Towards facilitating reliable recovery of JPEG pictures?

P. De Smet

(edited for public release)

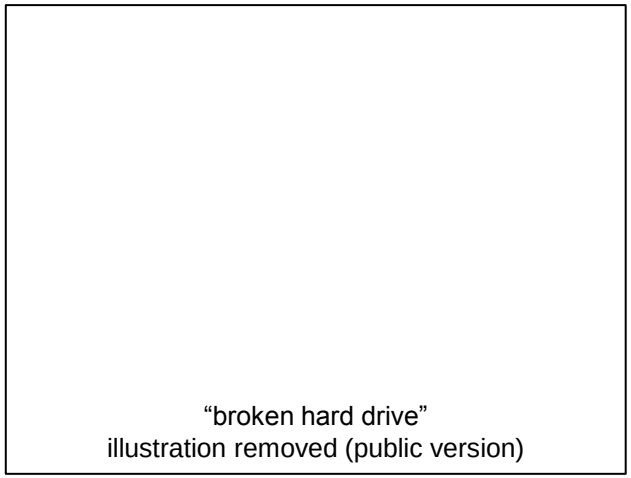
patrick.desmet@just.fgov.be

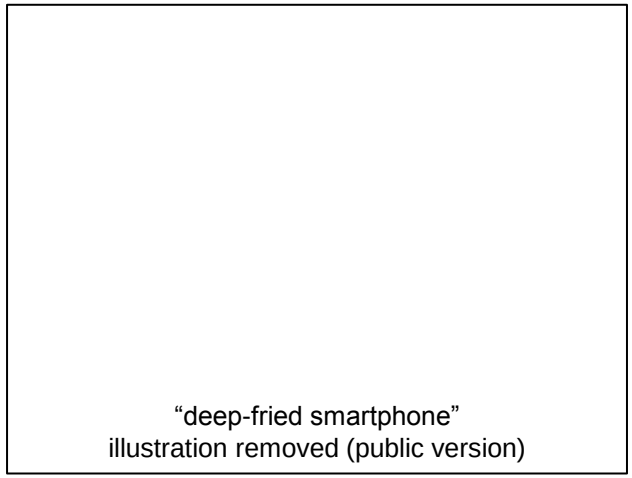
http://nicc.fgov.be/datarecovery/

Federale Overheidsdienst **Justitie**
Service public fédéral **Justice**

Introduction & disclaimer

- Aim of this talk: discuss simple possibilities which may have a major impact on data recovery
- This talk is not only about law enforcement but it's also about "consumers", society: digital heritage/preservation, you:
- What to do when you (accidentally) damage/erase your HDD/SD/USB-stick/… ?
- We have limited knowledge about JPEG (beyond ISO/IEC 10918); so maybe solutions are trivial?



