SWISS FTS
Forensic Technology Solutions

Advanced deduplication, beyond digital fingerprints

Irène Wilson

21 February 2019

Speaker

irene.wilson@swiss-fts.com

SWISS FTS
Forensic Technology Solutions

- Irène Wilson
  - Bsc in Forensic Sciences
  - Msc in Laws, Criminality and Security of New Technologies
  - EnCe
  - RCA, RSCP
  - Nuix eDiscovery Specialist
  - More than 9 years experience

Swiss FTS

SWISS FTS
Forensic Technology Solutions

www.swiss-fts.com

- Independent consulting firm for:
  - eDiscovery/Litigation Support
  - IT Forensics and Cyber Security/Investigations
  - Information Governance
- Offices in Singapore and Switzerland
- Operates Relativity datacenters in an ISO 27001:2013 certified environment in Singapore and Switzerland
- Uses standard and proprietary software solutions
- Staff certified on market leading software and hardware solutions
- Recognized by Who'sWho Legal and GIR

Agenda

SWISS FTS
Forensic Technology Solutions

- Introduction
- Benefits of deduplication
- MD5 hash: definition and limitations
- Problematic situations and solutions:
  - Email headers
  - Different file formats
  - Inclusive content within a same thread
  - Email format and variations
  - Email archives
- Risks
- Conclusion

Introduction

“Digital fingerprint”

SWISS FTS
Forensic Technology Solutions

Benefits of deduplication

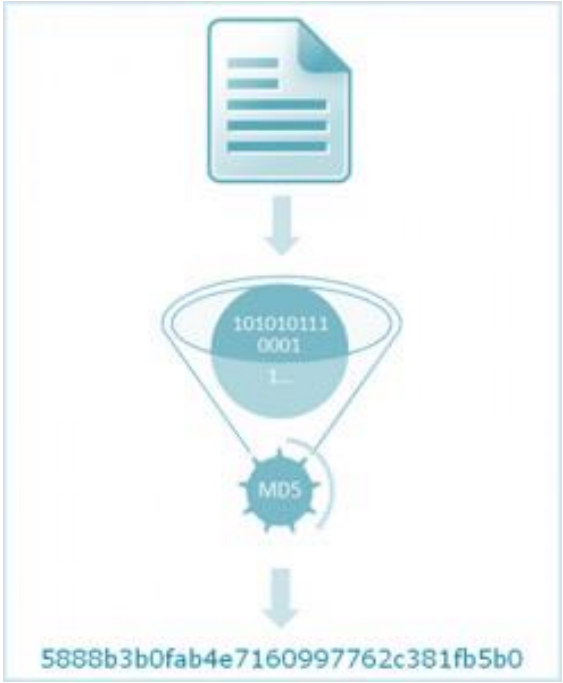
SWISS FTS
Forensic Technology Solutions

- Meet deadlines while saving money
- Increase consistency of review decisions and redactions on identical and similar documents

MD5 hash: definition and limitations

SWISS FTS
Forensic Technology Solutions

- What?
  - 32-characters value based on binary content
- Why used in forensics?
  - Reproducible
  - One-way algorithm
  - Low chances of collision
  - Short identifier
- Limitations?
  - Format-sensitive
  - Based on the content only

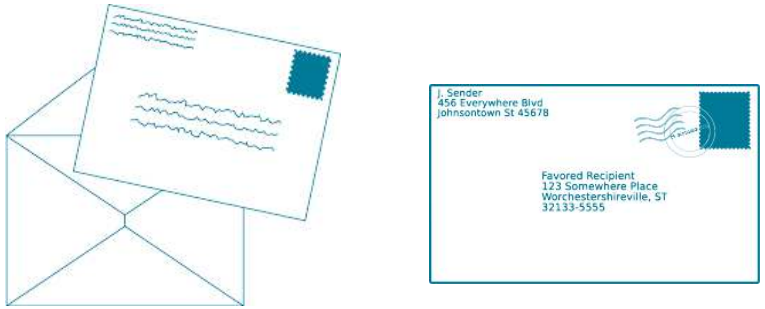


Problematic situations and solutions

Email headers

SWISS FTS
Forensic Technology Solutions

- Problem:
  - The route an email takes impacts its header, therefore its binary content.



- Solution:
  - Nuix custom MD5
    - Subject
    - From
    - To
    - Cc
    - Email body (text tokenized so whitespace and irrelevant characters are removed)
    - Binary streams of all attachments

Problematic situations and solutions

Different file formats

SWISS FTS
Forensic Technology Solutions

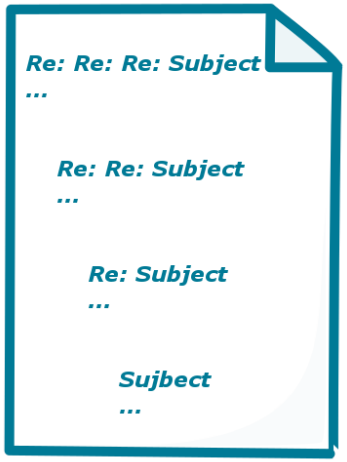
- Problem:
  - Same content into different file formats

- Solutions:
  - Near duplicates (shingles, textual duplicates)
  - Fuzzy hashing (SSDeep)
  - PhotoDNA (law enforcement forces only)

