Recovering Files using File Carving

Ziad Saif Alrobieh
[1]*Department of Communication & Computer Engineering, Alsaeed Faculty for Engineering & Information Technology, Taiz University, Taiz, Republic of Yemen;*
ziadrh@yahoo.com

ABSTRACT

Recently, file carving has become one of the most important topics of digital forensic to recover files from their fragments without the existence of the file where the restore mechanism does not depend on the file system or metadata. Smart Carving achieves higher efficiency and accuracy when compared to many rudimentary file recovery systems.
Keywords: File Carving, Digital Forensics, File Systems, Smart Carving.

1. Introduction

There are a large number of studies in the field of file carving, but most of these studies do not address the mechanism of this technique extensively. This technique does not depend on the knowledge of the metadata, but rather through a group of methodologies such as the header and the footer methodology, and such methodologies encountered problems such as fragmentation. Therefore, when trying to understand this technology, the file structure and the file systems must be understood, as there are many file systems, and thus the advantage of file carving was that it does not depend on the file system, and this is what this research seeks to clarify. The remainder of this paper is organized as follows: Section 2 explains the digital forensics, Section 3 discusses the file structure, Section 4 describes the file systems, Section 5 addresses the file carving technology, and Section 6 summarizes the conclusion.

2. Digital Forensics

2.1 History and Concept:

Digital forensic evidence began nearly forty years ago in the seventies of the twentieth century. Due to the increase of cybercrimes in the eighties, this made investigation community look to digital evidence as a source of criminal evidence. Computer evidence began in 1984 when the US Federal Bureau of Investigation (FBI) created a program (Magnetic Media Program), which has now become known as CART. In the same year, the Computer Analysis and Response Team (CART) provided assistance to the FBI in the search and confiscation of evidence. In the nineties of the last century, in 1997 in particular, and with the emergence of the Internet and thus the emergence of crimes via the Internet, the Scientific Working Group on Digital Evidence (SWGDE) [15] has established standards for the digital evidence. In the early of the year 2000, there was a widespread of electronic crimes as a result of the

proliferation of the computer and mobile devices [1]. Electronic investigations are used to uncover facts about some crimes, such as issues of intellectual property rights, cases of child molestation, financial fraud, and other crimes that use some electronic tools. *Digital Forensics Science (DFS)*, which is one of the branches of forensic science, includes the retrieval and analysis of data that are contained in electronic devices or have a link to computer crime. *Computer or Cyber Forensics (CF) evidence* includes computer investigations and analytical techniques to solve a problem, the crime, by obtaining evidence to support case procedures. Electronic criminal investigations have three stages [2]:

1- Evidence acquisition: It is obtaining evidence while preserving it from any alteration or distortion from the moment it is acquired until the evidence is analyzed.
2- Evidence analysis process: the evidence is examined, matched, and conclusions drawn. The examination process may be separate from analysis sometimes.
3- Presentation of the evidence: it is the presentation of the evidence in full, without deficiency, and the presentation of the results of the analysis to those concerned with the results of the guide.

Fig. 1. Stages of electronic criminal investigations

2.2 Digital Evidence:

Digital evidence is that which is produced using a digital device or stored in it [3], for example:

- Computers: including desktop, laptop, and servers.
- Peripheral devices: such as a webcam, printer, scanner, and others.
- Network devices: such as switches, routers, and others.
- Storage media: such as internal and external hard drives and removable media such as Flash drives, USB stick, and any devices through this port, as well as CD/DVD, Floppy disks, memory cards such as micro SD card and tapes.
- Portable devices: such as digital camera, video camera, MP3 player, voice recorder, digital calculator, smartphones, GPS, and others.

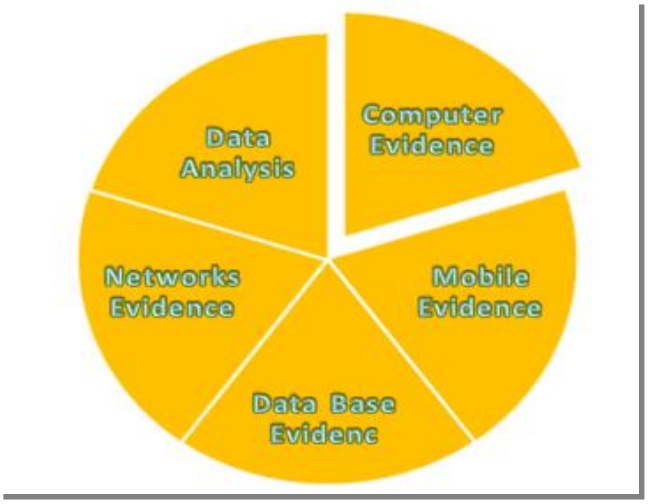

Fig. 2. Types of digital evidence

2.3 Criteria for Accepting Digital Evidence:

There is a criterion for accepting evidence through several steps, namely Daubert Hearing [4], which is one of the criteria that appeared in 1993 and has several factors, including:

1) The testing mechanism and its results.
2) Responsible and impartial review of evidence.
3) Establish controls and standards and protocols.
4) Special acceptance of the evidence by specialists.
5) Error rate detection.

