SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TNC TRANSACTIONS ON NETWORKS AND COMMUNICATIONS

VOLUME 8, NO. 1
ISSN: 2054 -7420

File Carving
Survey on Techniques, Tools and Areas of Use

[1]Ali Mohammed Abdullah Ali Raqpan, [2]Ziad Saif Alrobieh

[1]Department of Information Technology, Alsaeed Faculty for Engineering & Information Technology, Taiz University, Taiz, Republic of Yemen;

[2]Department of Communication & Computer Engineerin , Alsaeed Faculty for Engineering & Information Technology, Taiz University, Taiz, Republic of Yemen;

ali1raqpan@gmail.com; ziadrh@yahoo.com

ABSTRACT

In digital and computer forensics, file carving is a very hot research topic. That is the main reason why the research is needed to be focused on improving file carving techniques, so that digital investigation can obtain important data and evidence from damaged or corrupted storage media. In the digital forensic investigation, analyzing the unallocated space of storage media is necessary to extract the deleted or pre-written files when the file system metadata is missing or corrupted. Data carving can be defined as a method to recover the file from unallocated space based on different factors such as file type, information of the file (Header/Footer), or the contents of the file. Research in this area focuses on technological improvements in terms of tools and techniques over the past years. The studies examine different techniques of data carving, especially multimedia files (eg. images and videos). The work file carving is classified into three categories classic carving techniques, intelligent carving techniques and smart carving techniques. Moreover, there are seven popular multimedia carving tools that are mostly used and experimentally evaluated are presented. We conclude that proposing new advanced method for carving multimedia files still open and new direction for future research. This is because the fragmentations and compression are very commons used and useful for these kind of files.

Keywords: Digital Forensics, Data Recovery, Multimedia File Carving, Data Carving, File Carving Tools.

1 Introduction

In the last ten years, the rate of information has obtained great importance, so the idiom was introduced" Information society"[1] to refer to society these days. Information generation, processing, and distribution have become essential activities in different areas such as economics, politics and culture. The use, development and dissemination of electronic equipment as well as the interchange of information has increased steadily. In recent years according to the law "Kryder "[2] the space of magnetic storage media like hard disks grow even faster than processor speed. The processer speed doubles every eighteenth months, storage capacity has increased fifty million times since the drive began in 1956. In 2008 the number of computers used have exceeded one-billion-mark top of mobile phones worldwide. In 2009 the number of devices was 3.9 billion devices and in 2010 was 4.2 billion devices. Those numbers continue to increase by 7% in just one year. The number mention by the region affected by the offense related to Information Technology (IT) infrastructure and computers in general. In 2011, the federation of

DOI: 10.14738/tnc.81.7636
Publication Date: 04th March 2020
URL: http://dx.doi.org/10.14738/tnc.81.7636

German police to combat crimes announced. The total number of crimes committed using computers growing by 12.5% while the detection rate decreased by 1.7%[3]. These developments can be traced back to the increased rate with computing devices and the vast amount of data to be processed in any case. The recent verification conducted by the Francophone Association for Digital Investigation(AFSIN) each suspect in a criminal case has an average of 140 hard disks and 140 Compact Discs (CDs) or Digital Video (DVDs) and four memory cards and Universal Serial Bus(USB) [4]. When analyzing business cases, the volume of data is higher through up to 13 hard disks and 14 terabytes for one case [5]. Whereas the purpose of this research paper is to conduct an analytical study of the survey on File Carving, through which the work of extracting and retrieving structured data, which depends on a set of characteristics most prominent Header-Footer, size and some mechanisms, and where it does not depend on the type of the operating system, is inferred. The instrumentation measurement mechanism is based on finding both Recall and Precision. This mechanism was adopted through the Digital Forensics Research Workshop (DFRWS) challenge, and these tools work on different operating systems and have different efficiency and performance. One of the aims of the study is to clarify some legal frameworks for this technology, which were based on the Daubert criterion to calculate the resulting evidence of this technology, and some aspects of this study were dropped on the Yemeni community in order to contribute to the consolidation of this technology.

