File carving

From Wikipedia, the free encyclopedia

File carving is the process of reassembling computer files from fragments in the absence of filesystem metadata. The carving process makes use of knowledge of common file structures, information contained in files, and heuristics regarding how filesystems fragment data. Fusing these three sources of information, a file carving system infers which fragments belong together.

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 the fragmentation point.

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 are 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. 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 point. Richard and Roussev presented Scalpel, an open-source file carving tool.