The Kumho standard is an extension of the Daubert standard applied in 1999, but with including special technical and scientific knowledge for specialists.

2.4 Computer Role in Investigations:

The computer in any crime may be a [5]:

1- Target: for example, to be exposed to electronic attacks and hacking campaigns, and harm is done to the computer and its owner, in other words, the victim.
2- Tool of the crime: that is, it is used as means to commit the crime.
3- Assistant in collecting and storing evidence: it is when used to help specialists in criminal investigations to find out some evidence related to the crime.

In the context of this research, we focus on the assistance provided by computers through a set of specialized tools for file carving.

3. File Structure

The file structure is divided into three types depending on the operating system when performing file processing operations such as adding, deleting, modifying, embedding, and other operations [6]:

1- Contiguous files: These are files whose data are stored in the form of blocks and are stored sequentially in storage media.
2- Fragmented files: which form fragments (chunks) of the stored file inconsistently, and fragmentation may be divided according to the header and footer distribution between the blocks to:

- Linear fragmentation: the files are divided into parts, but these parts, despite their fragmentation, remain on a specific sequence, i.e. the first part of the file contains the header, then comes a file that breaks it up, and then the second part contains the footer.
- Non-linear fragmentation: it is the opposite of linear fragmentation where fragmentation occurs but in a non-sequential manner, i.e. the first part of the file contains the footer, then a file comes that separates it from the second part that contains the header.

There is also what is known as bi-fragmented, which is a fragmentation that occurs to the file but divides the file into two parts only [7].

3- Embedded: they are files that contain other files that have been stored in these files, for example saving files such as audio and video files inside zip files.

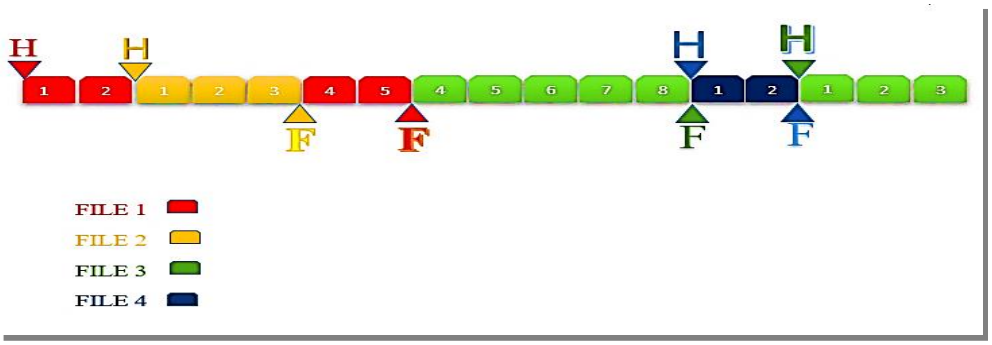

Fig. 3. Contiguous, fragmented and embedded file structure

4. File Systems

File systems differ and vary according to the different operating systems, and any file system has several functions [8], the most important of which are:

- Identify and customize files.
- Read files.
- Maintain and preserving metadata associated with files.

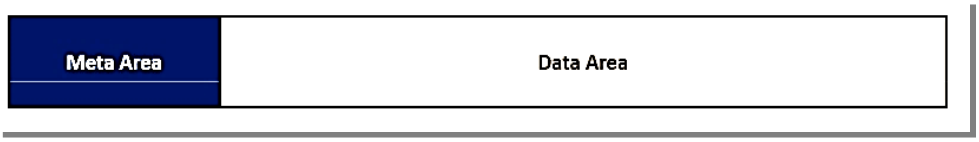

Fig. 4. General file system architecture

Table 1. Some types of file systems with operating systems.

Windows	Linux	Apple (Other)
FAT12, FAT16, FAT32, FAT64, NTFS, TFAT, and exFAT.	Ext2, Ext3, Ext4, NExt3, XFS, JFS, btrfs, UBIFS, JFFS2 and YAFFS.	(HFS) Hierarchical File System, UFS, HFS+, HPFS, APFS, ISO 9660, Files-11, Veritas File System, VMFS, ZFS, ReiserFS and UDF.

