



# **CARVING UP COMPUTER CONTENTS**

*Morgan Davies*

***COMP3012 – Digital Forensics***

Where technology strides, so does digital media and the subsequent inevitability of data loss and hiding [4]. Combating this are data carving recovery tools, increasingly popular software that has been growing over the past decade. Almost to the point where they are nearly a must-have for convicting those guilty of committing digital offences [5].

However, what use are these tools if there is no research regarding which one is most effective for the situation? In this report, each data carving tool (Scalpel, Autopsy and FTK) will be evaluated for its effectiveness against various file types and structures. Accompanying discussions will reflect on the implications of the findings and recommendations and suggestions for which tools are most appropriate for specific situations.

# METHODOLOGY

This experiment will be multi-layered and hold weight that each carving scenario is unique. One single experiment will not reflect the effectiveness of each tool in every carving situation due to almost infinite variability in computer configurations [6].

Each tool will be judged against NIST's Forensic File Carving Test Assertions Plan 2014 [7], a set of standards and tests good forensic recovery tools should adhere to and certify under. By running and grading each tool out of 5 against each of the tests in Figure 1 and their performance against each other, an accurate picture should be formed concerning its flexible viability [14]. At the end of the process, there will be an overall winner but this does not imply the tool is excellent against every testcase.

|   |  |
|---|--|
| <b>FC-CA-01.</b> The tool shall report all accessible metadata entries for deleted files.   | <b>FC-01.</b> Sector aligned contiguous files with no padding data between files. (No padding) |
| <b>FC-CA-02.</b> The tool shall return one carved file for each supported file header signature from a source file that is present in the search arena. | <b>FC-02.</b> Byte aligned contiguous files with no padding data between files. (Byte shifted) |
| <b>FC-CA-03.</b> A carved file shall only contain data blocks from the search arena.  | <b>FC-03.</b> Contiguous files with padding data between files. (Cluster padded)               |
| <b>FC-CA-04.</b> All data blocks in a carved file shall originate in a single source file.  | <b>FC-04.</b> Sequential fragmented files. (Fragmented in order)                               |
| <b>FC-CA-05.</b> The file type of a carved file shall match the file type of its contents.  | <b>FC-05.</b> Non-sequential fragmented files. (Fragmented out of order)                       |
| <b>FC-CA-06.</b> The tool shall return carved files in a state that conforms to a valid file of the carved file type.                                   | <b>FC-06.</b> Intermixed fragments. (Braided pair)   |
| <b>FC-CA-07.</b> The tool shall ignore padding between targets.   | <b>FC-07.</b> Partial files. (Incomplete)  |
| <b>FC-CA-08.</b> The tool shall ignore padding between file fragments.  |  |
| <b>FC-CA-09.</b> The tool shall reorder non-sequential file fragments.  |  |
| <b>FC-CA-10.</b> If a target is incomplete, the carved file shall contain all source file blocks present in the search arena.                           |  |

*Figure 1: NIST's specification for how each tool will be evaluated as well as the test cases that produce evidence for this evaluation [7]*

Each tool will be working against the same image, DFRWS 2006, which contains a set of JPEG, ZIP, HTML, TXT and DOC files in various states of file fragmentation and filler data in between clusters. These files will form the backbone of the test cases FC-01 through 07. It is worth noting that some of these test cases will require advanced data carving, due to the difficulty of marrying up fragments scattered across a file system [8].

The hypothesis is broad considering the multitude of scenarios that could be produced by the experiment, but to touch on the core expectations:

- ◆ The propriety software (FTK) is more likely to perform better overall vs the open-source tools (Scalpel and Autopsy), partly due to previous experiments on the image [9] but also because private software is generally less prone to tech-debt bugs from newcomers to the codebase [10].
- ◆ However, it would be surprising if Scalpel and Autopsy did not perform as well in sectors more suited for advanced data carving. After all, they are tools specifically designed for carving (FTK is also designed for traditional digital forensics and password cracking) and FTK has been notoriously poor at carving fragmented files with little options for configuring optimisation options [11].
- ◆ Finally, the implications should correlate with recent anti-forensics trends in metadata obfuscation and altering [12, 13].