The reminder of this paper is organized as follows. Section 2 presents the prospecting file “file carving”, Section 3 the Idea. Section 4 reviewed the tools used with file carving. Section 5 presents ethical and framework of file carving. Section 6 discusses legal framework for file carving. Section 7 explains digital evidence in Yemen legally. In section 8 technical framework of file carving are presented. Section 9 displays the smart carving. Section 10 logistic framework of file carving are presented. Finally, section 11 the conclusion is presented.

2 Prospecting Files "File Carving"

File Carving: a procedure to read-out data from driver or other storage device used in digital forensics without the need to help the file system that originally created the file. It is away to recover files with unallocated copies without any files information and is used to recover data and execute criminal investigation. The process is called "carving" is a generic term for extracting structured data from meta data based on the coordination of specific properties displayed in structured data. [6] [7].

3 The Idea

When we delete files from our computer or device. It is not completely lost until a site is deleted from its memory while scanning the device. Many of its fragments remains in unallocated memory and can be rebuilt in theory. Simpler, most file system do not erase data altogether.

3.1 The Difference Between the Recovery & File Carving

Recovery: it revives the file from the remaining fragments in the internal memory using only recovery mechanisms depending on the different operating system [8].

Files carving: operating systems are able to recognize and process file in storage disks. The ability of the operating system to distinguish files types is done by reading some of the values that exist at the beginning and end "header and footer" these values specify the file type so that the operating system or program reads these values and is identified as a file.

For example, a file for an image of type (JPG); the file extension that is written at the end of the name after a period such as (MP3) it never determines the file type.

Then comes the analysis phase of the result related to the file being excavated and to extract more valuable information from the file in it's from and image. Just help to understand the file link for the purpose for which the recovery and analysis work is called. The file carving mechanism has a number of steps to complete data extraction and calculate evidence from this data so that the tools recovery process first begins, then the analysis process and this step come in order to know the integrity of the evidence from distortion, then the sorting step comes to the evidence group, and finally the verification process of matching what has been retrieved With what was intended to be retrieved, and from these steps it is clear the integration process between the human factor represented by the specialist for this technology and the tool used.

Is a cooperative integration process between the various components to accomplish the purpose intended to use the process of prospecting, which is named the highest goals are in the criminal investigation of electronic, In the Fig. 2. Turns out the integration between human & machine in the process of exploration for files and remarkable. That it is a process of repetition we see verification automatically and to human and also analysis, despite the difference in the time of each process.

For example, (file) command in the operating system "Linux" it can recognize almost kind of file by reading this values. Example if we open an audio file of type (MP3) using the editor (hexa-decimal) we will find the value 494433 it is the head as head as in the Fig.1

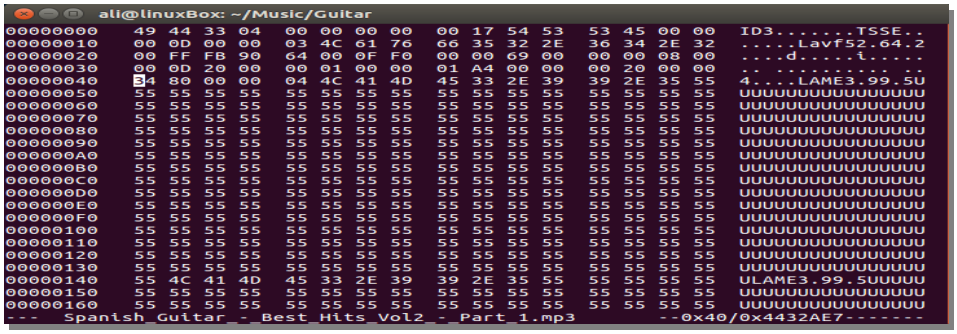


Figure. 1. Header-Footer of File in Linux

3.2 Problems Fixed by File Carving

This method is used by digital forensic experts to retrieve in some cases related to pornography, especially concerning children and minors. As well as digital fraud, law enforcement agents are often able to retrieve more images from suspected hard drive using "techniques of carving " as case of the hard disks and memory storage that the US navy took from Osama Bin Laden's place during their attack. Techniques were used "carving of file " extract very bit of information from the information media.

