Advanced file carving

How much evidence are you ignoring?

Bas Kloet, Hoffmann Investigations
September 2010

Who am I?

- Bas Kloet:
  - Digital Forensic Investigator at Hoffmann Investigations since 2007
  - Master project: File carving…
- Hoffmann Investigations:
  - Founded in 1962
  - Currently about 80 employees, 1000 cases per year
  - Fraud, theft, industrial espionage

About this presentation

Based on a full day training

Contents:

1. Carving and basic file information
2. File Systems and Fragmentation
3. General File Carving Techniques
4. Measuring File Carving Quality
5. Specific Purpose Carving Tools

Topic 1 - (File) Carving

Carving is a general term for extracting structured data (files) out of raw data, based on format specific characteristics present in the structured data.

Topic 2 - File Systems and Fragmentation

- Files are stored in file systems
  - Windows (FAT 12/16/32, NTFS)
  - Linux (Ext2/Ext3/Ext4, Reiser)
  - Mac (HFS, HFS+/HFSX)
- File systems store data in clusters or blocks
- Files are usually stored sequentially by the OS on media

FAT File Allocation

FAT File System Structures

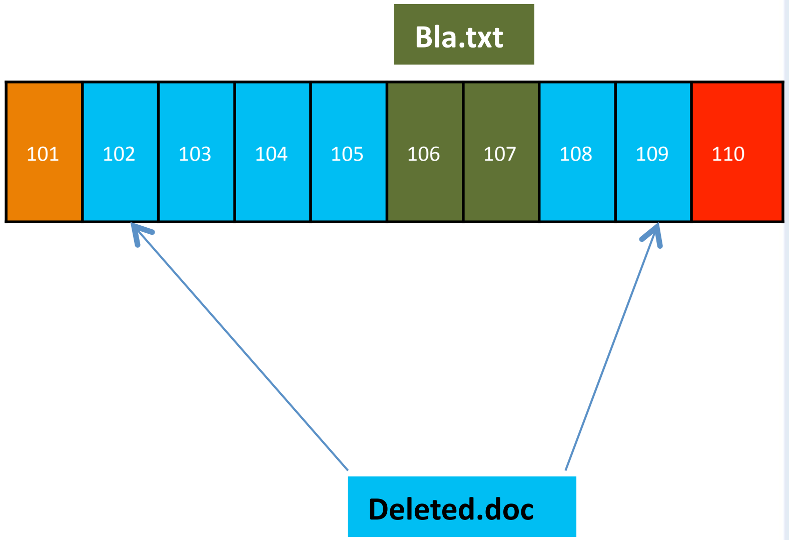
Root Directory Entries

File name	*Starting block*
Deleted.doc	102
Bla.txt	106
Archive.pst	110

FAT

Block Index	*Next Block*
101	Free
102	103
103	104
104	105
105	108
106	107
107	EOF
108	109
109	EOF
110	111

Media Data Block Area



Bla.txt

101	102	103	104	105	106	107	108	109	110

Deleted.doc

FAT File Allocation

FAT File System Structures

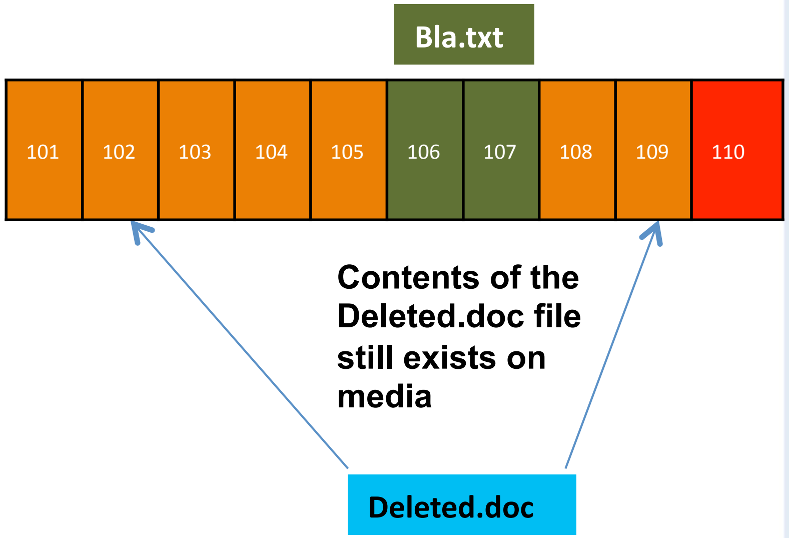
Root Directory Entries

File name	*Starting block*
_eleted.doc	102
Bla.txt	106
Archive.pst	110

FAT

Block Index	*Next Block*
101	Free
102	Free
103	Free
104	Free
105	Free
106	107
107	EOF
108	Free
109	Free
110	111

