April 2014

Forensic File Carving Tool Specification

Draft Version 1.0 for Public Comment

National Institute of Standards and Technology
U.S. Department of Commerce

Abstract

This document defines requirements for digital file carving forensic tools that extract and reconstruct files without examination of file system metadata. The specification is limited to tools that identify inaccessible (deleted or embedded) files from file data content. Such tools exploit the unique data signatures of certain file types to identify starting and ending data blocks of these file types. In addition, file system allocation policies often keep file data blocks contiguous and sequential. For such contiguous sequential block placement identification of starting and ending data blocks may be sufficient to carve complete files. In other non-contiguous or non-sequential block placement, file reconstruction by carving is problematic.

CONTENTS

1 Introduction

There is a critical need in the law enforcement community to ensure the reliability of computer forensic tools. A capability is required to ensure that forensic software tools consistently produce accurate and objective results. The goal of the Computer Forensic Tool Testing (CFTT) project at the National Institute of Standards and Technology (NIST) is to establish a methodology for testing computer forensic software tools by development of general tool specifications, test procedures, test criteria, test sets, and test hardware. The results provide the information necessary for toolmakers to improve tools, for users to make informed choices about acquiring and using computer forensics tools, and for interested parties to understand the tools capabilities. Our approach for testing computer forensic tools is based on well-recognized international methodologies for conformance testing and quality testing. This project is further described at http://www.cftt.nist.gov/.

The CFTT program is a joint project of the Department of Homeland Security, the National Institute of Justice, and the NIST Law Enforcement Standards Office and Information Technology Laboratory. CFTT is supported by other organizations, including the Federal Bureau of Investigation, the U.S. Department of Defense Cyber Crime Center, U.S. Internal Revenue Service Criminal Investigation Division Electronic Crimes Program, U.S. Department of Homeland Security's Bureau of Immigration and Customs Enforcement, U.S. Customs and Border Protection and the U.S. Secret Service. The objective of the CFTT program is to provide measurable assurance to practitioners, researchers, and other applicable users that the tools used in computer forensics investigations provide accurate results. Accomplishing this requires the development of specifications and test methods for computer forensics tools and subsequent testing of specific tools against those specifications.

Frequently during a forensic examination, data is discovered on the target media that is not part of any active or visible file. Although this data can still be examined at the byte level (e.g., string searching), the higher-level information is not apparent. If the data associated with a particular file could be identified and examined in its usual presentation format for the given file type, e.g., as a picture or video, this may provide more complete information. An example of this would be where a graphics file, carved from unallocated space, could be viewed—potentially providing more information than a simple string search. Many of the forensic tools used by investigators identify files that have been deleted and allow the operator to recover them by file carving. This allows the investigator to examine the carved file in the original format (e.g., a graphics file viewer).

A fundamental problem is that the potential uncertainty present in any recovery effort leads to a reduced level of confidence in the information recovered. Specifically with file carving, the data recovered may be commingled with data from other deleted files, allocated files, or even from non-allocated space.

2 Purpose

This document defines the functional requirements for tools used within forensic investigations to carve files. That is reconstructing deleted or extracting embedded files based on file content.

These requirements were developed through a combination of processes including but not limited to file carving research, personal interviews with forensic investigators, and informal discussions with individuals who are experts in the field of forensic investigation and depend on the results of file carving tools. Additionally, as this document evolves, feedback will be incorporated from a variety of sources, and will be posted to our web site at http://www.cftt.nist.gov for comments.

These requirements are used to derive test assertions and test methods used to determine whether a specific tool meets the requirements. The assertions are described as general statements of conditions that can be checked after a test is executed. Each assertion generates one or more test cases consisting of a test protocol and the expected test results. The test protocol specifies detailed procedures for setting up the test, executing the test, and measuring the test results. The test assertions, test methods and test protocols are found in an accompanying document, *Forensic File Carving Tool Test Assertions and Test Plan*, located on the CFTT web site, located on the CFTT web site, http://www.cftt.nist.gov/.

3 Scope

The scope of this specification and requirements document is limited to software that is used for file carving. The proper or improper use of a tool is not within the scope of this specification.

The specifications and requirements for file carving are high-level, and are based on the following assumptions.