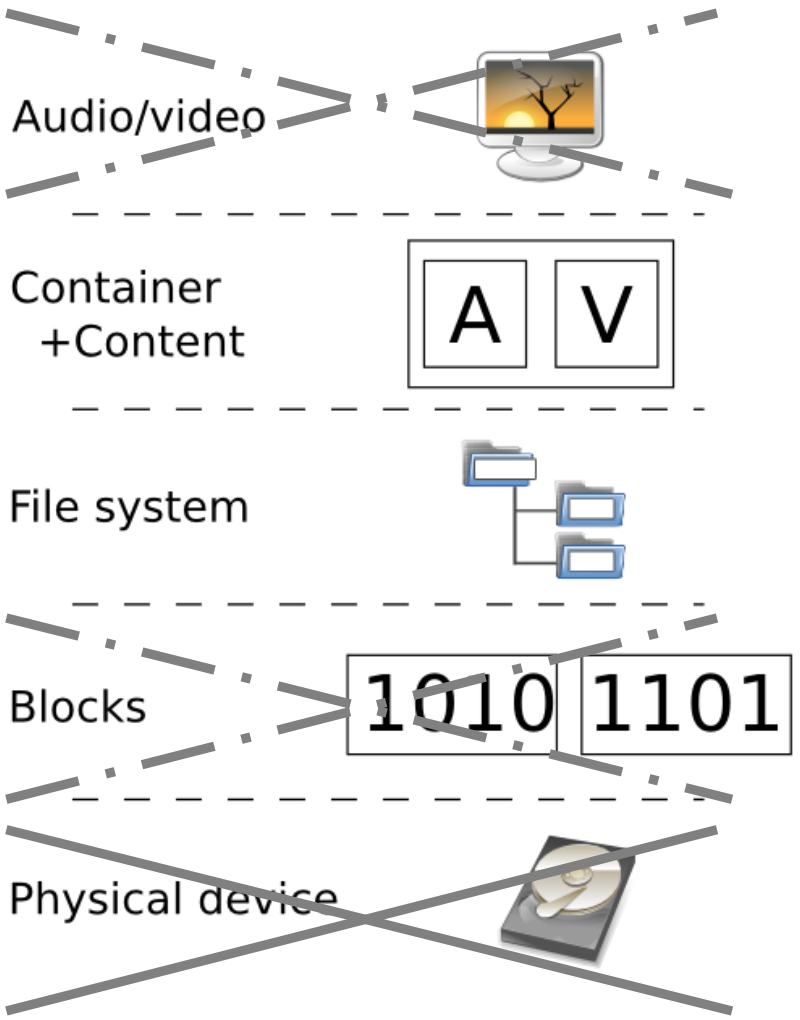


Introduction

Forensic data recovery:
recover and/or verify information from damaged devices, file systems, or data files.

In this presentation focus only on:
software based data recovery using a proper forensic copy or "data dump" D, obtained for a given electronic device

We focus on non-allocated data; i.e., if you have an active file system the first step is reading those files

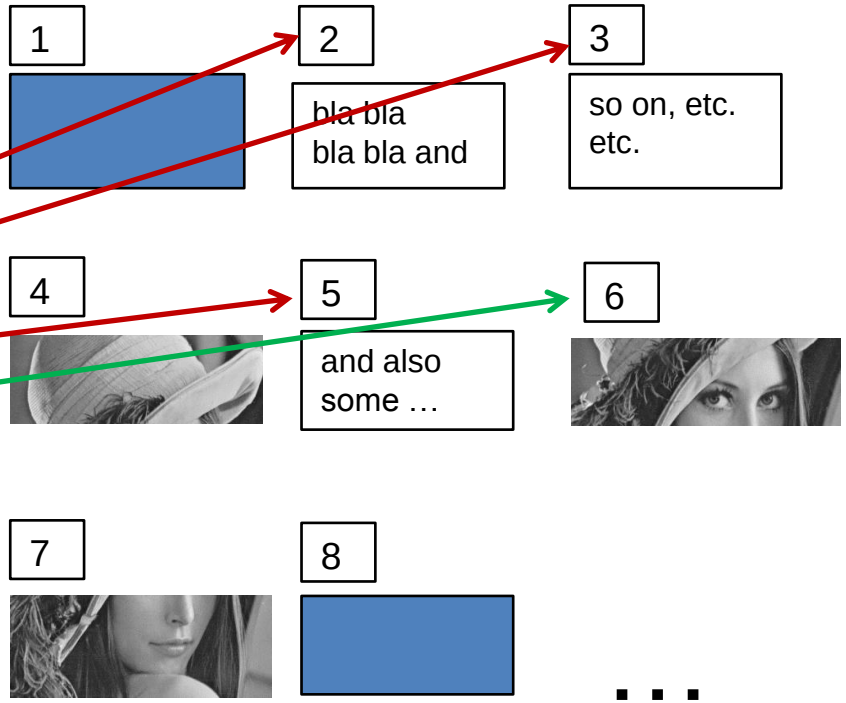


Introduction: file fragmentation

File systems typically try to recycle storage space, so files may be split into multiple fragments that can written anywhere when some free space has been located

File system meta-data/"table of contents"

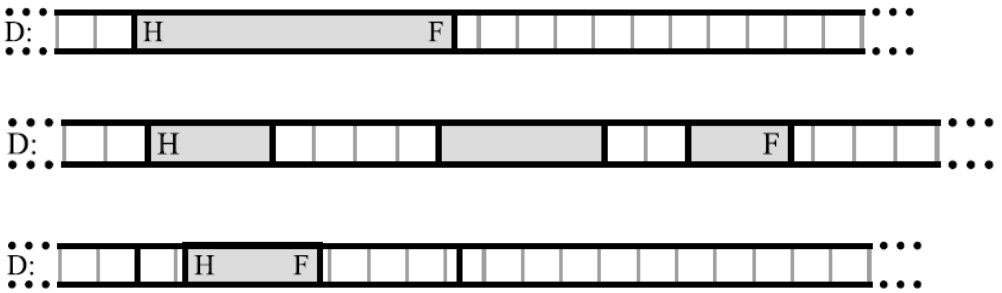
File	Address in storage space
myText.txt	2, 3, 5
Lena.jpg	4, 6, 7
…	…



Both files are fragmented, i.e., they are not stored as sequential/linear blobs

Introduction -Types of Recovery

- Complete file carving:
  - single fragment (SF)
  - multi-fragment (MF)



- Embedded files:
  SF, MF, partially re-encoded (e.g., JPEG in zip)
- Incomplete files: SF and MF
- "Trace evidence" : triage and examination of remnant data: JPEG or not JPEG?