3.3 Areas of Use for File Carving

Used in digital forensics to read-out "extract" data from the drive or other storage devices without the need to help the file system to originally create the file it is also a way to recover file with unallocated space without any file information, it is used to retrieve data and execute digital criminal investigation.

4 The Tools are Used for File Carving Process

Data recovery tools function are a significant role in most criminal investigation because evil users are smart and will always try to delete anything that indicates their illegal action.

Some remarkable tools for carving of file:Scalpel, FTK, Encase, Foremost, Photorec, Revit, Test Disk, Magic Rescue and F-Engrave.

Table 1. Tools with Recall and Precision of File Carving

Tools of File Carving	Recall	Precision
Scalpel	0.333	0.003
FTK 1.81	0.4	0.6
Encase 6.7	0.47	0.5380
FTK 3.0	0.676	1.0
Foremost	0.8	0.857
Photorec	0.933	1.0
Revit	0.933	1.0

5 Ethical and Societal Framework for File Carving

The bad aspect is to retrieve photos or personal file of someone. Before you decide to get rid of your electronic devices either computer or phone. You have to know that you may give a chance for intruders to get to know all your secret through deleted picture and file after restoring them from the file recovery software. Keeping your important photos and file on your computer or phone may make you a victim of snoopers.

It is a software for recovering files and image which is used primarily to recover files that have been deleted files by mistake others can know your secrets by returning deleted file from your computer or phone, which may be where the image is important or special for some girls or young people or contain important evidence for large institutions. Hence note that the use of this technique is a double –edged sword and therefore may be supports of this technique and there may be opponents. it is one of the issues that make this technology opposition to the issue of violation of privacy, for example which has become the majority of societies codified and taken into account when issuing any technology so as not to exceed.

5.1 Problems may be Caused by File Carving

Personal Computer (PC) is different from a desktop computer because it contains private information, and sometime confidential linked to the work or life of the other, therefore, laptop owners prefer to have a password or a specific code that does not open the device except through it. ether in the cause of disposal of the methods of sale , for example they simply delete the files or private data from it by returning it to factory or so-called colloquial mince thinking of then can prevent the existence of such files without being able to be restored and detected by anyone else , but what many do not know is that file and data can be restored .They exist on the computer using specific software is on the market , which warns him experts and specialties in computer technology , it causes problems in society . The problem lies not only in the presence of devices restoring deleted files, instead experts could not create devices that would permanently delete those files. That is, a one who owns a personal computer and wants to sell it or give it to anyone else or even bring it back to work after they have finished using it in their professional field. it will not be able to perform the final deletion of personal files. Naturally, human souls can’t be

guaranteed. Someone may tamper with memory of computer that was previously used. To extract old files belonging to the first user and contain personal photos or videos or confidential information should not be owned by the only the owner. The issue may increase its severity to the extent that major problems occur in society because of this what calls people to be-ware of their personal computers and mobile phone, and do not store images or very special files on them to avoid problems in the future. The bad side of them is that may be used for bad purposes that are immoral like extortion when someone bought a phone for a girl in advance and had done work to restore files from this phone it is important files these pictures. Where he blackmailed some institutions when found important data for these institution. For example, this may happen and may be repeated very significantly, especially after the advancement of technology and the emergence of new techniques for recovery and easy to use, which has become accessible to many people. Therefore, their use should be restricted to the computer authorities only and will not become a double – edged sword in the hands of person.

5.2 The Look of our Yemeni Society on File Carving

To know this aspect of the study must take a sample and slice of people through which we may be able to access a small part of the knowledge of the (Yemeni) society with this technology in some sample, we wanted a sample and rationing chip and that interested information and technical in addition to the sample of the people of the law .and random questions were clear and straight for word and get to know the results:

The question (1): what is the file carving?

The result (1): most probably did not know what the question originally not to mention the meaning, but someone linked to files.

The question (2): we have repeated the question from in Arabic what is the technique of file carving?

The result (2): there is a responsiveness and knowledge that is a computer technology when some only and without what this technique or the area used. Then we clarify this technique and its definition how to use it and how some of this sample and ask new and different question.

The question (3): what do you think of this technique and use in Yemeni?