# RESULTS

Table Of Results containing individual tool performance against each testcase

| Scalpel (3/5)  | Autopsy (4/5)  | FTK (2/5)  |
|--|--|--|
| <p>With the default settings, there was little success carving useful data. Although scalpel.conf does contain entries for relevant files, it produced 8 empty ZIP files as a set of false positives (Appendix 1).</p> <p>With custom configuration adjustments, better results were obtained (including the addition of the Excel file type and modification of the doc and jpg file types to accept more alternative likely scenarios). It produced duplicate (Figure 2) but clean carves of the HTML, Excel, Doc and some image files.</p> <p>However, these carves were only partial in some aspects. While the HTML and Excel files were cleanly cut (surprising considering some were fragmented or intertwined), most doc files were not able to be opened in a word processor and some JPG files were only partially complete (Appendix 1). Crucially, only ZIP files that were not fragmented were carved successfully.</p> | <p>One immediate obvious advantage was the recovery of file metadata as well as content (Figure 3). With it's helpful UI, it pinpointed deleted files and even modification/creation timestamps (although, the MFT would also need to be successfully carved out to confirm these stamps are accurate [1]).</p> <p>Autopsy was also capable of extracting random or slack space data that was intermingled with valid files. Although ZIP files carved were not extractable, the compressed contents were. Perhaps not fully, but images were still visible and text files were nearly completely preserved, along with the metadata crucial for reconstructing large system structure maps [2].</p> <p>However, Autopsy did not produce any information regarding the Excel documents and does not appear to have a plugin to detect them (Figure 4).</p> | <p>Regarding initial setup, FTK is arguably the poorest performing tool in this category. It has default settings for Zip, Html and JPG files but any others needed to be custom made. These settings are also flawed compared to the open-source tools when it comes to carving larger than standard image files (Appendix 2).</p> <p>That being said, the settings and UI that are there are easy to use and effective at carving out the same files Scalpel and Autopsy has done. What supercedes it above those tools however, is the ability to carve and display image EXIF data (Appendix 2). FTK is able to carve out fragmented files within a certain threshold of random data and even the contents and structure of ZIP files (Appendix 2), although it also struggles with archives that have been fragmented.</p> <p>However, FTK is quite poor at data carving Doc and Excel files in this instance. Only one xls file was carved and was completely unreadable, with none of the doc files being picked up (even with custom carvers [3]).</p> |

|                                   |   |           |                                      |                                      |          |                  |               |                  |               |
|-----------------------------------|---|-----------|--------------------------------------|--------------------------------------|----------|------------------|---------------|------------------|---------------|
| # GIF and JPG files (very common) |   |           |                                      |                                      |          |                  |               |                  |               |
| gif                               | y | 5000000   | \x47\x49\x46\x38\x37\x61             | \x00\x3b                             | doc-4-0  | 05/04/2021 10:17 | File folder   | Views            | File Types    |
| gif                               | y | 5000000   | \x47\x49\x46\x38\x39\x61             | \x00\x00\x3b                         | doc-5-0  | 05/04/2021 10:17 | File folder   | By Extension     | Images (26)   |
| jpg                               | y | 500000000 | \xff\xd8\xff\xe0\x00\x10             | \xff\xd9                             | doc-6-0  | 04/04/2021 15:36 | File folder   | Videos (0)       | Audio (0)     |
| jpg                               | y | 500000000 | \xff\xd8\xff\xe1                     | \xff\xd9                             | doc-7-0  | 04/04/2021 15:36 | File folder   | Archives (2)     | Databases (0) |
| jpg                               | y | 200000000 | JFIF\xff\xd8\xff\xe0                 | \xff\xd9                             | htm-16-0 | 04/04/2021 15:36 | File folder   | Documents        | HTML (10)     |
| jpg                               | y | 200000000 | JFIF\xff\xd8\xff\xe0                 | \xff\xd9                             | jpg-0-0  | 04/04/2021 15:36 | File folder   | Office (6)       | PDF (0)       |
| # Word & Text documents           |   |           |                                      |                                      |          |                  |               |                  |               |
| doc                               | y | 50000000  | \xd0\xcf\x11\xe0\x1b\x1a\xe1\x00\x00 | \xd0\xcf\x11\xe0\x1b\x1a\xe1\x00\x00 | xls-9-0  | 04/04/2021 15:36 | File folder   | Plan Text (2)    | Rich Text (0) |
| doc                               | y | 20000000  | \xd0\xcf\x11\xe0\x1b\x1a\xe1\x00\x00 | \xd0\xcf\x11\xe0\x1b\x1a\xe1\x00\x00 | xls-14-0 | 04/04/2021 15:36 | File folder   | Executable       | By MIME Type  |
| doc                               | y | 50000000  | \xec\xa5\x10\x00 \xec\xa5\x10\x00    | \xec\xa5\x10\x00                     | xls-15-0 | 04/04/2021 15:36 | File folder   | application      | zip (2)       |
| doc                               | y | 10000000  | \xec\xa5\x10\x00 \xec\xa5\x10\x00    | \xec\xa5\x10\x00                     | zip-18-0 | 04/04/2021 15:36 | File folder   | image            | xhtml+xml (2) |
|                                   |   |           |                                      |                                      | audit    | 04/04/2021 15:36 | Text Document | octet-stream (5) | msword (6)    |