4.1 FAT file system:

File Allocation Tables (FAT) is a file system supported by Microsoft. It has a structure divided into a group of parts and each part has a specific role. These parts are Data Region, Root Directory, and Boot Record.

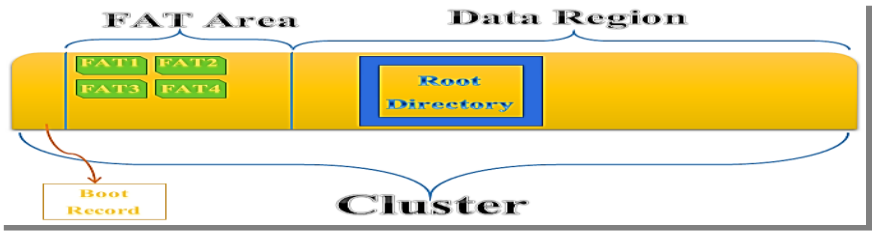

Fig. 5. FAT cluster

4.1.1 Structure of FAT:

In FAT file system, the folder is moved and the folder name is saved in the Root Directory provided that this folder is the root, and if it is not the root, the process of analyzing each Root Directory record begins to obtain the folder.

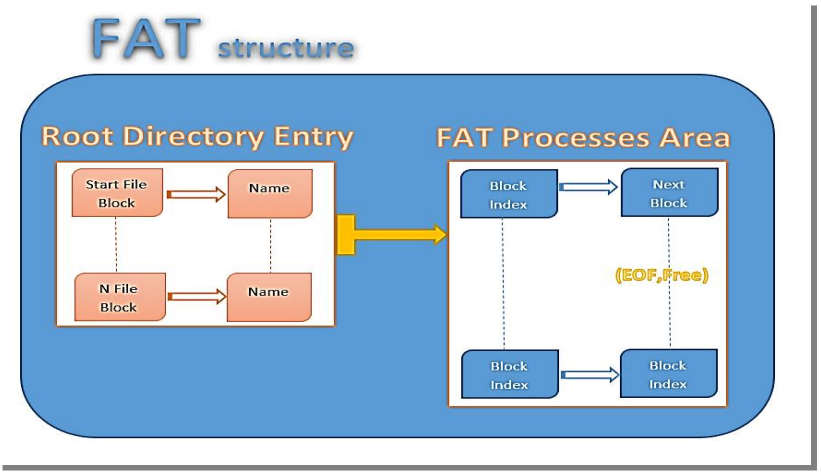

Fig. 6. FAT structure

4.1.2 FAT Deletion Mechanism:

When the file is deleted, the cluster becomes a pointer, and thus it is not specified (Unallocated) in the File Allocation Table and the change occurs in Data Region as the first byte of the file name changes to (a5) in the hexadecimal system. When deleting, the Next Block field becomes empty of the value it is indicated on. No change will occur in Root Directory Entry as the deletion, modification, and addition of files deal with the FAT Area.

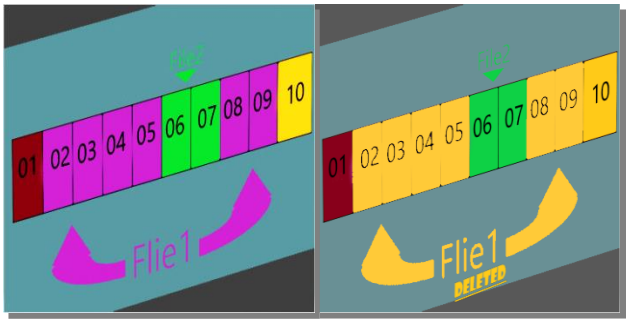

Fig. 7. File deletion

4.2 Extended File System:

Extended File System (Ext), which is a file system supported by Linux distributions, has a structure divided into a group of parts where each part has a specific role. These parts are Boot Block, a group of blocks (Super Block, I-node, Data blocks(, and Directory blocks.