The result (3) mostly expressed acceptance and clarification of the benefit of them, but some after knowing limited expressed concern about the wrong use.

The question (4): what do we need in order to successfully use this technique?

The result (4): here the result was good where some commented on the identification of the user and give the laws and powers and the work of its rule during the establishment of its own section.

6 Legal Framework for File Carving

File carving is used in computer forensic (digital forensic) read-out data from drive or other storage devices in legal investigation and criminal cases related to pornography and fraud also to retrieve evidence and data and perform digital forensic investigation. This method is specially used by digital forensic experts to recover evidence the law enforcement agent often is able to retrieve more images from suspected hard drives.

6.1 The Outlook for File Carving as tool

They are used legally and in criminal investigation as a tool such as criminal investigations related to financial fraud and issues of bullying, harassment extortion and porn channels. When these files are deleted by those responsible for hiding them are recovered and returned to the ground and used as a tool against its members and punished. The military side has been used to spy get information to the enemy. When a hard disk or enemy storage device is obtained it may contain important and confidential information or they may be deleted and recovered and get their information about the other party.

6.2 Deficiencies of Digital Evidence

Digital Evidence is inherently disturbing and controversial where they are digital data may be fragile weak and interchangeable and renewable and even damage. and as important as the use and increase in it that this thing is still flaws and mistake and still needs a lot of work to reach the efficiency, accuracy and durability of the evidence.

6.3 Criteria for Accepting Digital Evidence

There are two more famous criteria in this area:

1- standard "Daubert"[9] it is one of the criteria that emerged in 1993 and has several factors:

- The testing: the procedure to check-up the safety of evidence.
- Review the evidence fairly: the procedure that subjects evidence to tough review.
- Error rate: the procedure that defines defects of evidence.
- Setting standards and protocols: the procedure that defines method of using and accepting evidence.
- Admission from specialists: the acceptance from experts of investigation.

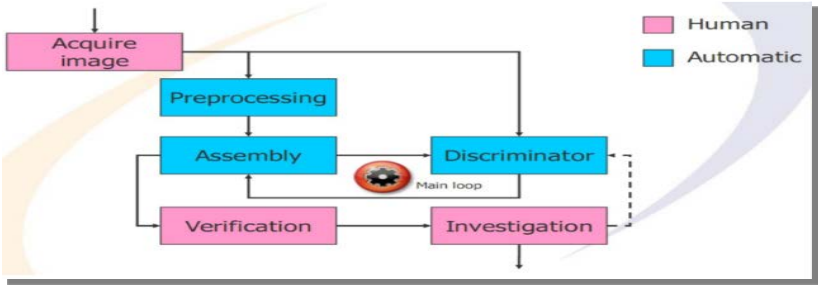


Fig.2.Integration Phases of Digital Evidence

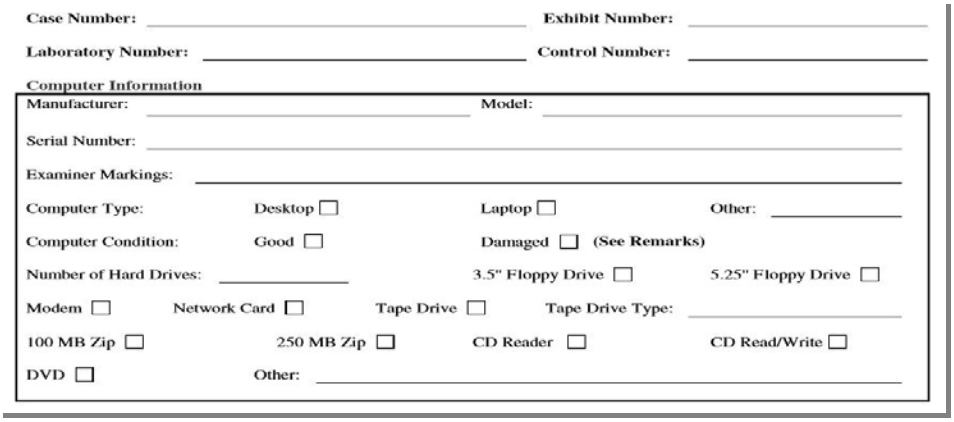
2- Standard "Kumho" it is an extension of the standard "Daubert", but the application includes a special technical and scientific knowledge of specialists was in the year 1999.

