International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Survey on Corrupted Video Recovery Using CODEC Specifications

Avinash Deshmukh[1], Manisha Desai[2]

[1]Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

[2]Professor, Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

Abstract: ***Digital forensics is one of the cornerstones to investigate criminal activities such as scam, computer security breaks or the distribution of criminal content. The significance and consequence of this research fields attracted various research institutes leading to substantial progress in the area of digital investigations. In digital forensics, recovery of a damaged or altered video file plays a crucial role in searching for evidences to resolve a criminal case. One essential piece of evidence is multimedia data. For this cause these papers gives an overview of the various available techniques to restore the corrupted video files. The main contribution of this paper is a discussion of existing and new approaches for the recovery of multimedia files.***

Keywords: Multimedia Data, Digital Forensics, file Carving, Data Recovery, and Survey.

1. Introduction

Recently, a large amount of video contents have been produced in line with wide spread of surveillance cameras and mobile devices with built-in cameras, digital video recorders, and automobile black boxes. Recovery of corrupted or damaged video files has played a crucial role in role in digital forensics. In criminal investigations, video data recorded on storage media often provide an important evidence of a case. As an effort to search for video data recorded about Criminal video data restoration and video file carving has been actively studied. Most existing video data restoration techniques attempt to restore the source data using meta-information recorded in the header of a file system. The meta-information of file system contains file information such as file name, time of modification, physical location, link, etc[1]. When the operator deletes a file, the corresponding file information in the meta-information of file system is updated as deleted although the video contents physically remain in the medium. Even though a video content exists in the media, it is challenging to recover the video data if the relevant meta-information is removed or altered. During digital investigations various types of media have to be analyzed. Relevant data can be found on different storage and networking devices and computer memory. Different types of data such as emails, electronic documents, system log and multimedia files have to be analyzed. Within this paper we focus on the recovery of multimedia files which are stored on either storage devices or computer memory using the file carving approach. File carving is a recovery technique which merely considers the contents and structures of files instead of file system structures or other meta-data which is used to organize data on storage media[1]. When part of the file was overwritten, restoration of a video file with meta-information only may not be successful in most situations. To tackle these problems, various techniques have been proposed such as File-Signature-Based Carving, Graph Theoretic Carvers, Approaches based on JPEG-Specifics, Smart Carving. We will see each of these techniques in detail in section II.

Mainly this paper focuses on the recovery of fragmented multimedia files. The recovery process is computationally complex; therefore we also discuss different optimizations to improve the overall performance of such file carvers. As video files can be seen as the prime example for multimedia files, we refer to these file types when we do not explicitly distinguish between file types that only contain one kind of media (visual or additive). The following section discusses formats of multimedia files that are currently used on the Internet.

1.1 Relevant Formats of Multimedia Files

This paper discusses the content-based recovery of multimedia files from various types of storage media. By "multimedia" files we refer to files that contain both additive and visual information, such as digital movies. Container files are files that may contain different files of different file types. The format of the container file describes how the different contents (sub files) are arranged within such a file. Multimedia files basically consist of header or metadata, as well as visual, additive, and text-based information. Therefore it can also be referred to as a "compound document". Video data is compressed by a so called Advanced File Carving Approaches for Multimedia Files Poisel, Tavolato and Tjoa "codec" program which reduces the required amount of data in a (usually) lossy fashion. With support for HTML5, web browsers support the playback of video files natively. According to a recent study more than 50% of now a day's web browsers support the playback of HTML5 videos. Since 2009 there was an increase of 66%. Therefore, in our study, we focused on the recovery of video formats supported by the HTML5 draft standard. According to the HTML5 draft standard and documentation of major web browsers, the following video formats are supported by web browsers:

- Ogg containers with Theora video streams,
- MPEG-4 containers with H.264 video streams and
- Web M containers with VP8 video streams.

Web browsers store the contents of surfed web pages in a cache for faster access. As with HTML5 multimedia files are part of web pages contents, it can be expected, that such files can also be found in the cache of web browsers. Cohen described the investigation of a browser's cache using the PyFlag framework. A browser's cache is typically organized in regular files on a file system and therefore it is possible to acquire content from visited web pages with tools from this field as well.