Figure 2: Duplicate entries in scalpel.conf that ultimately ended up carving the same files.

|   |  |  |                     |                     |
|---|--|--|---------------------|---------------------|
|  image_1.png |  |  | 0000-00-00 00:00:00 | 0000-00-00 00:00:00 |
|  image_2.png |  |  | 0000-00-00 00:00:00 | 0000-00-00 00:00:00 |
|  image_0.png |  |  | 0000-00-00 00:00:00 | 0000-00-00 00:00:00 |
|  image_0.png |  |  | 0000-00-00 00:00:00 | 0000-00-00 00:00:00 |

<

HexTextApplicationMessageFile MetadataContextResultsAnnotationsOther Occurrences

|                      |   |
|----------------------|---|
| Name                 | /img_dfrws.raw\ScarvedFiles\0034288.doc\image_1.png |
| Type                 | Local   |
| MIME Type            | image/png   |
| Size                 | 19893   |
| File Name Allocation | Allocated   |
| Metadata Allocation  | Allocated   |

Figure 3: Carved Image file metadata

|               |                  |               |
|---------------|------------------|---------------|
| File folder   | Views            | File Types    |
| File folder   | By Extension     | Images (26)   |
| File folder   | Videos (0)       | Audio (0)     |
| File folder   | Archives (2)     | Databases (0) |
| File folder   | Documents        | HTML (10)     |
| File folder   | Office (6)       | PDF (0)       |
| File folder   | Plan Text (2)    | Rich Text (0) |
| File folder   | Executable       | By MIME Type  |
| File folder   | application      | zip (2)       |
| File folder   | image            | xhtml+xml (2) |
| File folder   | octet-stream (5) | msword (6)    |
| File folder   | image            | wmf (2)       |
| File folder   | png (4)          | jpeg (17)     |
| File folder   | text             | plan (2)      |
| File folder   | html (9)         |               |
| Deleted Files | File System (0)  | All (34)      |

Figure 4: Carved Autopsy files, showing no entry for xls files



# DISCUSSION

The results of the experiment compared to the hypothesis are mixed. Although the open-source tools did perform better against files that required advanced carving (due to fragmentation or location within file or drive slack space [15]), proprietary software did not perform better overall due to the absence of carved doc and excel files. This is huge as when considering where data carving is often employed (e.g. against an illegal data vendor holding spreadsheets of citizen data [16]), it is crucial that forensic investigators can view incriminating evidence.

Previous surveys have shown that forensic investigators generally prefer to use GUI tools over CLI [17]. However, this experiment has proven that the Scalpel CLI tool can be just as effective in some cases. Due to the high configurability level of Scalpel (thanks to its malleable configuration file [9]), experienced digital forensics experts would likely find it more useful than the more rigid Autopsy and poorer performing FTK. The opposite is true for Autopsy, the GUI interface and acceptable default settings likely proving to be more attractive to newcomers.

However, in each tool, there was a common theme. None were able to effectively conduct advanced data carving on every occasion (and especially not with the ZIP files). Perpetrators therefore would be more likely to attempt to hide incriminating data within fragmented archives to prevent detection or even de-rail the investigation entirely via Archive Bombs [18]. Carving tools need to be improved when working with fragmented files and those that are utilising file slack space to hide. One way of doing so could be by utilising machine learning to find patterns within fragments, allowing the clustering of fragments byte by byte [19]. Hidden files could be identified this way if there are lone fragments that can fit together.