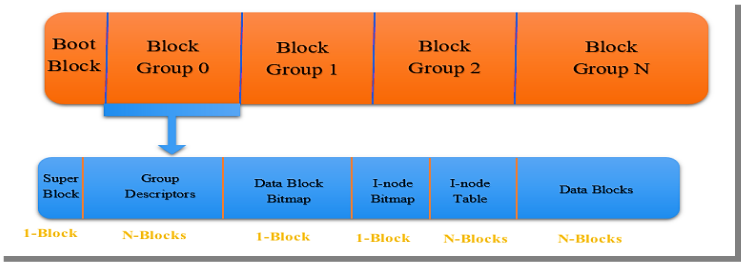

Fig. 8. Partition s of Ext

4.2.1 Structure of Ext:

Ext consists of (Directory Entry) folder entries, which include each of the following data:

- Name: it is a data container that contains the folder name and the names of the files that were saved.
- I-node: this is an index that contains instructions and a set of data, such as access time for each file, its size, and some other metadata. It also contains a group of blocks, which in turn are linked by pointers for all components of each file. There are also indirect blocks that are linked by pointers to the addresses of the files, which are stored in address blocks. Of course, this system contains file data or (files content).

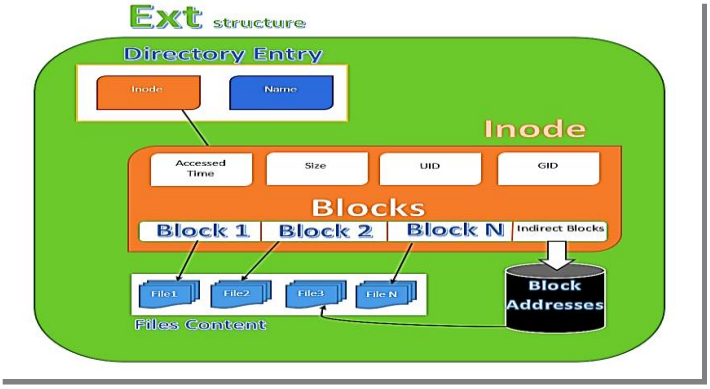

Fig. 9. Ext structure

4.2.2 Ext Deletion Mechanism:

Directory Entry is a feature on the basis that it has been deleted but not permanently. Upon deleting, all pointers of Blocks are deleted, as well as contents of Block Addresses and size. The value of I-node returns to I-node List and the data fields become Free Blocks.

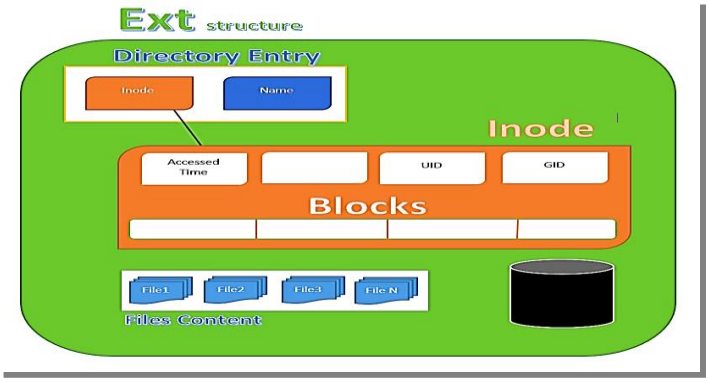

Fig. 10. Ext structure after deletion processes

4.3 NTFS File System:

New Technology File System (NTFS) is a file system supported by Microsoft. The information is stored in this system in the form of a B-Tree. This system consists of several parts with specific roles, and these parts are Master File Table (MFT) area, Boot Sector, File Parts.

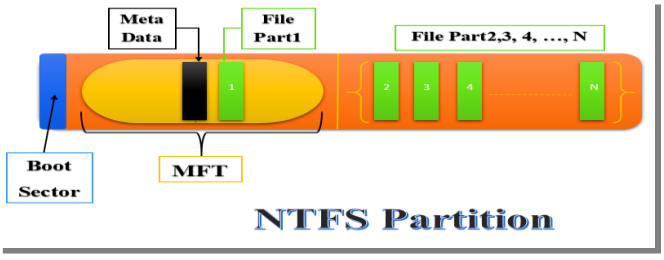

Fig. 11. NTFS Partition

5. File Carving

File Carving Technology is an innovative technology that works to recover files that are considered to have been deleted and does so through the structure of the file structure and the content of the file. It thus eliminates the dependency on the file system and recovers files from unallocated space. This technique is used as a powerful electronic criminal investigation technique [9].

File carving is one of the most powerful techniques used to revive files and data because it does not depend on the file system and can recover files even if the meta-data of the file system is completely destroyed. This metadata is presented in the MFT for NTFS system, in I-node for Ext system, and in a Root Directory for FAT system [19].

5.1 Algorithms of File Carving:

The algorithms of the file carving are the Brute Force, Boyer-Moore, Knuth-Morris-Pratt, Raita, Berry-Ravindran, Karp-Rabin, Commeritz-Walter, Wu-Manber, and Aho-Corasick.

