White paper on Incremental Disk Drive Imaging

Abstract

Hard disks often partially fail in a way that some areas are unreadable, or have an unacceptably high error rate, while large areas of the disk can be read correctly. In order to do a forensic data recovery it is essential to create a workable disk image before the disk fails totally. Incremental imaging is a technique to allow and image of the disk to be created while not getting stuck on a bad area, or by putting excessive strain on the drive by doing many read retries.

Disk Image

There are two possible approaches to disk imaging, creating an image file, or a physical clone onto a second disk of equal or larger size. The solution in this paper will be to look at the file creation method, although similar principles can be applied to both methods.

The foundation for an incremental image is a raw image file, and the simplest to work with is equivalent to a Unix DD file. ie this is a file made up of every sector in sequence. There are two significant differences between a standard DD file, and the incremental image file, being how failed and skipped sectors are handled, and how the image is built up.

Image file structure

Any image file should be compatible with standard data recovery tools, and so basing a new structure on the DD standard is a sound starting point. There need to be three types of sector stored

- A sector that has been read successfully
- A sector that has failed to be read
- A padded sector for one not yet read

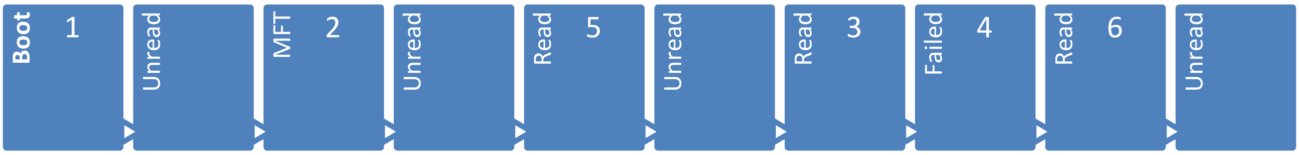
Ultimately, the image file will represent the complete disk, and so the first stage maybe to create a complete blank file, or rather easier will be to append to the file when extra sectors have been read. How skipped sectors, padded sectors are stored is key to an incremental imaging system. The easiest way would be to pad each sector with zeros. A complete file can be built up but it is then impossible to indicate which sector had failed to be read, and which had been skipped. The solution is therefore is pad any skipped and failed sectors with a pattern that can be recognised, but also one that is not common in data. Ideally it should also be clear that when data has been recovered, that a file may contain padded data. Visually, the easiest solution is a sector full of a fixed character, and one chosen by CnW Recovery is an upper case 'Z' (0x5a). This process can be taken a stage further by making use of the final bytes in a sector. These bytes can be used to store the sector number and a flag to indicate if the sector is a padded one, or indicates that the original sector had failed. The file is still a DD compatible file, but there is now useful data on the status of each sector.

Other approaches could be to have a separate status file, giving data for each sector. This however limits the number of recovery programs that can process this data. For that reason, a single file approach is a sensible compromise.

Imaging sequence

By having an intelligent image file structure, there is no requirement to image a disk starting at sector 0, and going to the end in a single pass. Sector 0, the master boot record (MBR) is actually a common sector to fail. There are a few considerations as for the most appropriate sequence to use. A good starting point is to get an idea of where the disk is failing by reading different areas of the disk. This could be by hand, or by a simple program to scan the disk very quickly with the lowest number of retries, and never trying to read many sectors in a bad area. It is very typical for a disk to work well at the end of the disk, but have large failures around the MFT directory, or Macintosh Catalogue area. The second pre image test is to try and establish how full the disk is, and where the data storage may stop. There is no point spending time initially imaging a blank area of disk, because at all times it must be considered that the drive could fail fatally on the next sector read.

The next level of decision has to be determined by the type of data recovery or investigation required. If a single specific file is being looked for then it pays to work hard on the MFT or Mac catalogue until the file entry is found. If though it is a case of general investigation of the whole disk it will make most sense to image areas of the disk that read without error, though the directory area is always an area with a considerable about of useful information.



The chart above shows a possible sequence to image a disk, starting with the boot sector, followed by the MFT, followed by good areas of data. The image at that stage will include known good data, known bad data, as well as skipped areas. A very valuable aspect of incremental imaging is being able to work out where to image next. If the disk is in a very poor state then just three sectors can be of immense use, the boot sector, the Bios Parameter Block and first block of the $MFT. In most instances, this will allow the location of all fragments of the $MFT to be found, and hence imaged. With a complete $MFT, the location and names of all the files can be determined, without imaging the complete disk. It is important that any recovery program will work in conjunction with an incremental image, as the image may be shorter that the original disk, and will have blank areas.

Read skipping

A critical aspect of imaging damaged drives is not to dwell on bad areas of the disk. A useful tool when imaging is to set a threshold for the number of failed sectors to read. Once the number of sequentially failed sectors has been reached, the program should pad an area of data, and continue trying to read further down the disk. The padded areas can always be filled in by using the shadow technique described below.

Shadow Drive

A second approach to incremental imaging is to use a shadow drive along with the file image file. The procedure is that a sector is read from the image file, but if it is marked as unread, then the physical drive is read. If the sector can be read, then this is added to the image file, otherwise the image file is marked as an unreadable sector. It is recommended that this approach can be used after an initial imaging as described above. Once for instance the MFT or Catalogue has been established, it is then possible to selectively recover maybe just .JPGs, or .XLSX documents. By using the image file s the main reference, physical disk access is limited to only new sectors on demand.

As the shadow drive reading updates the main image file, all the sectors added should be hashed and these hash values logged. This makes for a messy log file and so it is preferable, but not essential, to do as much imaging as possible in large blocks.

Recovery software

Incremental imaging works best then the recovery software is very tolerant of both missing areas of the disk and also if it can allow for skipped sectors, and distinguish between failed and skipped sectors. The software must also be capable of reconstructing critical sectors (eg boot sector) that are missing, or in very bad cases resorting to just data carving.

Forensic considerations

An incrementally imaged disk will almost certainly be incomplete so forensically this has to be considered. The emphasis of any report, must be what has been found, rather than what was not found. Ideally the report should indicate the percentage imaged, skipped and failed.

The keyword in forensic investigation is repeatability. Ideally someone else must be able to get the same results, but with a failing disk this may be impossible. A sector can occasionally be read and produce different contents, or more commonly not be able to be read again. It is therefore essential that every sector run that is added to the image file is hashed. This way, one can state that sectors xx to yy did produce a hash of zz, and this can be verified. A final stage of any incremental image must be a hash of the complete DD file.

Summary

Incremental imaging is a useful technique to recover as much data as possible from a failing disk drive, or one with damaged areas. The aim is to read any sector of the drive only once, and where possible focusing reading on the most useful areas of the disk. Forensic limitations have to be considered, but files found will still be valid.

Michael Cotgrove
www.cnwrecovery.com

About the author

Worked for InterMedia / eMag Solutions from 1984 until 2004. Developed MM/PC tape program. Since 2004 has been developing CnW Recovery forensic data recovery software.