The main advantage of hiding data within slack space is due to its volatility (can be easily and quickly overwritten if the suspect wishes) and inability of forensic tools to work well if the MFT is compromised [20]. Anti-forensic tools and research already exist to achieve this [21], so it would be wise for forensic experts to develop counters and ensure slack space is searched for the sort of information that could be found there (e.g. cryptographic keys).

# CONCLUSION

Data hiding techniques continue to evolve and some data recovery tools seem unable to keep up with their advancements. Advanced data hiding using slack space, fragmentation and magic number obscuration continue to pose a significant counter towards recovering incriminating data. The experiment has shown that this is true, and that some tools can conduct this advanced carving, they are not always 100% successful. Further development is needed on these tools to be able to extract files that have been purposely or unintentionally, well hidden.

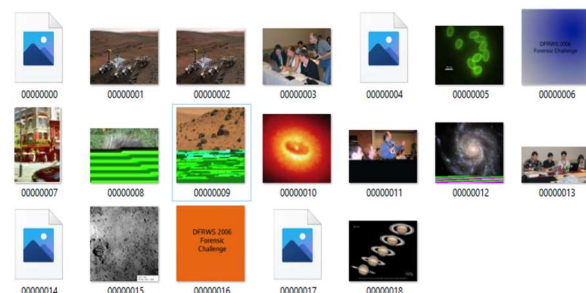
The discussion has shown that guilty parties hold an interest in anti-forensics as a means of claiming ignorance in crimes they have committed. Efforts should be made to unearth the employed techniques successful suspects have used, mainly to reverse engineer their methodology and develop generic countermeasures against them.