5.2 Methodologies of File Carving:

There are several methods and mechanisms used by the file carving tools when excavating, and these mechanisms include: Header Based method, File Structure, Fragment-Recovery Carving, Bi-fragment Gap Carving, Block-Based Carving, Repacking Carving, and File Validation.

5.3 Smart Carving:

Smart Carving technique is one of the modern methods in this field which also helped in solving the problem of fragmented files through three main processes that take place before restoration, namely pre-processing, collecting, and then recombining the file (Reassembly) [10][21].

1- Pre-processing: it is the first process where it is concerned with working on decrypting files if they are encrypted and decompressing compressed files in order to facilitate dealing with files directly. In this process, allocated and unallocated clusters are determined, and upon completion of these operations, the file is ready to move to the next process.
2- Collecting and Comparison: similar clusters are classified according to the File Signature factor as each file, depending on its format, has this unique factor. Here, string matching algorithms can be relied upon, or entropy can be taken as it helps to reveal the probability of the file type.
3- Reassembly: the parts are arranged correctly then combined together, such as the original file, and then the output of this assembly is checked whether it is correct or not [11].

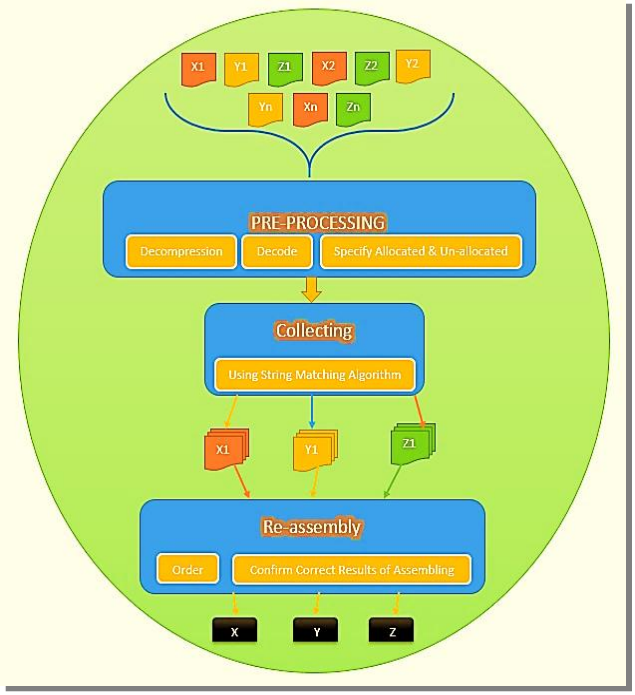

Fig. 12. Smart carving

5.4 File Carving Analyses:

File carving analyses are divided into are three types based on the type of storage media:

1- Hard-Drivers Analysis: which depend on machine learning and entropy of file blocks, header and footer, and target all file formats such as (ZIP, PDF) or file system such as (NTFS).
2- Volatile Memories Analysis: which search for special strings or parameters (Signatures) or interpretation of the structure of the inner core (Kernel), and work to recover from finished and running processes as well as from open ports and hidden data.
3- Non-volatile Memories Analysis: they are used with embedded devices and uses a simple technique to interpret the data recovered from the embedded devices called the dump.
   - Embedded devices have self-data and are considered as low-cost devices, such as credit cards, electronic passports, etc.

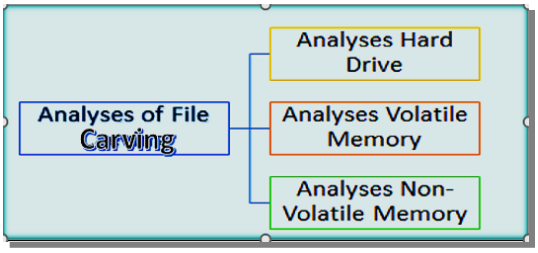

Fig. 13. File carving analyses

5.5 Disadvantages of File Carving:

1- File Carving has a high time-complexity, which is why it cannot perform quick verification and retrieval. The reason is due to the collection of different parts and paths.
2- It depends on the quality of the tool used, as each tool has different advantages and disadvantages.
3- Waste of space by storing data that is recovered and may not be useful or may not be the required data.
4- Some parts and files are not readable or corrupted.
5- In the case of file fragmentation, each cluster is disconnected and non-sequential. This is why the carving process fails. It is considered a defect of file carving, but not for all carving methods or tools, as this problem has been addressed in several ways.

