering Deleted Files

- On some filesystem types, recovery may be trivial (**`undelete`** command)
- File "carving" software such as **`photorec`** can be used
- Scans media for recognisable signatures within known file formats
- Can also recover files hidden in slack space within/outside filesystems
- File fragmentation (and loss of fragments) can hamper recovery
- Defragmentation is sometimes used to prevent recovery, but can have limited effectiveness
- Secure erase must overwrite multiple times[1] before zeroing file content **and** metadata
- e.g. **`shred`** command, macOS securely empty Trash

[1] 7 times according to FIPS

Extreme Data Recovery: Magnetic Force Microscopy, etc.

- Advanced techniques exist to recover overwritten tracks
- Can potentially recover data going back several generations
- Magnetic Force Microscopy (e.g. NanoScope) is one example
- Extremely expensive and specialised
- Likely only used where national security at stake
- Defeatable by secure erase, physical destruction of disk, strong cryptography



Source: Digital Instruments, Veeco Metrology Group

Technically Advanced Hiding Places

- Text segments of binary files (**DLL**, **EXE**, etc.)
- NTFS (Windows filesystem) alternative data streams
- macOS resource forks/bundles
- Unallocated space on disk (use file carving to discover)
- Hidden partitions (use **testdisk** to discover)
- Steganography (diffusion of secret data within other files, e.g. images/sound)
- Non-user sectors on hard drives (e.g. spare sector regions, host protected area, device configuration overlay)

Physical Concealment Devices

Consider microSD cards:

- Tiny form factor (15 mm × 11 mm × 1 mm)
- High capacity (gigabytes to terabytes)
- Robust (no moving parts)
- Easy to conceal physically (e.g. classic spy hollow coin trick)

Section 4

Cryptography

The Need for Confidentiality

"Arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say."
— Edward Snowden

Cryptography in Brief

- Cryptography is the study of secret messages
- Mainly concerned with ensuring confidentiality of recorded or communicated data
- Also has a role to play in assurance of integrity (authentication of message content, sender identity)
- History of cryptography has become entwined with the history of information technology and computing (and politics, commerce, etc.!)
- Can present significant obstructions to a forensic investigation

Caesar Cipher

- Scheme used by Julius Caesar for military communications
- Simple *substitution cipher* (regular replacement of characters)
- Based on rotating the alphabet (A→X, B→Y, C→Z, D→A, ...)
- Probably OK if your enemies are generally illiterate!
- Vulnerable to *cryptanalysis*

Gaius Julius Caesar
100 BC – 44 BC

By Gautier Poupeau from Paris, France - César, CC BY 2.0,
`https://commons.wikimedia.org/w/index.php?curid=35036729`

Modern Digital Cryptography

Two major classifications:

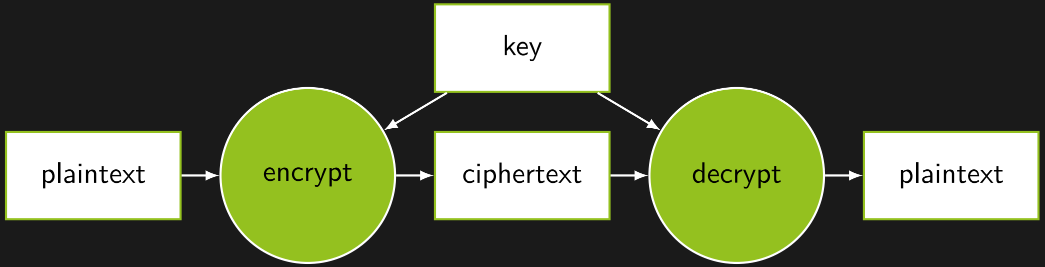
Symmetric or Shared-Key Cryptography uses the same key (or an easily derivable key) for both decryption and encryption.

Asymmetric or Public-Key Cryptography uses different (but mathematically related) keys for encryption and decryption.

Symmetric Cryptography

- Same key (or easily derivable key) used for decryption and encryption
- Also known as shared-key cryptography (key is a *shared secret*)
- Key-sharing can be problematic (if you have a secure channel, why not just use that?)
- Suitable for encrypting your own data for your eyes only
- Suitable for communicating with a group with equal read privilege
- Also effective in encrypting secret keys in asymmetric schemes
- Generally does not provide authentication of the message or sender!

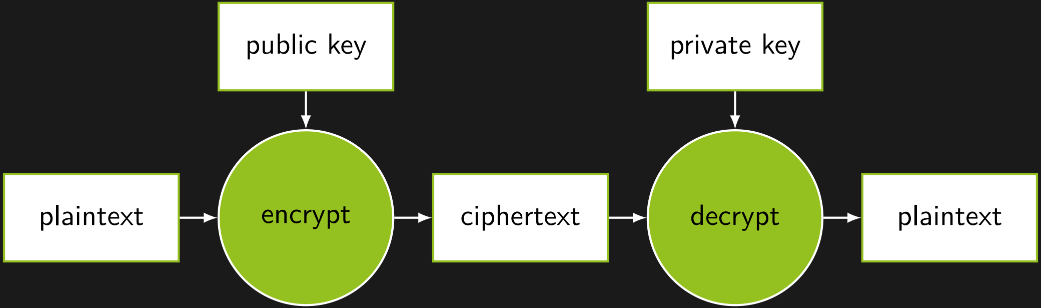
Symmetric Cryptography



Asymmetric Cryptography (aka Public-Key Cryptography

- Different key used for encryption and decryption
- Mathematically-related key-pair: private key and public key
- Private key cannot be deduced from public key
- Public key is used to encrypt
- Private key is used to decrypt
- Ideal for confidential one-way communication
- Need two-way communication? Just use two one-way channels!
- Frequently permits digital signatures for authentication of message and sender (no tampering or masquerading, and *non-repudiation*)