6.4 Ways to Preserve Digital Evidence

There are several different ways to preserve the digital directory:

1- create a special body to deal with digital evidence, whether physical such as media or hard copies (**Fig. 3**) or digital or software manual such as webpages for interested website this body shall be composed of specialists (technicians and legal) [10].

2- Developing a preservation mechanism commensurate with each guide in terms of its type accuracy or even its strength.


Case Number: _____ Exhibit Number: _____

Laboratory Number: _____ Control Number: _____

Computer Information

Manufacturer: _____ Model: _____

Serial Number: _____

Examiner Markings: _____

Computer Type: Desktop ☐ Laptop ☐ Other: _____

Computer Condition: Good ☐ Damaged ☐ (See Remarks)

Number of Hard Drives: _____ 3.5" Floppy Drive ☐ 5.25" Floppy Drive ☐

Modem ☐ Network Card ☐ Tape Drive ☐ Tape Drive Type: _____

100 MB Zip ☐ 250 MB Zip ☐ CD Reader ☐ CD Read/Write ☐

DVD ☐ Other: _____

Fig. 3. Computer Evidence Worksheet

7 Enable File Carving and Digital Evidence in Yemen Legally

In order to make any change it is self – evident to know the reasons that lead to this change and the consequences of the change. It is this principle that you should know the reasons for introducing digital evidence into the legal field Yemen, like other third world countries uses globalization.

As a result of cybercrime, it exists in Yemen as a result of the spread of electronic commerce and the connection to the internet for the sake of remittances or, for example a Yemen citizen concealed the corruption files on his computer. Does he close the case for lack of concrete evidence?

Enabling Requirements:

- The establishment of joint department in the first stage between the faculty of law and technical colleges to graduated, learn with cybercrime and electronic evidence at a later stage the possibility of its independence.
- Issue the necessary laws so that the cases are subject to the use of electronic evidence or the work of exploration notes for files.
- Work to establish organization and institution of the state in order to reduce the indiscriminate use of these technologies.
- Provide software and hardware needed and work with the sophistication of the technique.
- Identification and definition of nomenclature such as electronic harm, electronic directory electronic investigations and other concepts.

8 Technical Framework for File Carving

In this part we will discuss the technique and the basics of dealing with files to be recovered and analyzed in relation to the file structure or the operating system or stored media will be touched on several different methods. [11]

These techniques are divided into two groups:

1- check the formula " format validation"

2- rebuild the file " file reconstruction"

First “format validation” it is one of the easiest ways where a block of data matches the formula "format". This method is considered useless with large data volume up to “terabyte " manual verification can take up to several months, but it may be fusible to use automated verification. This automated method depends on the file format of the ID structure, but the disadvantage is if there is a file structure that is free and unspecified.

We don’t use a verification mechanism such as a file size, for example in the case it is impossible to use automatic verification.

There are several methodologies for automated verification:

i- Match the Magic Number:

They use digital file formats (hex-decimal, binary) and rely on (header-footer) [12] to file any beginning and end formula, regards evidence (scalpel) developed from "foremost” one of the tools used for this methodology (**TABLE 2**.).

Table 2. Extension and Magic Number of File

File Type	Extension	Magic Number
JPEG	.jpg	FFD8
GIF	.gif	47 49 46 38 [GIF89]
PNG	.png	89 50 4E 47 .PNG
MIDI	.mid	4D 54 68 64 [MThd]
Icon	.ico	00 00 01 00
AVI	.avi	52 49 46 46 [RIFF]
FLASH shockwave	.swf	46 57 53 [FWS]
FLASH video	.flv	46 4C 56 [FLV]
Mpeg4	.mp4	00 00 00 18 66 74 79 70 6D 70 34 32 [… ftypMP42]

i- Dependency Data Resolving:

Using this methodology allows the file format to define its own format, where file size occur “of size file " this will identify the file.

ii- Verification Internet Checking:

It is the opposite of the previous two approaches where the content of the file is not taken into account, but the circular code is taken.

iii- Output Algorithm Analysis:

In this methodology, the compressed or encrypted data stored in the media is examined, for example picture "JPEG" use its "Huffman Code" to compress data it is easy to identify this data through "Sequence Bits".

