**File carving** is the process of reassembling computer files from fragments in the absence of [filesystem metadata](https://en.wikipedia.org/wiki/Filesystem#Metadata).

All filesystems contain some metadata that describes the actual file system. At a minimum, this includes the hierarchy of folders and files, with names for each. The filesystem will also record the physical locations on the storage device where each file is stored. As explained below, a file might be scattered in fragments at different physical addresses.

File carving is the process of trying to recover files without this metadata. This is done by analyzing the raw data and identifying what it is (text, executable, png, mp3, etc.). This can be done in different ways, but the simplest is to look for the signature or "magic numbers" that mark the beginning and/or end of a particular file type. For instance, every Java class file has as its first four bytes the hexadecimal value CA FE BA BE. Some files contain footers as well, making it just as simple to identify the ending of the file.

Most file systems, such as the FAT family and UNIX's Fast File System, work with the concept of clusters of an equal and fixed size. For example, a FAT32 file system might be broken into clusters of 4 KiB each. Any file smaller than 4 KiB fits into a single cluster, and there is never more than one file in each cluster. Files that take up more than 4 KiB are allocated across many clusters. Sometimes these clusters are all contiguous, while other times they are scattered across two or potentially many more fragments, with each fragment containing a number of contiguous clusters storing one part of the file's data. Obviously large files are more likely to be fragmented.

Simson Garfinkel reported fragmentation statistics collected from over 350 disks containing FAT, NTFS and UFS file systems. He showed that while fragmentation in a typical disk is low, the fragmentation rate of forensically important files such as email, JPEG and Word documents is relatively high. The fragmentation rate of JPEG files was found to be 16%, Word documents had 17% fragmentation, AVI had a 22% fragmentation rate and PST files (Microsoft Outlook) had a 58% fragmentation rate (the fraction of files being fragmented into two or more fragments). Pal, Shanmugasundaram, and Memon presented an efficient algorithm based on a greedy heuristic and alpha-beta pruning for reassembling fragmented images. Pal, Sencar, and Memon introduced sequential hypothesis testing as an effective mechanism for detecting fragmentation points. Richard and Roussev presented Scalpel, an open-source file-carving tool.

File carving is a highly complex task, with a potentially huge number of permutations to try. To make this task tractable, carving software typically makes extensive use of models and heuristics. This is necessary not only from a standpoint of execution time, but also for the accuracy of the results. State of the art file carving algorithms use statistical techniques like sequential hypothesis testing for determining fragmentation points.

In most cases, when a file is deleted, the entry in the file system metadata is removed but the actual data is still on the disk. File carving can be used to recover data from a hard disk where the metadata was removed or otherwise damaged. This process may be successful even after a drive is formatted or repartitioned.

File carving can be performed using free or commercial software and is often performed in conjunction with computer forensics examinations or alongside other recovery efforts (e.g. hardware repair) by data recovery companies. Whereas the primary goal of data recovery is to recover the file content, computer forensics examiners are often just as interested in the metadata such as who owned a file, where it was stored, and when it was last modified. Thus, while a forensic examiner could use file carving to prove that a file was once stored on a hard drive, he or she might need to seek out other evidence to prove who put it there.

SBCTF{n1c3\_c4rv1n6\_w3ll\_d0n3}