Aymmetric Cryptography



Cryptographic Vulnerabilities

- Weak cryptographic algorithms
- Weak keys or passwords (can be guessed or brute-forced)
- Poor key/password management
- Re-use of passwords
- Plaintext interception
- Social engineering

Cryptographic Terminology

plaintext Readable, unencrypted information comprising a single message.

ciphertext A unit of unreadable, encrypted information.

cipher A method (algorithm) for translating information into a form that is unreadable without the key.

encryption The process of scrambling a message according to a cipher.

decryption The reverse process of encryption: applying the rules of a cipher to a ciphertext in order to recover the original plaintext.

key Some (usually secret) information that is used with a cipher scheme to encrypt or decrypt information.

cryptanalysis Systematic ways of deducing what cipher and/or keys might have been used in producing some ciphertext.

Section 5

Steganography (“concealed writing”)

Steganography: Motivation

- Cryptography alone provides confidentiality, but not secrecy of communication
- Overt use of cryptography may draw attention, esp. if not legal (as in some parts of the world)
- Need a way to disguise sensitive data without drawing attention

Steganography: General Concept

- Covert embedding of information within a carrier/container/cover file
- Cover file is often an image or sound file of suitable file format
- *Diffusion* of information: dilutes, spreads, and combines the payload with the original data
- Encoded cover file can then be transmitted over potentially insecure channels, e.g. Facebook or e-mail
- Changes due to embedding are generally imperceptible
- (of course, what’s perceptible is not the same as what’s measurable)

Detecting Steganography

- Generally, steganography should evade detection.
- Even if a file is suspect, there is little to indicate the use of steganography
- Naïve algorithms may leave statistical traces, noise, unevenness
- Biggest vulnerability is probably the use of a known cover file (differences can indicate concealed data)

Original Cover Image

Text to be Diffused

"I have grown to love secrecy. It seems to be the one thing that can make modern life mysterious or marvelous to us. The commonest thing is delightful if only one hides it."

—Oscar Wilde, The Picture of Dorian Gray

"There is not a crime, there is not a dodge, there is not a trick, there is not a swindle, there is not a vice which does not live by secrecy."

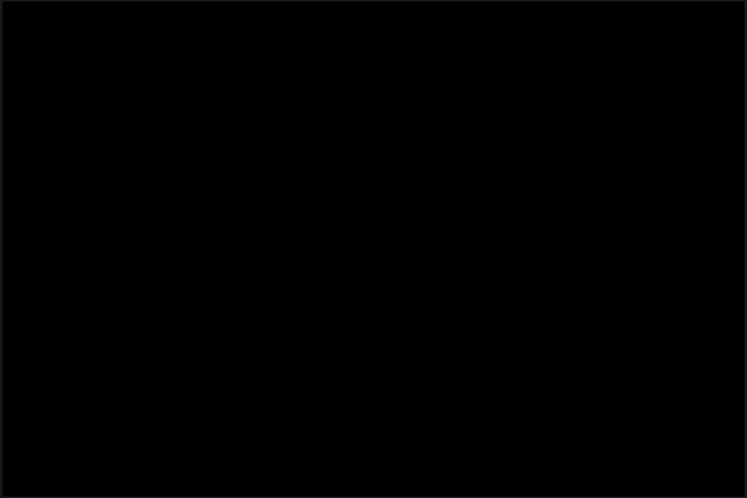
—Joseph Pulitzer

Steghide Command Line

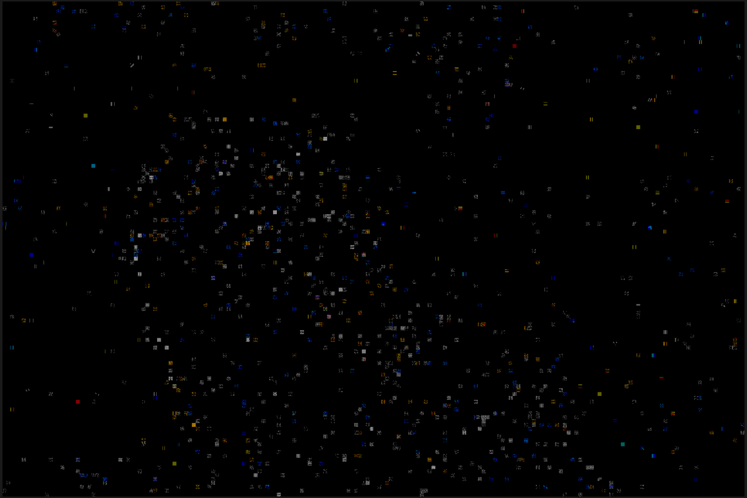
```
steghide embed -cf steg-cover.jpg -ef secret.txt
```

Cover Image with Concealed Content

Digital Image Subtraction Result



Digital Image Subtraction Result (increased contrast)



Analysis

Note how the differences are concentrated at areas of higher detail, which makes them harder to see/detect in the encoded cover image.

This kind of analysis is only feasible if the unmodified cover image is available. Careful steganography use will always use unique cover files.

Statistical analysis can also be used, but again because of the nonuniform diffusion, this can be challenging.

Combining Steganography with Cryptography

- Can be very effective, as ciphertext should resemble random noise anyway
- Even if stego is discovered, the message content remains protected
- Data compression also typically used, as compressed data more closely resembles random noise, and the smaller size means less change required to cover file

Section 6

Wrap-Up

Hints for Success

- Know your tools and choose them wisely
- Take your time—rushing causes mistakes
- Document everything thoroughly
- Be methodical—one step at a time
- Observe recommended practices
- Stay educated—tools and techniques are always changing
- The "bad guys" will also be trying to stay one step ahead!

The End

Thank You!