- The tools are used in a forensically sound environment.
- The individuals using these tools adhere to forensic principles and have control over the environment in which the tools are used.
- The carving tool input is a file or set of files that might be produced by a forensic acquisition tool acquiring digital media such as secondary storage or volatile memory.
- The files used test input to carving tools were created in a process that places file data blocks in a manner similar to how end-user activity would locate file data blocks.

4 Definitions

This section contains definitions of terms used in this specification document. Although there may be commonly accepted definitions for some of the terms, the context of this document may require a specific meaning.

Carved File: A file created by a carving tool purported to be one of the source files present in the search arena.

Data Block: File system specific data allocation unit (block), usually a multiple of 512 bytes. Some file systems may use other terms to describe a *data block* such as, *cluster* in FAT file systems.

File Carving: Reconstructing deleted files from unallocated storage or extracting embedded files from a container file, based on file content; file system metadata may be a secondary consideration or completely ignored.

File-footer signature: A data string that identifies the end of a file. The string must be unique for a given file type. The string may begin anywhere within a data block.

File-header signature: A data string that identifies the beginning of a file. The string must be unique for a given file type. The string usually begins on a data block boundary, but it may begin anywhere within a data block.

Metadata: The associated periphery information or attributes that describe a file such as name, time-based metadata (creation, modification, and last accessed times), access rights, ownership, and location.

Search arena: An acquisition file to be searched, e.g., the file obtained by acquiring unallocated space from a secondary storage device or acquiring primary memory from a running system. The search arena is composed of source file data blocks and other unspecified data blocks. A given source file may be complete, incomplete, fragmented, contiguous, sequential or non-sequential.

Source file: One of several files used to construct the search arena. All or part of a source file might be used. A carving tool should return a carved file for each complete source file in the search arena. The carved file returned by the carving tool should be visually identical to the original source file.

5 File Carving Background

File carving is widely used in digital investigations to extract information from unallocated storage. Usually file carving is applied to file types with a recognizable structure so that unallocated space can be scanned for file components that are reassembled into complete files. Under some conditions this is an easy task. If the file has easily identified beginning and ending content and is contiguously allocated then carving is simple. However, the reality of file fragmentation complicates the task considerably.

Categories of files that are common targets of file carving include:

- Still Picture: JPG, GIF, PNG, BMP & TIF
- Videos: MP4, AVI, MOV, 3GP, OGV & WMV
- Audio: MP3, WAV, AU & WMA
- Document: DOC, DOCX, XLS, XLSX, PDF, PPT & PPTX,
- WEB: HTML, SQLite & chat
- Archive: ZIP, RAR, 7Z, GZ & TAR
- Misc: exec, logs, etc.

For the most part, common file system block allocation policies assist in the recovery of data on the drive, regardless of the type of file system the data resides on. Files can be completely recovered if at least three conditions are present:

1. There is a uniquely identifiable start data block.
2. The file is contiguously and sequentially allocated.
3. There is a uniquely identifiable final data block.

Several problems may occur in practice that file carving tools might be required to deal with:

- Not all file types have a uniquely identifiable final data block and may require tools to guess where the end of the file is located.

- If a complete source file is present in the search arena, but the file is fragmented then the carving tool needs to be capable of identifying all file fragments and assembling the fragments in the correct order. This is not an easy task and may not be possible is many cases.

- If a source file is incomplete within the search arena then it may be possible to assemble the first or last part a file from the available data, but this may not be possible is many cases.

5.1 References (Informative)

It is important to note that these references are primarily informative.

Carrier, (2003). "File System Analysis Techniques: Sleuth Kit Reference Document." Available at http://www.sleuthkit.org/sleuthkit/docs/ref_fs.html.

Crane, (1999). "Linux Ext2fs Undeletion mini-HOWTO." Available at http://www.tldp.org/HOWTO/Ext2fs-Undeletion.html.

Erdelsky, (1993). “A Description of the DOS File System.” Available at http://www.alumni.caltech.edu/~pje/dosfiles.html.

Himmer, (2000). “File Systems HOWTO.” Available at http://www.faqs.org/docs/Linux-HOWTO/Filesystems-HOWTO.html.