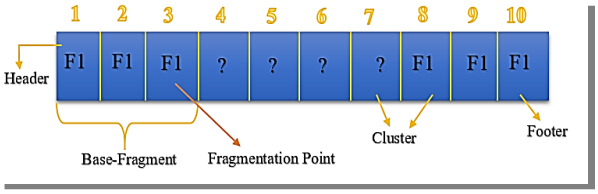

Fig. 14. File fragmentation

5.6 File Carving Operations:

They are operations that evidence or file to be restored goes through and explain the integration between the human element and the automatic methods [12].

- Preprocessing: it is extracting information about the file by specifying the beginning and the end of the file and the length of the file size, as well as specifying all the sectors in which the parts of the file are located.
- Assembly: by generating a copy of the file temporarily, taking all sectors and arranging them in sequential order.
- Discriminator: ensure the correctness of the results and avoid mistakes.

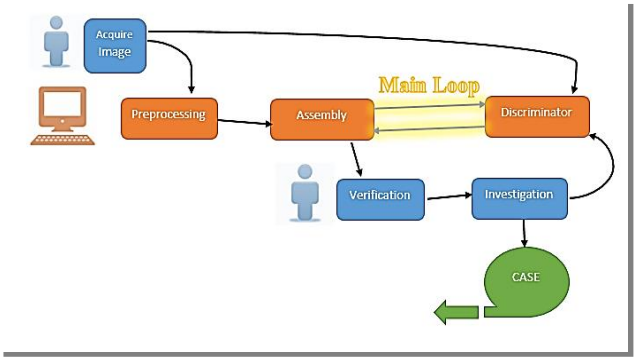

Fig. 15. File carving operations

5.7 File Carving Tools:

File carving tools have the best performance among the electronic criminal investigation tools for several factors, including [13]:

1- The percentage of recovered files.

2- Accuracy and reliability of the outputs of these tools.
3- The speed of processing carried out by the carving tools.

5.7.1 Foremost:

Foremost is a tool that works in the form of Commands, uses a number of Parameters and user-inputs. It works on recovering files through Internal Data Structures & Headers-Footers by dealing with storage media directly or images of storage media, such as Encase, dd, Safeback, and others [14].

Table 2. File Types Supported by Foremost

Support Mechanism	File Types
Header-Footer	ART, asf, chm, dat, dbx, fws, idx, java, cpp, lnk, mail, mbx, MP3, mpg, ost, pgd, pgp, ppt, pst, ra, rdp, rpm, tif, txt, wma, wmv, wpc, COOKIE and XLS Files.
File Structure	BMB, HTML, DOC, PDF, GIF, JPG, PNG, MOV, AVI, WAV ,ZIP Files and RAR

5.7.2 Scalpel:

It is an improvement and development of Foremost in order to increase performance and reduce the required storage space. It shares some code scripts, but it is considered faster. It works as follows [17]:

1- First, it creates a database for each header with all the files, and then searches for possible footers. When the header is found, the maximum length or file size must be specified.
2- Then, we do a match between all the headers and footers to create the file. The creation process is through the queues for each chunk of the file.
3- Finally, all files are extracted through the queues for all the fragments.

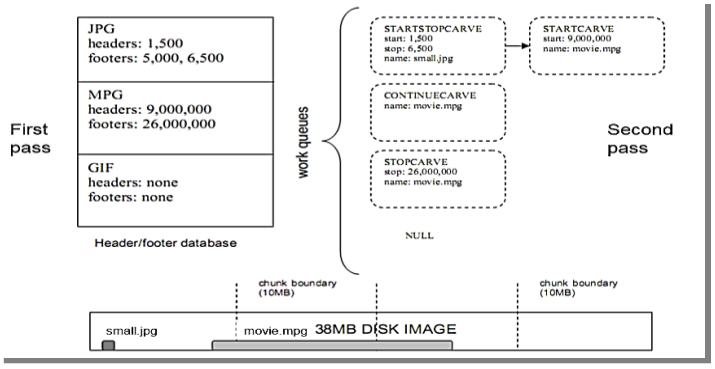

Fig. 16. Mechanism of Scalpel tool

5.7.3 Autopsy:

It is a multi-use investigation tool that is built on the Sleuth Kit platform. It works on retrieval and sorting of files, and displaying files in the form of windows or via commands with a statistic on the total number of files, the number of files recovered, and the number of all files of each type that have been recovered. This tool is easy to use due to the clarity of its interfaces.

5.7.4 Photorec:

Regardless of the tool's name, this tool can recover many types of files such as audio, text, video files, and archived files in addition to image files. It supports approximately 82 types of files and their formats and works to recover files through a combination of two techniques of file carving, File Structure-Based & Header-Footer [20].