Media Data Block Area



Bla.txt

101	102	103	104	105	106	107	108	109	110

Contents of the Deleted.doc file still exists on media

Deleted.doc

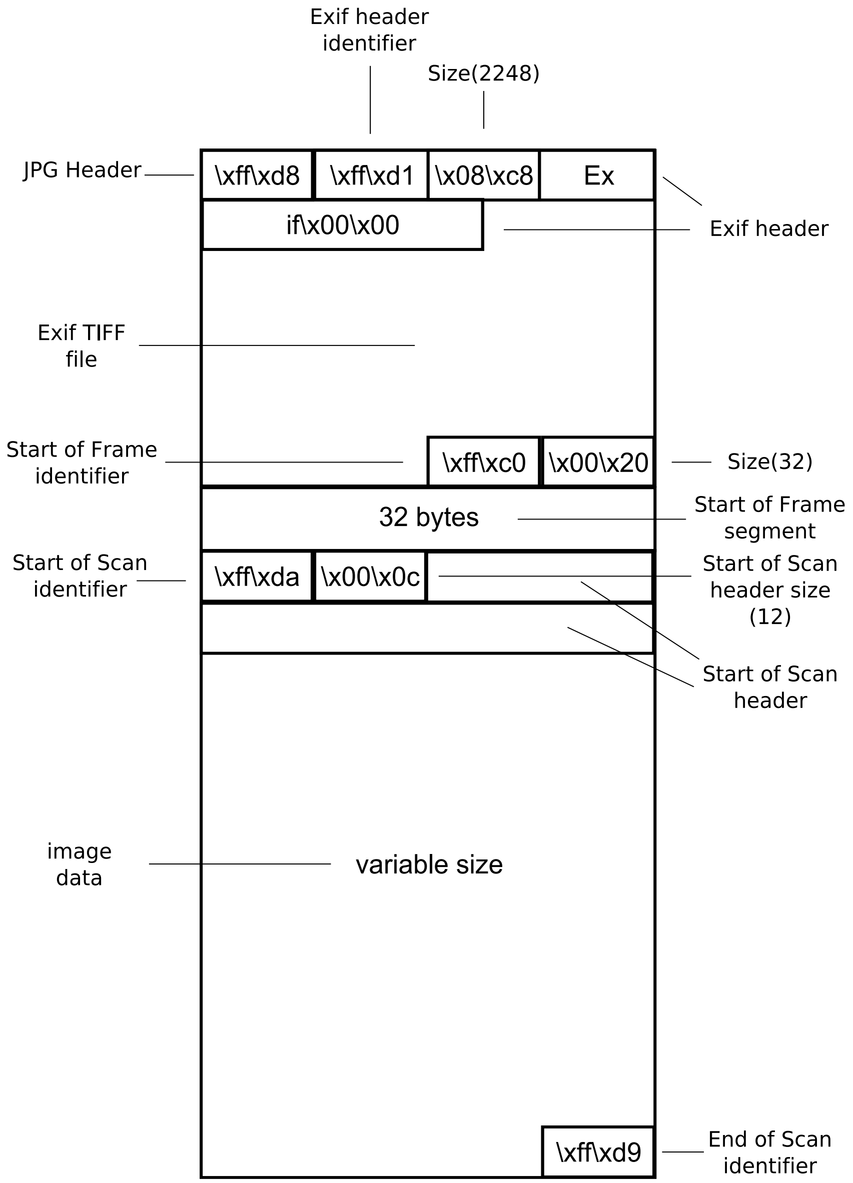
Topic 3 – General File Carving Techniques

The most common general file carving techniques are:

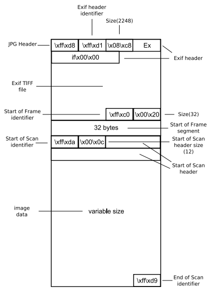
- Header-footer or header- “maximum file size” carving
- File structure based carving
- Content based carving

JPEG file structure

- JPEG header
- Exif header identifier
- Exif header
- Exif TIFF data
- Exif JPEG Thumbnail
- Start of image data (Start of scan)
- Image data
  End of image data
  (End of scan)

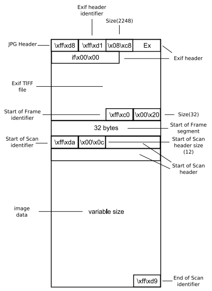


Header-footer Carving



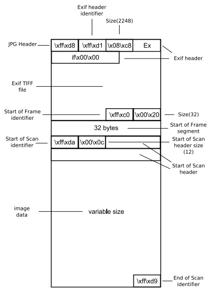
- Recover files based on known Header and Footers or maximum file size
  - JPEG: “\xFF\xD8” header and “\xFF\xD9” ‘footer’
  - GIF: “\x47\x49\x46\x38\x37\x61” header and “\x00\x3B” footer
  - PST: “!BDN” header and no footer
- If the file format has no footer a maximum file size is used in the carving program
- Known header footers carvers are Scalpel, Foremost and File finder (EnCase)

File Structure Based Carving



- This technique uses the internal layout of a file
- Elements are header, footer, identifier strings and size information
- Known carvers which use this technique are Foremost and PhotoRec

Content-based Carving



- Content structure
  - Loose structure (MBOX, HTML, XML)

- Content characteristics
  - Character count
  - Text/Language recognition
  - White and Black listing of data
  - Statistical attributes (Chi^2)
  - Information entropy

Carving problems

- Time consuming
- Many unreadable invalid and partial results
- More data out than in
- No offset/sector reference to input data
- Quality of the tooling is unclear

Course of action:

Measure quality of (file) carving

Topic 4 - Measuring quality

- Determine quality criteria
  - Required features
  - Quality of the results
- Determine the carving quality of a tool
  - Tools and datasets
  - Results

Quality of the results

In dataset Recovered	Yes	No
Yes	Positive	False positive
No	False negative	-

“Recall”: What proportion of the available files is recovered?

“Precision”: What proportion of the recovered files is correct?

How to check the results?