Importance of JPEG

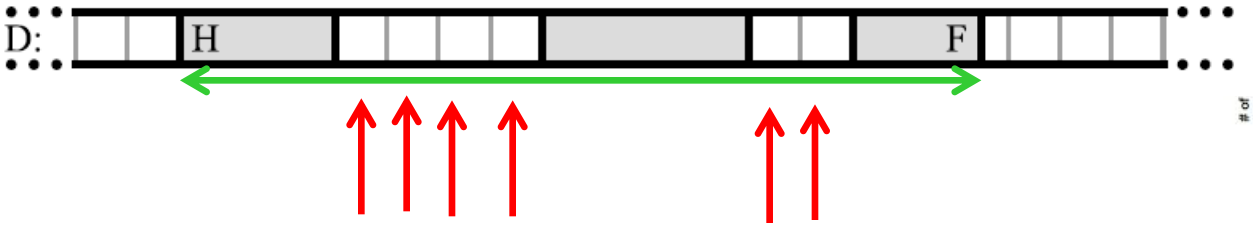
- "Big data" requires triage
- JPEG is everywhere
- (Recovery of) a picture is worth 1000 …
  … man hours of work

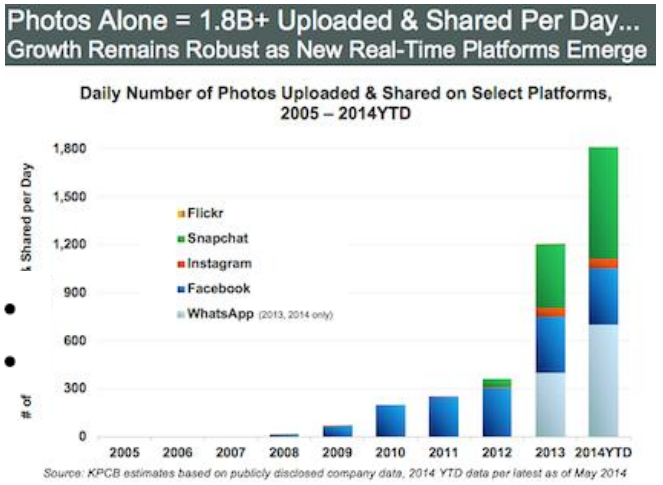
  (what if someone took a snapshot of a criminal stabbing one of your family members?)
- Getting one format right will help recovery of all others

  E.g. multi-fragment files may require exhaustive search for missing fragments

  E.g. simple **header footer** carver may destroy (many) other files

"Big data" joke:
illustration removed (public version)





JPGcarve vs. APF

Data dump	Cluster size	Number of complete multi-fragment JPEGs			Time [s]	
		Ground truth	JPGc	APF	JPGc	APF
APFD	4096	10	10	9	12.76	67.02
	2048	10	10	9	13.54	98.50
	1024	10	10	3	15.17	132.12
	512	10	10	2	18.35	146.17
DFRWS2006	512	7	7	7	2.10	31.02
DFRWS2007	512	13	13	13	28.50	67.32
SKLprof	16384	4	4	3	4.51	74.64
	8192	4	4	3	4.68	75.38
	4096	4	4	3	5.27	77.20
	2048	4	4	3	6.19	79.17
	1024	4	4	3	8.20	80.89
	512	4	4	1	13.48	98.63
SKLval	16384	5	5	4	5.57	90.00
	8192	5	5	4	5.69	81.91
	4096	5	5	4	6.19	86.52
	2048	5	5	4	7.01	87.89
	[illegible]	5	5	4	9.62	102.93
	[illegible]	5	5	0	14.71	115.94

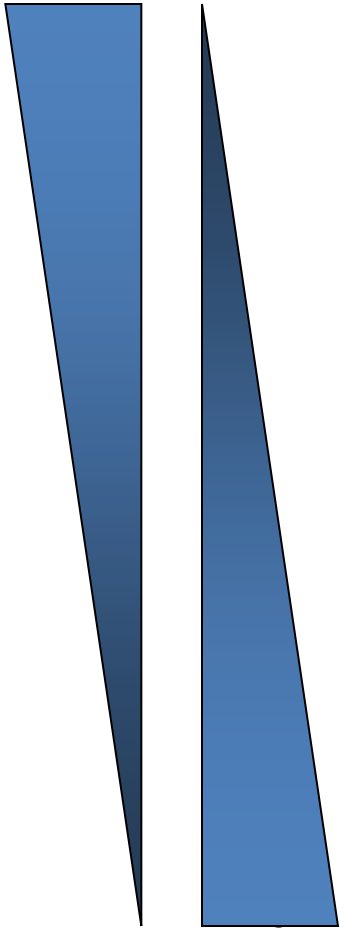
http://dx.doi.org/10.1109/TIFS.2015.2475238