5.8 File Types Supported by File Carving Tools:

As each tool supports specific types that can be restored, there are also tools that have the ability to adjust the type of files that can be recovered. Each data type has a special header and is distinct from the rest of the other types and may have footer [16].

Table 3. File types supported by the file mining tool

Documents	Web Files	Archived Files	Videos	Audio	Pictures	Other Types
PDF ,DOC, PPT , TXT & XLS	HTML ,CHAT , E-MAIL & SQLITE	RAR , GZ , TAR, 7Z & ZIP	MP4, AVI , MOV, WMV , 3GP & OGV	WAV , WAM, AU & MP3	GIF,PNG, TIF, BMP&JPG	LOGS , EXEC & PROGRAMS FILES

5.9 How to Carve Files:

The file carving technique works by knowing the file type first for the file to be distinguished, for example a compressed (ZIP) file with a header value (504B0304140000000800), then determining the location of the header and footer, and finally retrieving all data between the header and the footer. If the footer is not present, the Maximum Length is taken [18].

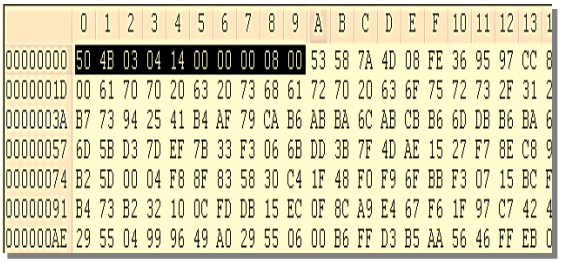

Fig. 17. ZIP file's header

Carving techniques can also be classified into two main categories of basic carving techniques and advanced carving techniques. The basic carving techniques use the file structure in recovering files and the most common basic techniques are signature-based, content-based, and structure-based carving. On the other hand, the advanced carving techniques utilize the individual files contents beside the file structure, such as graph theoretic-based carving and weightage Technique [22].

6. Conclusion

File carving technology has helped many governments and organizations that rely on electronic investigations to solve cases. Its tools have become the most used tools due to their high-accuracy results in retrieving evidence, as well as its continuous development of new mechanisms to solve the problems the technology addresses. When the mechanisms that do not depend on metadata have been adopted, they have become more flexible and effective than traditional restoration. Since each operating system has its own metadata and a special file system, which leads to the inability to recover if the system itself is deleted, several algorithms have been improved and developed, such as String Matching Algorithm, which in turn helped innovate many new carving methodologies. With all these changes, the need for evaluation mechanisms for the technology and its tools emerged. File carving technology adopted a methodology based on three factors: recall, precision, and f-score in addition to additional factors such as the time of execution of the restoration process or the speed of the tool in execution, etc. In this research, we added a new factor called images of the storage medium. The importance of the images in this process is that they keep a copy of the evidence and work on it in order not to harm the original evidence.

REFERENCES

[1]. B. V. Prasanthi, “Cyber Forensic Tools: A Review”, International Journal of Engineering Trends and Technology (IJETT) – Volume-41, ISSN: 2231-5381, Number-5 - November 2016, https://www.ijettjournal.org.

[2]. T. Gougeon, M. Barbier, P. Lacharme, G. Avoine, C. Rosenberger, “Memory Carving in Embedded Devices: Separate the Wheat from the Chaff”, 14thInternationalConference on Applied Cryptography and Network Security (ACNS 2016), Jun 2016, Guilford, United Kingdom. 10.1007/978-3-319-39555-5_32. hal-01338109v2.

[3]. Y. LeClaire, “The Forensic Process Examined: Creating cases for classroom use”, Lewis University, MSIS 68-595.

[4]. B. Carrier, “Open Source Digital Forensics Tools the Legal Argument”, carrier@cerias.purdue.edu .

[5]. A. Sivaprasad, Prof. S. Jangale, "A Complete Study on Tools & Techniques for Digital Forensic Analysis", Information Technology V.E.S.I.T Mumbai, India, abi.lecturer@gmail.com, smitajangale@yahoo.com.

[6]. T. Laurenson, "Performance Analysis of File Carving Tools", 28th Security and Privacy Protection in Information Processing Systems (SEC), Jul 2013, Auckland, New Zealand, pp.419-433,10.1007/978-3-642-39218-4_31, hal-0146384.

[7]. S. Garfinkel, "Carving Contiguous and Fragmented Files with Object Validation", The Digital Forensic Research Conference DFRWS 2007, USA Pittsburgh, PA (Aug 13th - 15th), http:/dfrws.org .