Second rebuild the file "file reconstruction" it is a method that employ inference to know the retail sample "fragment" or file properties without this metadata. it is difficult to locate or know the file in the media. it has several methodologies: [13]

- Fragment Reordering: divided into Bi-fragment Gap carving (BGC) and Advanced carving (AC) whereas (BGC) it is restricting the search and its area exploration and drilling only in fragmented files which may be are added to the file.
- Classifying of Fragments: A stand-by tactic to decreasing the collections of fragments to consider when rebuild files is "Classifying". Singular fragments are count and, based on their contents, either contain or eliminate from further rebuild. This method take-in well with fragment reordering or any other technique of file carving, as it merely decreasing the fragments to consider.

9 Smart Carving

The concept of file carving that is not bounded to recovering files with dual fragments. renowned the issues deep-rooted with their incipient research into retrieval of fragmented partitions at random using precomputed weighing and provided the scope and conclusion for a carver that takes into account the typical fragmentation behavior of a disk and one that scales to huge disks. Approach of file carving that can expatiate fragmented and un-fragmented on media-storage and any kind of file systems. The first stage is "preprocessing", where clusters of file are de-compressed or de-crypted.The next stage is "collation", where clusters of data are categorized as pertinence to a type of file. The final step is "reassembly", where clusters collective in the collation process are joined jointly to rebuild files [14].

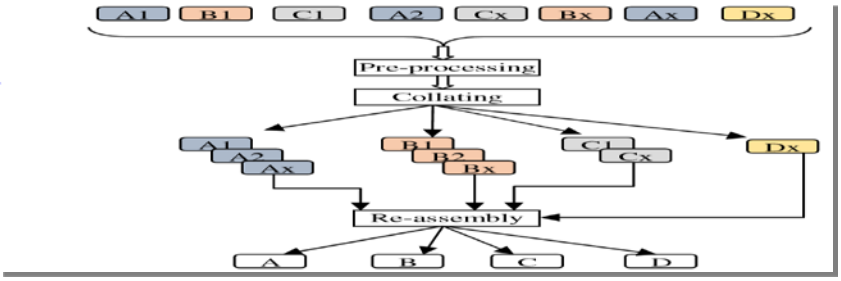


Figure 4. Structure of Smart Carving

10 Logistic Framework for File Carving

When you take the technical side of this technology, you should focus on what is returned to the output of this technology. The user need auxiliary tools such as some science knowledge and algorithms, but we must focus on what the specialist needs in this technique.

Strength Points needed by the specialist:

1- logical thinking

2- dealing software and databases

3- operating system and file system

4- Internal structure of processors and storage media for devices.

11 Conclusion

The methods for recovering deleted files play an important role in the field of digital forensics. This research paper seeks to clarify from several aspects of society such as Yemeni Society as model, and one of the methods of recovery that play this role is File Carving , which the study intends to clarify the

mechanism of its work and its techniques and tools used where it was done making a comparison of tools through previous experiences and research , most notably (DFRWS) Digital Forensics Research Workshop challenge in 2006 to clarify the strength and weakness of the most important tools through to main factors (Recall and Precision) . Then this study set out a number of frameworks for this technology in order to consolidation and a solid foundation for the work of this technology, including the societal framework, which we have made clear that the community does not know according to a mini field study and clarify the problems that may arise through the indiscriminate use of File Carving technology, then we developed an idea for this technology to be with in a legal framework by presenting a set of requirements, most notably the establishment of a common section between a group of legal and technical persons, then move this study to the most important framework which is the technical framework that explains the techniques used in File Carving, which is the most important method for it is the Header-Footer and with an explanation of development of technology via Smart Carving .