1.2 Related work

Tools that focus on the recovery based on file system structures have been presented by Carrier. His findings have been compiled into a suite of small programs "The Sleuth Kit". Each of these programs has its dedicated purpose, such as calculating file/directory listings or storage units of specific files. One of the first file carvers which has been mentioned in scientific research papers is "Scalpel: A Frugal, High Performance File Carver". It only ropes the recovery of unfragmented files by defining signatures for the beginning and end of file types that should be recovered. Data between these defined signatures is extracted and represents the recovered files. Garfinkel extended the signature based approach by considering meta-data in files intended for the recovery. Some file headers contain information about the file length. In case a file format does not have a distinctive signature for the end of a file this information can be interpreted to extract the files' contents. Further, Garfinkel proposed the "Bifragment Gap Carving" approach for files that are fragmented into exactly two fragments. After determining the starting and end points of a file, the data between these two points is tested for its validity. If the test fails a gap which is excluded from the validation procedure is introduced. The boundaries of this gap are modified until the validation succeeds. As a subset of recovering files, Garfinkel discussed the recovery and validation of objects. Especially for container formats such as the Portable Document Format (PDF) it is possible to extract meaningful information from within such a file[2]. In this context the purpose of a validator is to ensure that data recovered by file carving can be decoded successfully. However, certain decoders which have been used as validator do not necessarily generate an error if they encounter invalid information. Further, it is possible to successfully decode syntactically correct, but semantically invalid data. Therefore the validity of investigated material has to be ensured by further analysis, e. g. semantic validation. Cohen considered the recovery of fragmented files and described the carving process as being equivalent to estimating a mapping function between bytes copied from an image of storage media to the recovered file. Files could be recovered using a generator that produces all possible mapping functions. Results of these mapping functions are evaluated for their validity. The downside of this approach is the vast number of combinations (see "Reassembly" later in this paper). Cohen also discussed another critical component of a file carving system: the discriminator. This component can tell if a recovered file is corrupt or likely to be correct. It's result is fed back into the mapping function generator. This way the recovery process can be improved by excluding mappings that are incorrect before they are evaluated for their correctness. Cohen described mapping function generators for the PDF and ZIP file formats. Further improvements have been proposed by Pal and Memon. For the recovery of fragmented files they introduced the Smart Carving architecture which is not limited to the number of pieces into which a file is fragmented by the underlying file system or storage algorithm.

2. Existing Techniques

2.1 File-Signature-Based Carving

File fragments are identified by comparing byte- sequences contained in headers and footers with values stored in a database containing well known values for precise file types. Former file carving approaches where computationally rigorous and required large amounts of memory. Scalpel was introduced to overcome these limiting factors. The operation of Scalpel is performed in two chronological passes. During the initial pass the whole disk image is indexed by reading chunks of several megabytes and searching for file headers. After finding headers in a large piece, footers are identified as well and stored in a database. This database is examined to only contain header-footer tuples which fulfill the constraints for the maximum size of files to be recovered[2]. The contents of the database are used to put up working queues which contain locations for the file extraction process in the second pass. During the second pass the disk image is again processed in chunks to copy recovered files to the place where recovered files are kept. Carving files using Scalpel has further been improved by removing the final step of copying recovered files. Instead a file system is developed using the FUSE library. The user accesses the investigated storage area by mounting an image using the Scalpel file system in which the contents of the header-footer database are presented as actual files. Further improvements for the carving of contiguous files have been categorized based on different properties for files to be recovered:

- Header/footer carving: for removing data between distinct start and end of file markers (string sequences),
- Header/maximum size carving: with further analysis for the longest valid string sequence that still validates,
- Header/embedded length carving: which is used for file formats that do not have distinctive footers for the end and
- File trimming: for "byte-at-a-time formats" that do not have obvious footers by trimming characters at the end until the file no longer validates.

2.2 Graph Theoretic Carvers

File carvers assembling fragments based on graph theoretical algorithms have been proposed especially for text-based media by Shanmugasundaram as well as for digital images by Pal. Approaches for text-based data proposed by Shanmugasundaram assign candidate probabilities for their adjacency to recovered file fragments. For text-documents these probabilities can be determined using a sliding-window algorithm which evaluates the statistics for symbol usage in a language or, for generic data, is based on statistical models used for data compression.

Possibility which have been assigned to fragments are then used to determine the permutation which maximizes the sum of candidate probabilities of neighboring fragments. This mathematical problem is equivalent to finding a maximum weight Hamiltonian path in a whole graph[3]. As this problem turned out to be intractable, heuristics have been introduced to provide the best solution. The approach proposed by Pal assigns probabilities to file fragments of digital images which are then put together using different graph theoretic algorithms, e. g. an adapted version of the Shortest Path First (SPF) algorithm which yielded the best results for seven datasets of images. This method can only be applied to a video file with two fragments and this technique has limitation when the gap between the two file fragments is large.

2.3 Approaches based on JPEG-Specifics

The number of different graphics formats used in web-content is low. With the JPEG-format being one among these lots of related work in this field concentrated on this format. Karres and et al. proposed a method which uses the so called restart markers (RST) of the JPEG file format to reassemble non-differential Huffman entropy coded baseline sequential Discrete Cosine Transform (DCT) JPEG image fragments. With restart markers being used the scan is interrupted at regular intervals by a exact bit pattern. Further unprocessed pixels of an image are grouped into 8x8 pixel blocks which are transformed into the frequency domain using Discrete Cosine Transformation (DCT). Most important item here is the first which represents the zero frequency DC coefficients. The data between RST markers is called Minimum Coding Unit (MCU) and it is the smallest part of an image which can be decoded if it is integral. Luminance DC values in all resume intervals are used to form DC value chains[4]. The DC component chains are then analyzed using a sliding window approach to identify the order of fragments of a specific image. Enhancements using different aspects, e. g. by considering the Define Huffman Table (DHT) segment, of the JPEG-format have been proposed.