Microsoft, (2004). “Description of the FAT32 File System.” Available at http://support.microsoft.com/default.aspx?scid=http://support.microsoft.com:80/support/kb/articles/q154/9/97.asp&NoWebContent=1.

NIST, (2004). “General Test Methodology for Computer Forensic Tools,” Available at http://www.cftt.nist.gov/.

Anandabrata Pal and Nasir Memon. (2009, March) www.smartcarver.com. [Online]. www.smartcarver.com/technology/research/pubs/ieee-spm-2009.pdf

Antonio Merola. (2008, November) www.sans.org. [Online]. http://www.sans.org/reading_room/whitepapers/forensics/data-carving-concepts_32969

Brian Carrier, Eoghan Casey, and Venema Wietse. DFRWS 2006 Forensics Challenge File Image Layout. [Online]. http://dfrws.org/2006/challenge/layout.shtml

Brian Carrier, Eoghan Casey, and Venema Wietst. DFRWS 2007 Forensics Challenge. [Online]. http://dfrws.org/2007/challenge/layout.shtml

Nicholas A. Mikus. Basic Data Carving Test #1. [Online]. http://dftt.sourceforge.net/test11/index.html

Nicholas A. Mikus. Basic Data Carving Test #2. [Online]. http://dftt.sourceforge.net/test12/index.html

S.J.J. Kloet, "Measuring and improving the quality of file carving methods," Department of Mathematics and Computer Science, Eindhoven University of Technology, Almere, Master's Thesis 2007.

Simson Garfinkel, Paul Farrell, Vassil Roussev, and George Dinolt, "Bringing science to digital forensics with standardized forensic corpora," in DFRWS, Montreal, 2009, pp. 2-11.

S. Garfinkel, "Carving Contiguous and Fragmented Files with Fast Object Validation," in Proceedings of Digital Forensic Research Workshop (DFRWS), Pittsburg, 2007, pp. 2-12.

G. Richard Golden III and Vassil Roussev, "Scalpel: A Frugal, High Performance File Carver," in Proceedings of Digital Forenwsics Workshop (DFRWS), New Orleans, 2005, pp. 1-10. [Online]. roussev.net/pdf/2005-DFRWS--scalpel.pdf

Ahmed Patel, Mustafa Mat Deris Kamaruddin Malik Mohamad, "Carving JPEG Images and Thumbnails Using Image Pattern Matching," in 2011 IEEE Symposium on Computers & Informatics , Kuala Lumpur, 2011, pp. 78-83.

Anabadrata Pal, Husrev T Sencar, and Nasir Memon, "Detecting file fragmentation point using sequential hypothesis testing," in Proceedings of the Digital Forensic Research Workshop (DFRWS), Baltimore, 2008, pp. 2-13.

Husrev T Sencar and Nasir Memon, "Identification and recovery of JPEG files with missing fragments," in DFRWS, pp. 88-98.

Kamaruddin Malik Mohamad, Ahmed Patel, Tutut Herawan, and Mustafa Mat Deris, "myKarve: JPEG Image and Thumbnail Carver," Journal of Digital Forensic Practice, vol. 3, no. 2-4, pp. 74-97, January 2010.

Simson L Garfinkel, Aleatha Parker-Wood, Daniel Huynh, and James Migletz, "An Automated Solution to the Multiuser Carved Data Ascription Problem," IEEE Transactions on Information Forensics and Security, vol. 5, no. 4, pp. 868-882, December 2010.

6 Requirements

The requirements section is divided into two parts. The first, *Requirements for Core Features*, are those features that should be present in all tools. The second is the *Requirements for Optional Features*. These features, on the condition they are present, are used to report on the tool capabilities. If a feature is not present, then requirements for those features will not be tested.

6.1 Requirements for Core Features

All file carving tools must support the following requirements.

FC-CR-01 The tool shall return one carved file for each supported file header signature from a source file that is present in the search arena.

FC-CR-02 A carved file shall only contain data blocks from the search arena.

FC-CR-03 All data blocks in a carved file shall originate in a single source file.

FC-CR-04 The file type of a carved file shall match the file type of its contents.

FC-CR-05 The tool shall return carved files in a state that conforms to a valid file of the carved file type.

6.2 Requirements for Optional Features

No optional features are identified at this time.