- Compare carving results to known correct files
- Chosen method: determine a similarity index by using ssdeep
- 99% match or better is a Positive

Tools and datasets

- Tools
    - Scalpel
    - Encase
    - FTK3
    - Foremost
    - PhotoRec
    - Revit
- Datasets
    - FAT carving test dataset (15 files)
    - DFRWS 2006 challenge image (32 files)

Tool quality – FAT

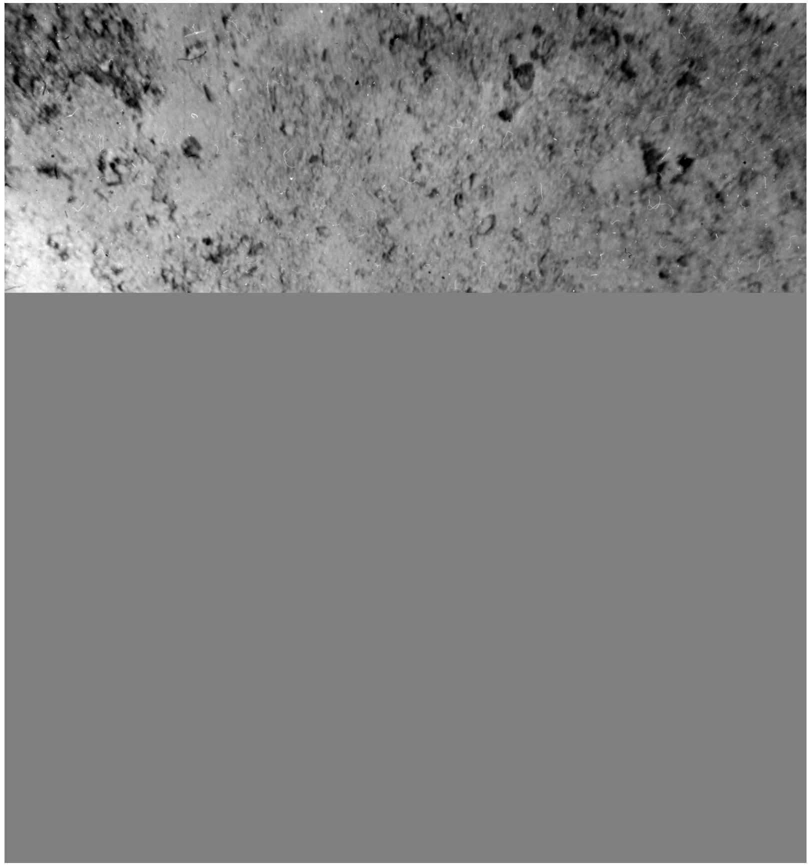
Tool	Carving Recall	Carving Precision
Scalpel	0.333	0.003
FTK 1.81	0.4	0.6
Encase 6.7	0.467	0.538
FTK 3.0	0.667	1.0
Foremost	0.8	0.857
Photorec	0.933	1.0
Revit	0.933	1.0

Tool quality – DFRWS 2006

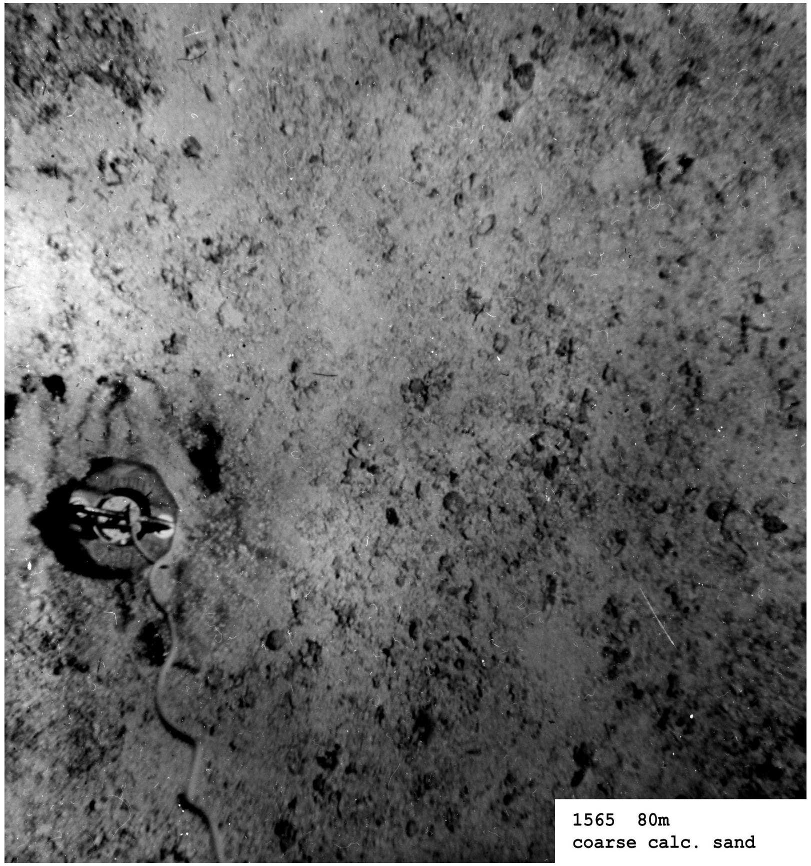
Tool	Carving Recall	Carving Precision
FTK 3.0	0	-
Scalpel	0.219	0.001
Encase 6.7	0.219	0.28
FTK 1.81	0.25	0.258
Foremost	0.281	0.36
Photorec	0.563	0.643
Revit	0.625	0.69

What does this mean in practice?