Problematic situations and solutions

Inclusive content within a same thread

- Problem:
  - Redundant content because of conversation history



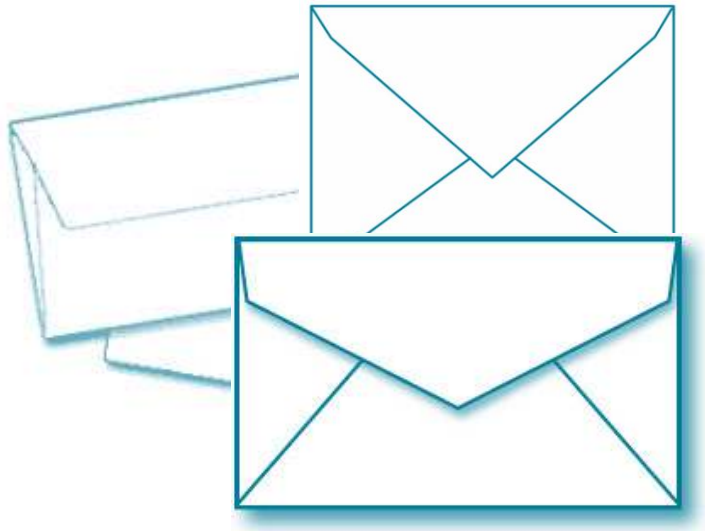
- Solution:
  - Email threading
    - What was not sent to threading (loose files for ex.)
    - What was ignored by the process
    - Unclustered items
    - Endpoint items
    - Endpoint-attach items

Problematic situations and solutions

Email format and variations

SWISS FTS
Forensic Technology Solutions

- Problems:
  - Internal vs. external email addresses
  - Attachment order
  - Etc.



- Solutions:
  - Custom deduplication using a metadata profile
  - Custom script

Problematic situations and solutions

Email archives

SWISS FTS
Forensic Technology Solutions

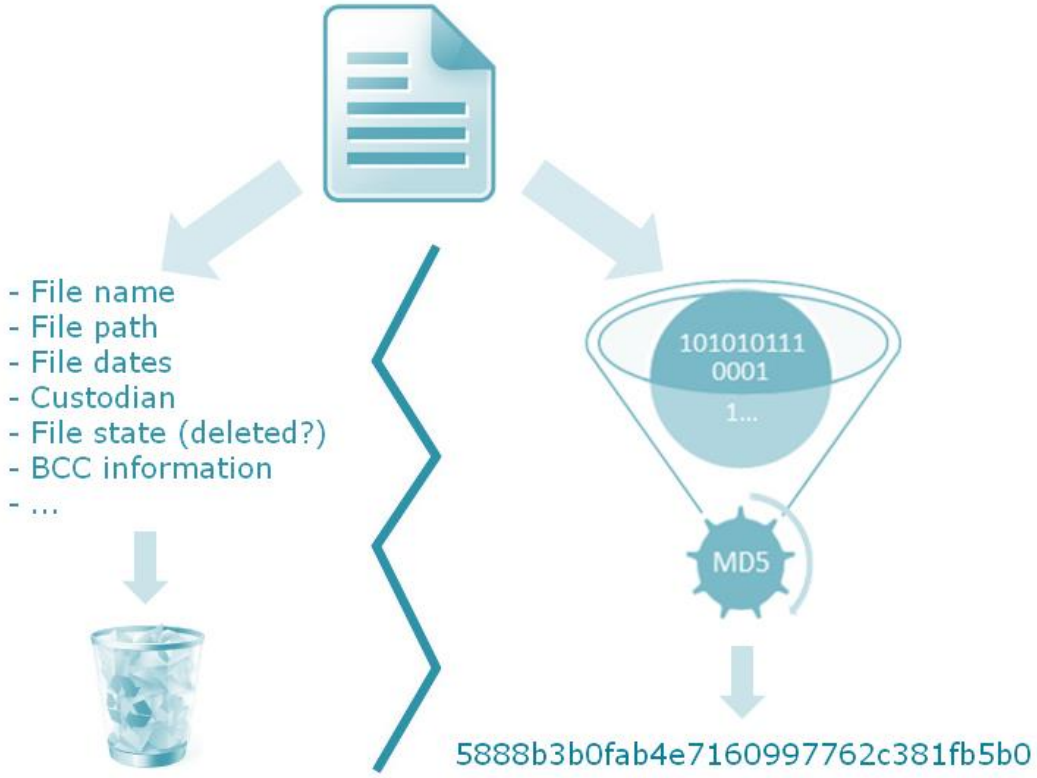
- Problem:
  - Original emails vs. archived emails

- Solution:
  - Custom script:
    - Search for “This message has been archived”.
    - Identify potential original emails with same Name, From, Item Date
    - Export text and compare (original text starts with archived text, without archiving message)

Risks

Ignored metadata

- Relevance of external or ignored metadata
  - Data leakage investigation
  - Date filtering
  - Custodian filtering
  - Evidentiary value based on location



Risks

False duplicates

- Empty files
- Calendar entries

Conclusion

SWISS FTS
Forensic Technology Solutions

- Deduplication and data reduction are a real need in eDiscovery.
- Advanced deduplication requires performant tools and creative approaches, while ensuring reproducibility.
- Deduplication risks need to be kept in mind all along the project.

SWISS FTS
Forensic Technology Solutions

CAN WE ASSIST YOU OR DO YOU HAVE ANY QUESTIONS?

If so, simply call or email us.

SWISS FTS AG
Europa-Strasse 19 | 8152 Glattbrugg | Switzerland
Phone +41 43 266 78 50 | info@swiss-fts.com | www.swiss-fts.com

SWISS FTS AG
Av. De Sévelin 46| 1004 Lausanne | Switzerland
Phone +41 21 510 53 81 | lausanne@swiss-fts.com

SWISS FTS (SINGAPORE) PTE LTD
55 Market Street | #10-01 | Singapore 048941
Phone +65 6950 1370 | singapore@swiss-fts.com