# References

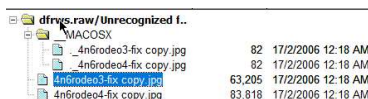
- [1] T. Knutson, "Filesystem Timestamps: What Makes Them Tick?," 23 3 2016. [Online]. Available: <https://www.sans.org/reading-room/whitepapers/forensics/filesystem-timestamps-tick-36842>. [Accessed 5 4 2021].
- [2] R. Bedoll and C. Kimball, "The importance of metadata in mass-storage systems," 10 5 1990. [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=113579>. [Accessed 6 4 2021].
- [3] A. Stochlia, "Custom Carvers," AccessData, 28 1 2015. [Online]. Available: <https://support.accessdata.com/hc/en-us/articles/203423159-Custom-Carvers>. [Accessed 6 4 2021].
- [4] Nurhayati and N. Fikri, "The analysis of file carving process using PhotoRec and Foremost," 10 8 2017. [Online]. Available: <https://ieeexplore.ieee.org/document/8320663>. [Accessed 31 3 2021].
- [5] N. Alherbawi, Z. Shukur and R. Sulaiman, "A survey on Data Carving in Digital Forensics," 2016. [Online]. Available: [https://www.researchgate.net/profile/Nadeem-Alherbawi/publication/316453903\\_A\\_survey\\_on\\_data\\_carving\\_in\\_digital\\_forensic/links/58ff02efaca2725bd71e321a/A-survey-on-data-carving-in-digital-forensic.pdf](https://www.researchgate.net/profile/Nadeem-Alherbawi/publication/316453903_A_survey_on_data_carving_in_digital_forensic/links/58ff02efaca2725bd71e321a/A-survey-on-data-carving-in-digital-forensic.pdf). [Accessed 31 3 2021].
- [6] E. Alshammary and A. Hadi, "Reviewing and Evaluating Existing File Carving Techniques for JPEG Files," 4 8 2016. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7600210>. [Accessed 31 3 2021].
- [7] National Institute of Standards and Technology, "Forensic File Carving Tool Test Assertions and Test Plan," 4 2014. [Online]. Available: <https://www.nist.gov/system/files/documents/2017/05/09/fc-atp-public-draft-01-of-ver-01.pdf>. [Accessed 31 3 2021].
- [8] M. Cohen, "Advanced carving techniques," 12 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1742287607000813>. [Accessed 31 3 2021].
- [9] T. Laurenson, "Performance Analysis of File Carving Tools," 2013. [Online]. Available: [https://link.springer.com/content/pdf/10.1007%2F978-3-642-39218-4\\_31.pdf](https://link.springer.com/content/pdf/10.1007%2F978-3-642-39218-4_31.pdf). [Accessed 31 3 2021].
- [10] A. Boulanger, "Open-source versus proprietary software: Is one more reliable and secure than the other?," 2005. [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5386727>. [Accessed 31 3 2021].
- [11] V. Roussev, "Scalpel: A Frugal, High Performance File Carver," 2005. [Online]. Available: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.77.6222&rep=rep1&type=pdf>. [Accessed 31 3 2021].
- [12] S. Berinato, "The Rise of Anti-Forensics," 8 6 2007. [Online]. Available: <https://www.csoonline.com/article/2122329/the-rise-of-anti-forensics.html>. [Accessed 31 3 2021].
- [13] A. Dewald and S. Seufert, "AFEIC: Advanced forensic Ext4 inode carving," 3 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1742287617300270>. [Accessed 31 3 2021].
- [14] A. Merola, "Data Carving Concepts," 2008. [Online]. Available: <https://www.sans.org/reading-room/whitepapers/forensics/data-carving-concepts-32969>. [Accessed 5 4 2021].
- [15] V. Roussev, "Forensics Knowledge Area," 2018. [Online]. Available: [https://cybok.org/media/downloads/Forensics\\_KA\\_-\\_draft\\_for\\_review\\_July\\_2019.pdf](https://cybok.org/media/downloads/Forensics_KA_-_draft_for_review_July_2019.pdf). [Accessed 6 4 2021].
- [16] R. Goichman, "Leaked Voter Info and Illegal Electioneering: Inside Netanyahu's Election Day App," Haaretz, 11 3 2021. [Online]. Available: <https://www.haaretz.com/israel-news/elections/.premium-leaked-voter-info-zero-privacy-inside-netanyahu-s-election-app-1.9611603>. [Accessed 6 4 2021].
- [17] H. Hibshi, T. Vidas and L. Cranor, "Usability of Forensics Tools: A User Study," 2011. [Online]. Available: <https://www.cs.cmu.edu/~hhibshi/pdf/HVC11.pdf>. [Accessed 6 4 2021].
- [18] R. Harris, "Arriving at an anti-forensics consensus: Examining how to define and control the anti-forensics problem," 9 2006. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1742287606000673>. [Accessed 6 4 2021].
- [19] A. Ai-Sadi, M. Bin-Yahya and A. Almulhem, "Identification of image fragments for file carving," 12 2013. [Online]. Available: [https://www.researchgate.net/publication/260509372\\_Identification\\_of\\_image\\_fragments\\_for\\_file\\_carving](https://www.researchgate.net/publication/260509372_Identification_of_image_fragments_for_file_carving). [Accessed 6 4 2021].
- [20] S. Kloet, "Measuring and Improving the Quality of File Carving Methods," 29 10 2007. [Online]. Available: [http://index-of.co.uk/readings/Kloet\\_2007.pdf](http://index-of.co.uk/readings/Kloet_2007.pdf). [Accessed 6 4 2021].

- # Appendix 1

*All jpg files were at least partially carved but none of the fragmented files were done so fully (missing pixels in the images).*

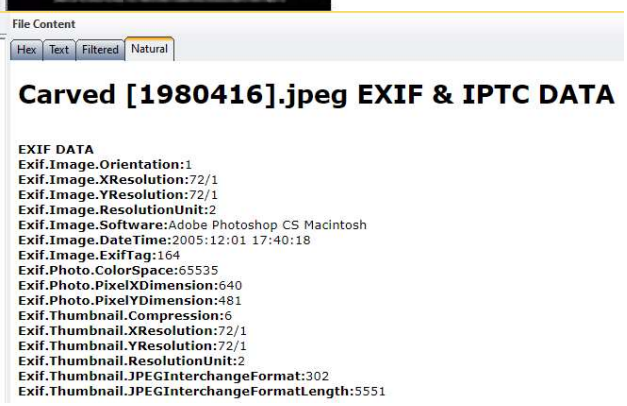


## Appendix 2



A photograph showing three people from behind, sitting at a table and looking at a laptop. The laptop screen displays a list of text, which appears to be a log or a code output. On the table, there are several bottles, including a beer, and some papers. The setting seems to be a casual meeting or a workshop.

*FTK view of a carved zip archive and one of the images that was inside it.*



*Carve of a large image file (>1mb) that had been fragmented.*



*Image EXIF data carved out (along with the image contents itself) by FTK.*