Encase on DFRWS 2006



Photorec on DFRWS 2006



Carving quality conclusion

- Huge difference in carver quality
- On “simple” datasets, tools like Photorec and Revit get very good results
- On more complex datasets the overall quality of the results is significantly lower
- Fragmentation of files can have a major impact on the quality of the results

Topic 5 - Specific purpose carvers

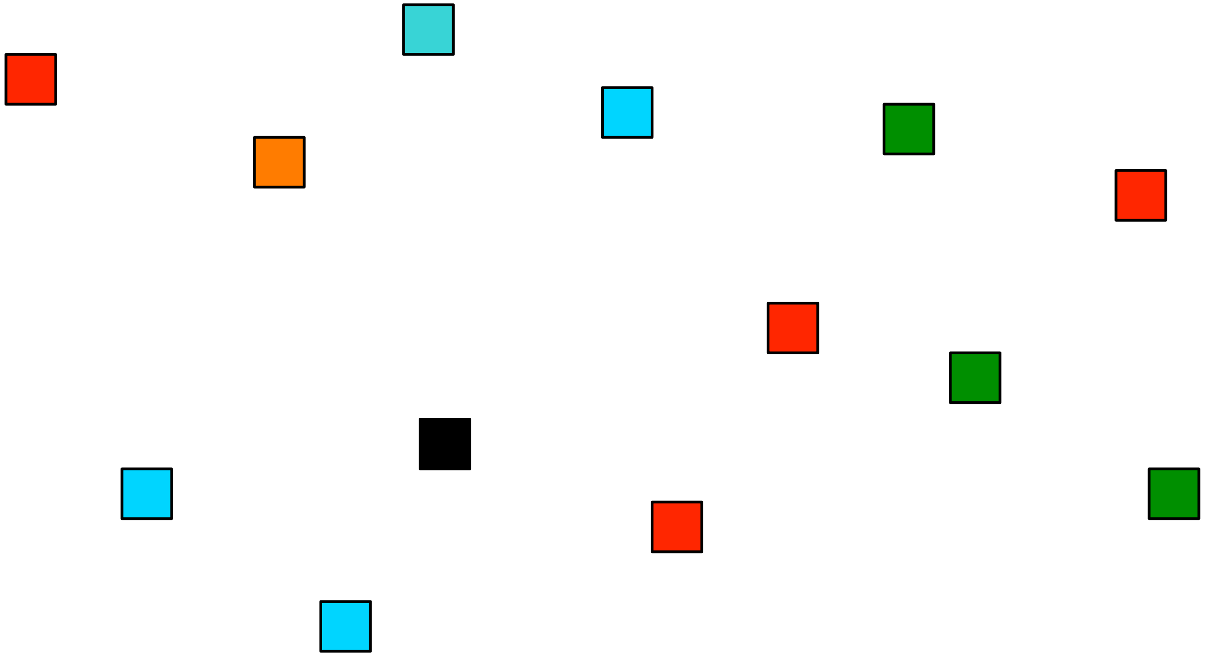
- Support one or more specific file formats
- Use techniques that are specifically based on characteristics of those file formats
- Usually more effective for those file formats than generic purpose carvers
- Some tools are created to carve *inside* of specific files

Specific purpose carver examples

- From raw data:
  - Adroit: very effective (Jpeg) carver
  - “Cohen carver”: very effective for Pdf and Zip
  - NTFS-compressed data carving
  - Netanalysis: recovers index.dat records
- Inside files:
  - Libmsiecf: Recovers removed index.dat records
  - Reglookup recover: Recovers removed registry entries
  - Libpff: Recovers removed e-mails from Pst/Ost files