JPGcarve

- Search for JPEG headers in data dump D
- Recover all single fragment files (header-footer carving and libjpeg(turbo) based file validation)
- For all remaining JPEG headers:
- Use search space S =D – {SF files}
- Eliminate clusters in S based on: JPEG markers, areas that have not been properly byte stuffed, entropy
- Find end positions of fragment, try matching end to new data, i.e., all remaining clusters (again using libjpeg(turbo) decoding: very slooooow)

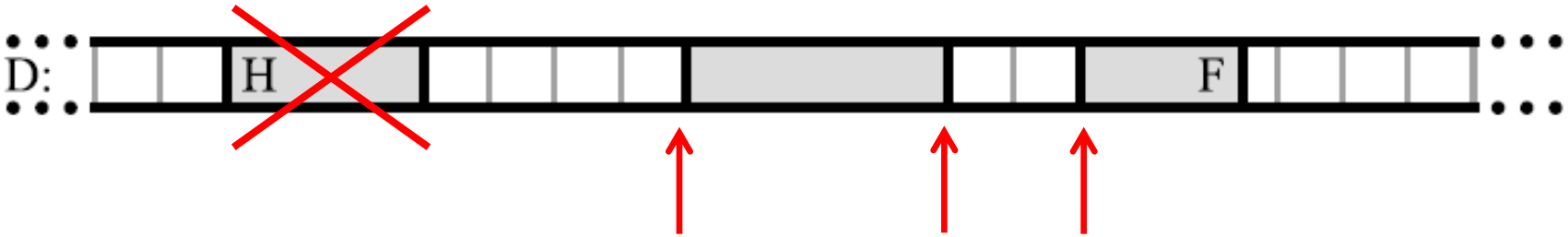
State-of-the-Art (?)

	Methods	Tools	Ref. test data	Tool compare	QA	Vol. of data	Complexity / Time
• Complete files:							
– single fragment (SF)	+	+	+	?	?		
– multi-fragment (MF)	+	+-	?	+-	?		
• Incomplete files:							
– SF	+-	-	-	--	-/?		
– MF	--	--	--	---	---		
• Embedded files:							
– SF	+	+	?	-	--		
– MF	+-	+-	?	-	--		
– re-encoding	--	--	?	--	--		
• "Trace evidence"	+-	--	--	--	---		



So …

- Please “ignore” any issues related to bit stream error detection/correction (please don’t, but:)
- Consider file system fragmentation: a file can be split at any point, into one or multiple fragments:
  - “Do we have reliable statistics? “ Yes, some in literature (HDDs), but what about SSDs, any wear-aware device?!
  - statistics are irrelevant (for HDDs, traditional sector API devices): depends on usage: backup media (linear blobs) vs. smartphone (shoot, delete, shoot, download, …)
- And: these fragments can be (partially) overwritten, or corrupt (e.g., due to hardware repair issues)…



Conclusion

- Please consider investigating any technical possibilities enabling (faster) JPEG file recovery, including:
  - IDing JPEG remnants: within data dump
  - IDing JPEG remnants: as being parts of the same or a similar file
  - regrouping and matching JPEG remnants to reconstruct files
  - robustness against loss of meta-data and/or any part of the data (certain fragments)
  - ….?

Random crazy/stupid solutions?

- Hurray for JPEG byte stuffing and markers!
- Recovery facilitating ID of any (sub)fragments (JPEG will be box based?); i.e., how to include/exclude (encrypted?) data in the recovery search space?
- Bulk of pixel data: additional integrity checks could be used to verify if chunks of data clusters belong together? (fragmentation/error detection without decoding)
- A “ToC” could be used to group chunks/fragments together, at correct offsets?
- **Main point: “…?”, if interested: let us discuss this**

References

- J. De Bock, P. De Smet, JPGcarve: an Advanced Tool for Automated Recovery of Fragmented JPEG Files, IEEE TIFS, to appear Nov./Dec.2015, http://dx.doi.org/10.1109/TIFS.2015.2475238 http://nicc.fgov.be/datarecovery/
- E. Uzun, H. T. Sencar, Carving Orphaned JPEG File Fragments, IEEE TIFS, vol. 10, no. 8, 2015
- B. Kloet, Measuring and Improving the Quality of File Carving Methods, MSc Thesis, Eindhoven University, Oct. 2007
- L. Huang, M. Xu, H. Zhang, J. Xu, N. Zheng, A Method of Approximately Reconstructing JPEG Quantization Table via Saturated Overflow, IEEE ICCSIT2011, China.
- D. Salamani, Header construction for carved JPEG fragments, EAFS2015.
- A. Pal, N. Memon, The Evolution of File Carving, IEEE SPM, March2009.
- H. Sencar, M. Memon, Identification and recovery of JPEG files with missing fragment, Digital Investigation, 2009.
- etc. (full reference list available upon request)

Thank you!

Q&A

Please join us in further
collaboration regarding this and related topics:

patrick.desmet@just.fgov.be / nicc-din@just.fgov.be