REFERENCES

[1]. R. Poisel, S. Tjoa and P. Tavolato, *"Advanced File Carving Approaches for Multimedia Files"* Insitute of IT Security Research St. Poelten University of Applied-Sciences, St.Poelten, Austria {rainer.poisel,simon.tjoa,paul.tavolato}@fhstp.ac.at (presented at the 6th International Conference on Availability, Reliability and Security (ARES'11), Vienna, Austria, August 2011).

[2]. C. Walter, "*Kryder's law"* (2005, July) http://www.scientificamerican.com/article.cfm?id=kryders-law. Scientific American. [Online; Status: October 15th 2011].Teng, T., M. Lefley, and D. Claremont, Progress towards automated diabetic ocular screening: A review of image analysis and intelligent systems for diabetic retinopathy. Medical and Biological Engineering and Computing, 2002. 40(1): p. 2-13.

[3]. German Federal Ministry of the Interior, *"Polizeiliche Kriminalstatistik 2010,"* http://www.bmi.bund.de/ SharedDocs/Downloads/DE/Broschueren/2011/PKS2010.html, May 2010, [Online; Status: October 6th 2011].

[4]. R. Poisel and S. Tjoa, *"Roadmap to approaches for carving of fragmented multimedia files,"* in Proc. of the 4th International Workshop on Digital Forensics (WSDF'11), Vienna, Austria. IEEE, August 2011, pp. 752–757.Haddouche, A., et al., *Detection of the foveal avascular zone on retinal angiograms using Markov random fields.* Digital Signal Processing. 20(1): p. 149-154.

[5]. McAfee Avert Labs, *"French Authorities Talk Up Digital Investigations,"* http://process-info.org/news/ security-news-blogs/archive/2010/10/french-authorities-talk-up-digital-investigations, October 2010, [Online; Status: March 29th 2011].

[6]. C. Veenman, *"Statistical Disk Cluster Classification for File Carving"* Intelligent System Lab, Computer Science Institute, University of Amsterdam, Amsterdam,Digital Technology and Biometrics Department, Netherlands Forensic Institute, The Hagu, Conference Paper · August 2007.

[7]. A. Pal and N. D. Memon, *"The evolution of file carving,"* IEEE Signal Processing Magazine, vol. 26, no. 2, pp. 59–71, March 2009.

[8].] S. Kloet, *"Master's Thesis Measuring and Improving the Quality of File Carving Methods"* Supervisor: Prof. Dr. W.J. Fokkink, Eindhoven University of Technology Department of Mathematics and Computer ScienceAlmere, October 29, 2007.

[9]. Daniel B. Garrie and J. David Morrissy, *"Digital Forensic Evidence in the Courtroom: Understanding Content and Quality"* 12 Nw.J. Tech. & Intell.Prop.121,https://scholarlycommons.law.northwestern.edu/njtip/vol12/iss2/5N Northwestern Journal of Technology and Intellectual Property, Spring 2014.

[10]. J. Ashcroft, Deborah J. Daniels and Sarah V. Hart, *"Forensic Examination of Digital Evidence: A Guide for Law Enforcement"* U.S. Department of Justice Office of Justice Programs 810 Seventh Street N.W. Washington, DC 20531, http://www.ojp.usdoj.gov/nij

[11]. R. Pahade, B. Singh and U. Singh, *"A SURVEY ON MULTIMEDIA FILE CARVING".* Department of Computer Science & Engineering, Defense Institute of Advanced Technology (DIAT), Pune, India Xiaohong, G., et al. *A method of vessel tracking for vessel diameter measurement on retinal images*. in *Image Processing, 2001. Proceedings. 2001 International Conference on*.

[12]. N. Škrbina and T. Stojanovski, *"USING PARALLEL PROCESSING FOR FILE CARVING"* European UniversitySkopje, Republic of Macedonia,2013.

[13]. L. Pereira and C. Romano, *"File carving in practice"* (Master Computing Engineering Dissertation supervised by Prof. Cunha) October2015.Pinz, A., et al., *Mapping the human retina.* Medical Imaging, IEEE Transactions on, 1998. 17(4): p. 606-619.

[14]. A. Pal and N. *Memon, "The Evolution of File Carving [The benefits and problems of forensics recovery]"* IEEE SIGNAL PROCESSING MAGAZINE MARCH 2009.