[8]. K. Hulin, "Digital Forensics III - File Carving", Department of Computer Science The University of Texas at Dallas, September 23rd, 2011.

[9]. A. Pal and N. Memon, "The Evolution of File Carving [The benefits and problems of forensics recovery]", IEEE SIGNAL PROCESSING MAGAZINE, 1053-5888/09/$25.00©2009IEEE, MARCH 2009.

[10]. A. Nur Elmi Abdi, "MP4-Karver: CARVING OF CORRUPTED MP4 VIDEOS USING ASMD REPAIRING TECHNIQUE", Faculty of Computer Science and Information Technology Universiti Tun Hussein Onn Malaysia, July 2016.

[11]. H. Lee, H. WooLee, "Block based Smart Carving System for Forgery Analysis and Fragmented File Identification", Journal of Internet Computing and Services(JICS) 2020. June.: 21(3): 93-102, http://dx.doi.org/10.7472/jksii.2020.21.3.93.

[12]. G. D. Cantrell ,J. R. Through, "Teaching Data Carving Using The Real World Problem of Text Message Extraction From Unstructured Mobile Device Data vice Data Dumps", Journal of Digital Forensics, Security and Law: Vol. 14 : No. 4 , Article 4 ,DOI: https://doi.org/10.15394/jdfsl.2019.1603 , Available at: https://commons.erau.edu/jdfsl/vol14/iss4/4 .

[13]. Nurhayati, N. Fikri, "The Analysis of File Carving Process Using Photorec and Foremost", Department of Informatics Engineering, Faculty of Science and Technology Syarif Hidayatullah State Islamic University Jakarta Jl.Ir.H.Juanda No.95 Ciputat 15412 Jakarta-Indonesia nurhayati@uinjkt.ac.id, fikriiki07@gmail.com.

[14]. E. Alshammary, A. Hadi, "Reviewing and Evaluating Existing File Carving Techniques for JPEG Files", Computer Science Dept. Princess Sumaya University for Technology israa88h@gmail.com,a.hadi@psut.edu.jo, 2016 Cybersecurity and Cyberforensics Conference, 978-1-5090-2657-9/16 $31.00 © 2016 IEEEDOI 10.1109/CCC.2016.21 .

[15]. D. Povar ,V.K. Bhadran , "Forensic Data Carving", Center for Development of Advanced Computing, Trivandrum, Ministry of Communications and Information Technology, Govt. of India {paward,bhadran}@cdactvm.in, I. Baggili (Ed.): ICDF2C 2010, LNICST 53, pp. 137–148, 2011,© Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2011.

[16]. M. N. Ashraf, "Forensic Multimedia File Carving", Master's Thesis, Department of Computer and Systems Sciences Royal Institute of Technology Summer 2012.

[17]. L. Aronson, J. van den Bos, "Towards an Engineering Approach to File Carver Construction", Netherlands Forensic Institute, The Hague, The Netherlands Centrum

Wiskunde & Informatica, Amsterdam, The Netherlands leon@holmes.nl, jeroen@infuse.org.

[18]. J. GUO, J. HE, N. HUANG "Research of Multiple-type Files Carving Method Based on Entropy", 4th National Conference on Electrical, Electronics and Computer Engineering (NCEECE 2015).

[19]. E. Durmus, P. Korus, N. Memon, Fellow, IEEE, "Every Shred Helps: Assembling Evidence from Orphaned JPEG Fragments", Copyright (c) 2019 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. E. Durmus and N. Memon are with Tandon School of Engineering, New York University, USA. (e-mail: {edurmus,memon}@nyu.edu). P. Korus is with Tandon School of Engineering, New York University, USA and also with the Department of Telecommunications, AGH University of Science and Technology, Poland. (e-mail: pkorus@nyu.edu).

[20]. A. A. SÜZEN , K. TAŞDELEN, "Recovering Multimedia Files from a Memory Image", RECOVERING MULTIMEDİA FILES FROM A MEMORY IMAGE … Politeknik Dergisi, 2018; 21 (3) : 731-737.

[21]. H. Lee and H.-W. Lee, "Block based Smart Carving System for Forgery Analysis and Fragmented File Identification," 인터넷정보학회논문지, vol. 21, no. 3, pp. 93–102, Jun. 2020.

[22]. S. Sari and K. Mohamad, "A Review of Graph Theoretic and Weightage Techniques in File Carving", Journal of Physics: Conference Series, vol. 1529, p. 052011, 2020. Available: 10.1088/1742-6596/1529/5/052011 [Accessed 17 July 2021].