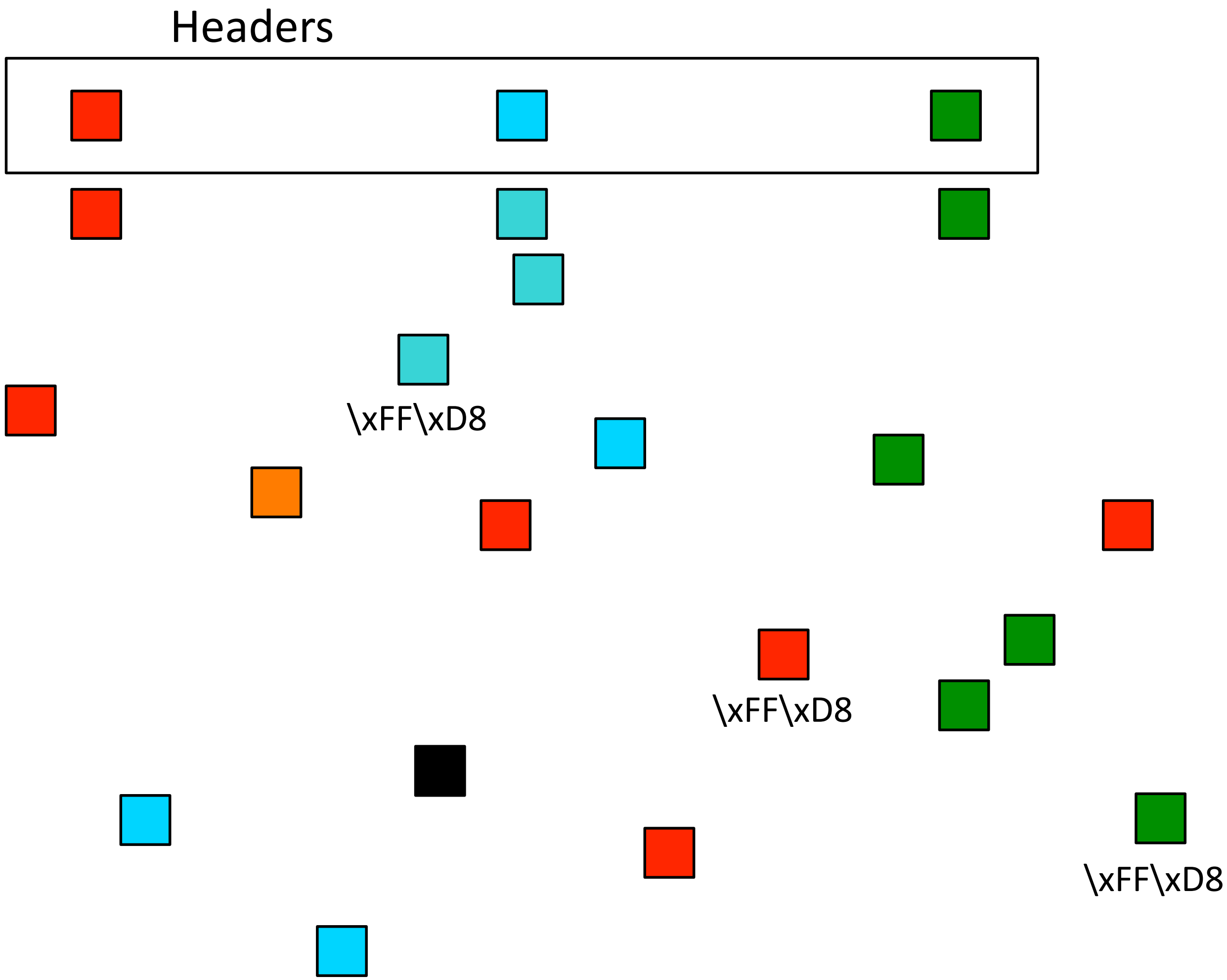
Adroit: Graph Theoretic Carvers

- Assume blocks are completely randomized



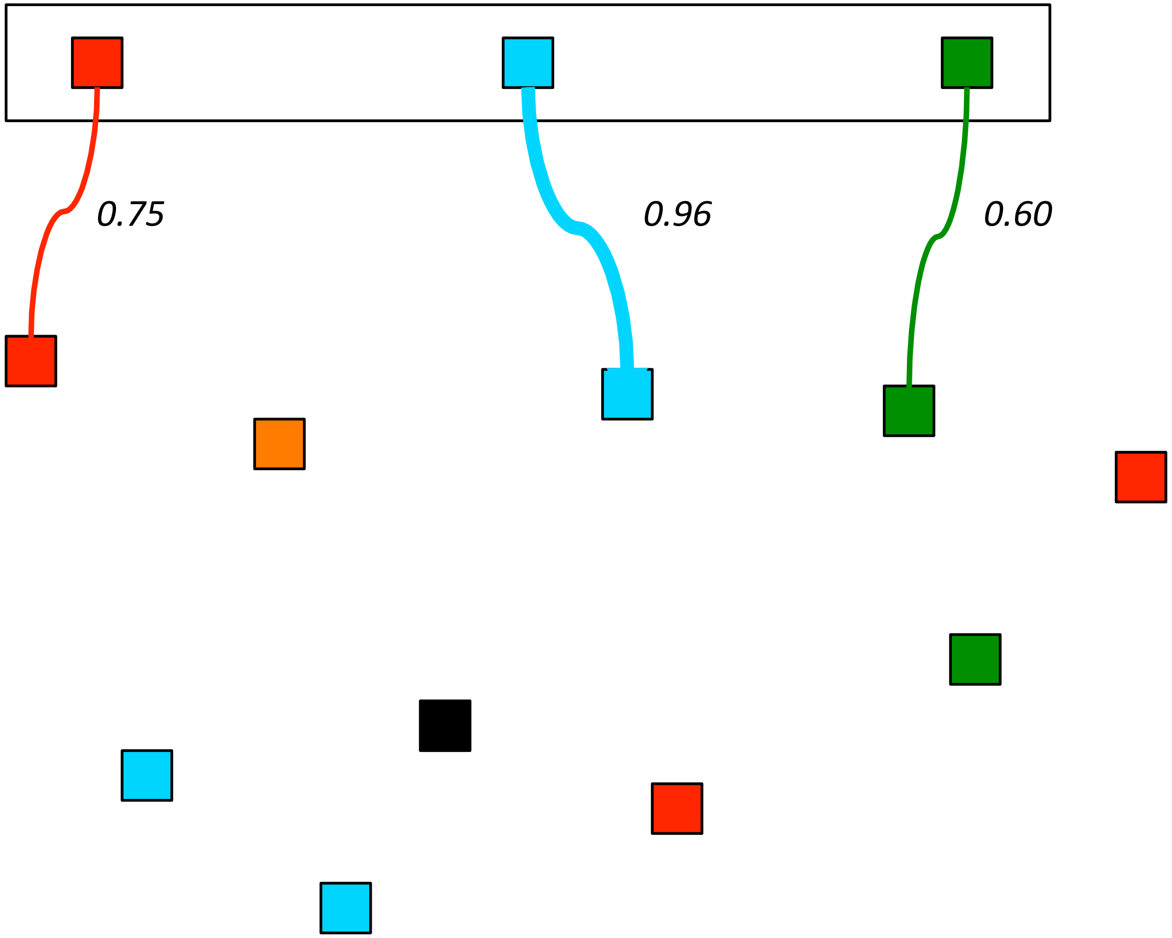
Adroit: Graph Theoretic Carvers

- Identify headers using keywords / signatures
- JPEG header is \xFF\xD8



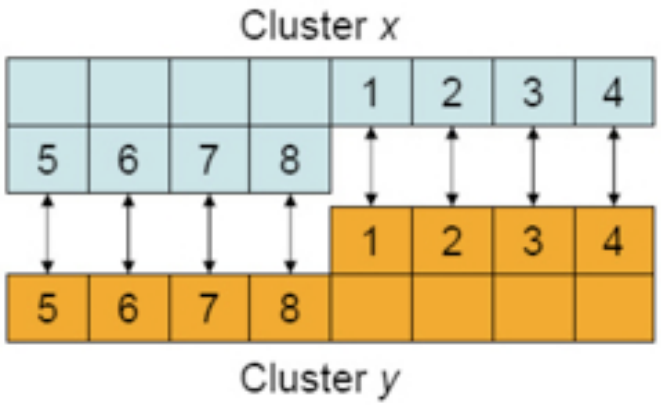