2.4 Smart Carving

Reassembling objects out of their fragments which are randomly mixed with fragments from other files is a problem that can be found in many different disciplines. To overcome the lack of research in the field of digital forensics Pal et al. Proposed a generic approach for images which comprises the following three steps for a document reassembly process (see figure 1):

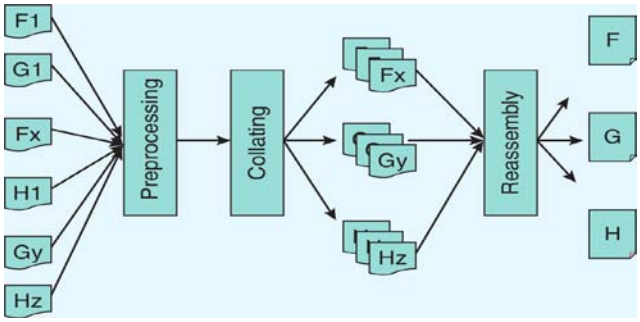

Figure 1: The three components of Smart Carving

- Preprocessing: During this step investigated data is prepared to be usable by forensics investigation methods. This comprises the decryption of encrypted devices, as well as the deletion of all known clusters base on file system metadata.
- Collating: The identification of fragments is performed during the second phase. After their identification fragments are collected in groups of the same type to be assembled into their original files in subsequent steps.
- Reassembling: Finally the fragments recovered in the previous phases are reassembled to their original files. It is therefore necessary to find the fragmentation points for each unrecovered file. In case of incorrect reassembly recent tools.

Smart Carving technique was proposed to restore a file without being restricted by the number of fragments [4]. This technique, if it identifies the occurrence of fragmentation, combines the permutations of the fragment components and searches for the order of the fragments. They technique consists of three steps: preprocessing, collation, and reassembly. In the preprocessing step, they collect the called block part, which was not allocated to a file, using the file system information to reduce the size of the data to analyze. The collation step categorizes the collected blocks in the preprocessing step according to a file format. The reassembly step determines fragmented parts and merges them into a file. In this they extended Smart Carving to apply to multimedia files. In the reassembly step, they increased the restoration rate of multimedia file by assigning a weight to each fragment using the decoded frame difference. However, the method presented in, which is also a file-based approach, has a limitation to restore a video file when a part of video file is overwritten. In addition, graph theoretical carving was proposed by which the k-vertex disjoint graph is created to piece together fragments. This technique proposed various greedy heuristic restoration techniques with which to use the matching technique and search for the sector/block order. The weight of all the fragment pairs should be calculated in advance, however, which is costly.

3. Conclusion

Thus From above review we come to know that all the existing techniques are based on file structure instead of frame structure. Such techniques can be used to restore video file which is severally fragmented but in case video file is partially overwritten restoration is not possible by such techniques. Hence Frame-Based recovery comes into picture. By using frame-based recovery we can recover partially overwritten as well as severally fragmented video files.

References

[1] Rainer Poisel and Simon Tjoa," Forensics Investigations of Multimedia Data: A Review of the State-of-the-Art", 2011 Sixth International Conference on IT Security Incident Management and IT Forensics.
[2] Nasir Memon and Anandabrata Pal" Automated Reassembly of File Fragmented Images Using Greedy

Algorithms", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, NO. 2, FEBRUARY 2006.
[3] Vrizlynn L. L. Thing, Tong-Wei Chua, and Ming-Lee Cheong," Design of a Digital Forensics Evidence Reconstruction System for Complex and Obscure Fragmented File Carving", 2011 Seventh International Conference on Computational Intelligence and Security.
[4] Rainer Poisel and Simon Tjoa,"A Comprehensive Literature Review of File Carving", 2013 International Conference on Availability, Reliability and Security.

Author Profile

Avinash Deshmukh Research Scholar RMD Sinhgad School of Engineering, University of Pune. He has received B.E. in Computer Engineering from University of Pune. Currently he is pursuing M.E. in Computer Engineering from RMD Sinhgad School of Engineering, Warje, University of Pune, Pune .

Prof. Manisha Desai received the B.E. from University of Pune and M.Tech Degrees in Computer Engineering from Defense Institute of Advance Technology, Pune. She is working as Assistant Professor in Department of Computer Engineering, RMD Sinhgad School of Engineering, Pune. She is having more than Three year experience.