Adroit: Parallel Unique Path

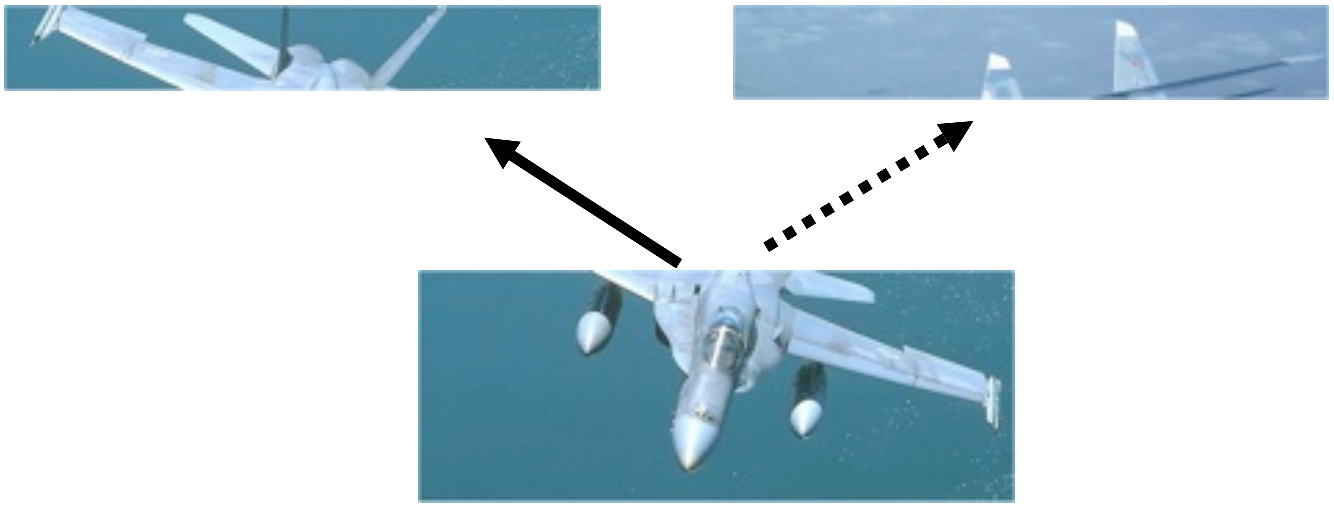
- For each header find best match (using matching metric)
- Choose the best overall match



Adroit: Matching Metric between blocks (Images)

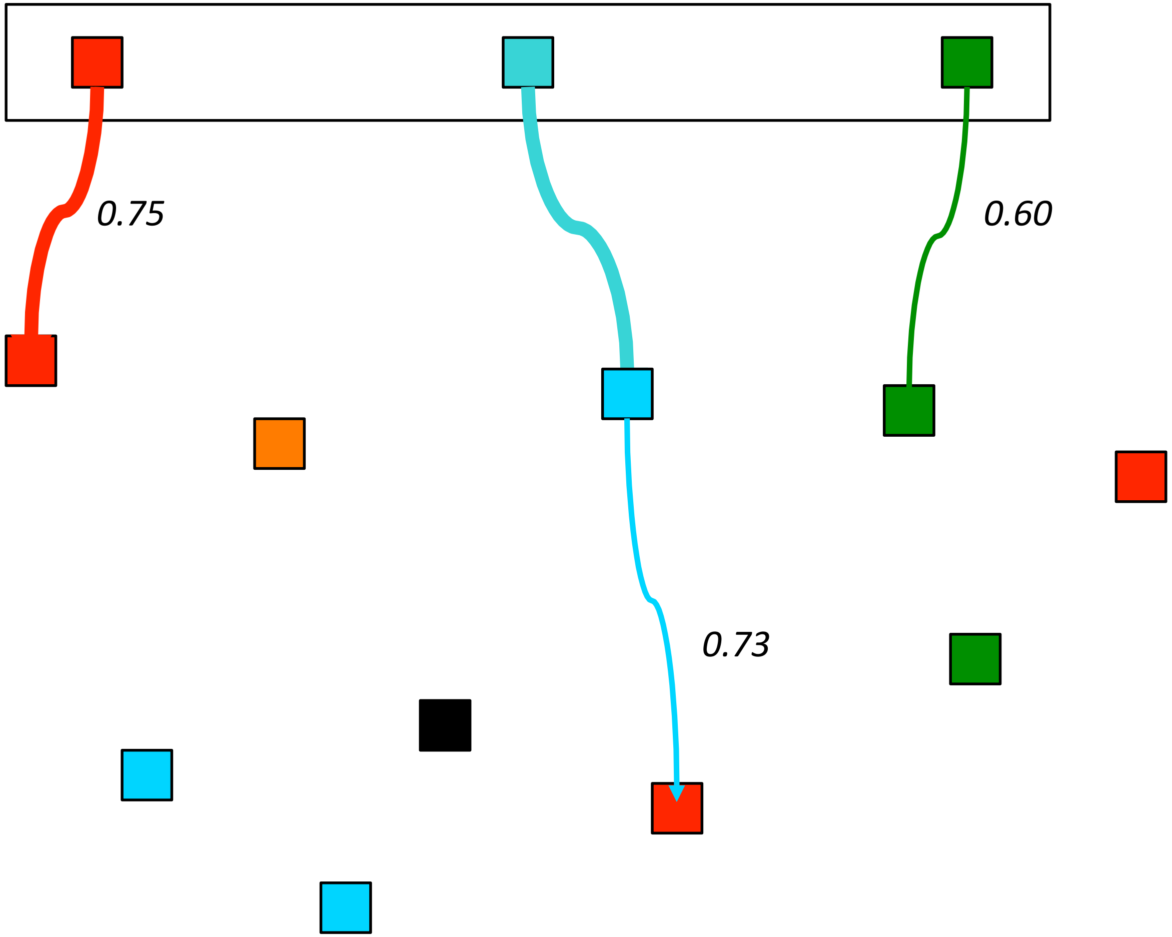
- For images: look at the boundary formed by the addition of a new block





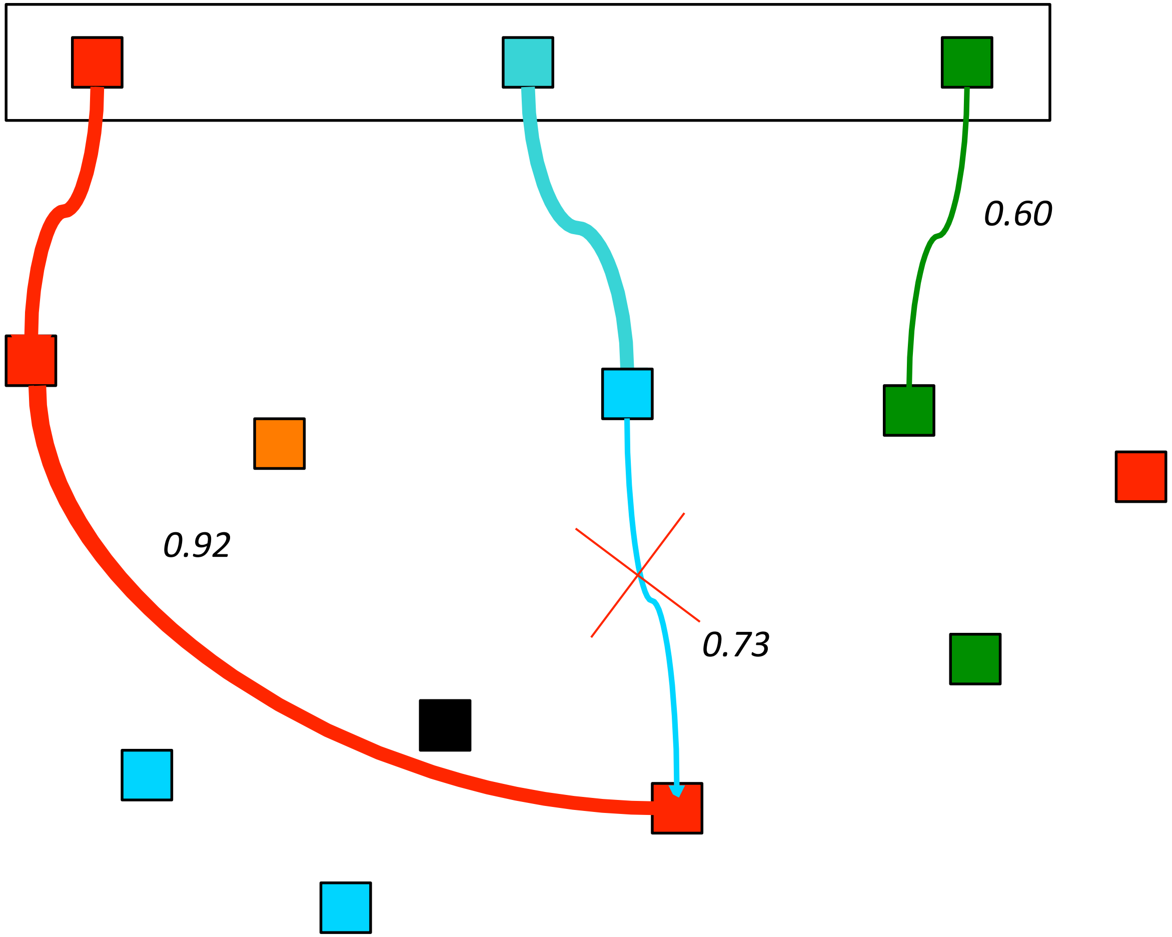
Adroit: Parallel Unique Path

- Find best match for recently added node
- Choose the best overall match again



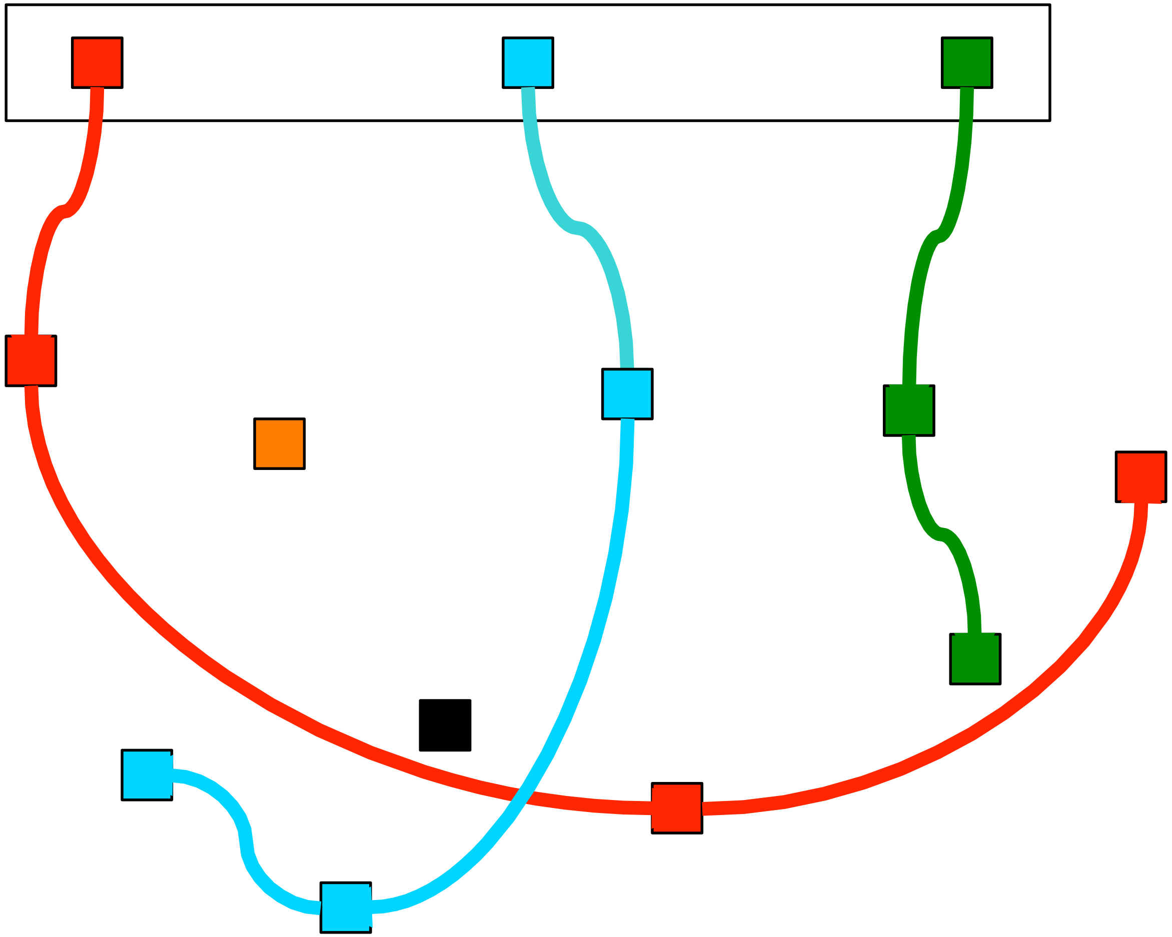
Adroit: Parallel Unique Path

- Repeat process
- Now a block is the best match for two files
- Choose the better of the two and continue



Adroit: Parallel Unique Path

- Repeat until all files are built or no more nodes can be chosen



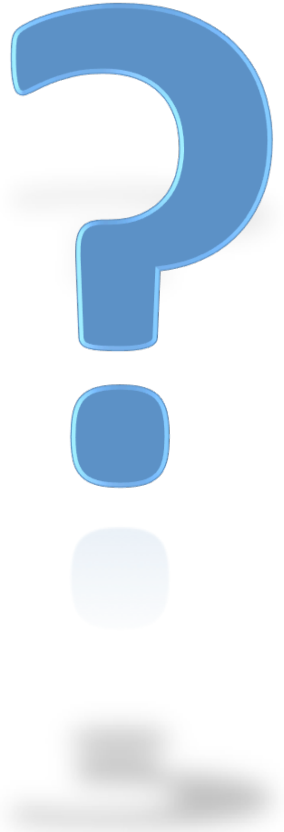
Adroit

- PUP by itself is too slow for effective real-world use in carving, but it is only part of the Adroit approach
- However, the rest is too complex for this short presentation…

Contents recap

1. Carving and basic file information
2. File Systems and Fragmentation
3. General File Carving Techniques
4. Measuring File Carving Quality
5. Specific Purpose Carving Tools

Questions?



Links 1/3

- General forensic tools:

Encase:	www.guidancesoftware.com
FTK:	www.accessdata.com

- General purpose carvers:

Scalpel:	www.digitalforensicssolutions.com/scalpel
Foremost:	foremost.sourceforge.net
Photorec:	cgsecurity.org
Revit:	revit.sourceforge.net

Links 2/3

- Specific purpose carvers and tools
  - Cohen carver: www.pyflag.net
  - Adroit: digital-assembly.com
  - Netanalysis: digital-detective.co.uk
  - Libmsiecf: libmsiecf.sourceforge.net
  - Reglookup: projects.sentinal-chicken.org/reglookup
  - Libpff: libpff.sourceforge.net
- Libraries and other tools:
  - Libewf: libewf.sourceforge.net
  - Ssdeep: ssdeep.sourceforge.net

Links 3/3

- Datasets:
  - FAT: dftt.sourceforge.net/test11
  - Ext2: dftt.sourceforge.net/test12
  - DFRWS 2006: dfrws.org/2006/challenge
  - DFRWS 2007: dfrws.org/2007/challenge
- Measuring file carving quality:
  - alexandria.tue.nl/extra1/afstversl/wsk-i/kloet2007.pdf