

Definitions

- Sectors



Created by leo-graph.com
from Noun Project

- File system



Created by BomSymbols
from Noun Project

NMAAHC Workshop, Caroline Gil and Eddy Colloton
2021

We're going to define and unpack terms that we will return to regularly throughout the workshop.

Sectors are a unit of measurement on a disc - granular discrete space that varies by media type, usually a multiple of 512 bytes, traditionally 2048 on CD-ROMs or DVDs. While measured in data, sectors actually refer to a physical space on a disk. Sometimes you'll hear blocks and sectors get used interchangeably. The terms do often serve the same purpose, they're referring to a granular, discrete, small amount of data on a drive. Blocks can be a single sector, and therefore literally refer to the same thing. However, a block can be multiple sectors (typically adjacent to one another), defined as a discrete unit by the file system.

A file system can be thought of as an "address book" for data, it explains where information is, how it is structured. The file system will say, "this file - workshop.pdf" is stored across these physical sectors," or "it's stored on this block and that block." Examples of file systems include FAT(12,16,32), HFS+, NTFS, ISO9660, etc.

Definitions

- Digital forensics
- Forensics bridge/write blocker



Created by Nanda
from Noun Project



Created by ade
from Noun Project

NMAAHC Workshop, Caroline Gil and Eddy Colloton
2021

A commonly used definition of digital forensics comes from the Digital Forensic Research Workshop (DFRWS) of 2001: “Digital forensics is the practice of identifying, extracting and considering evidence from digital media such as computer hard drives. Best practices in digital forensics aim to ensure that materials of evidentiary value can be effectively isolated and extracted in a scientific manner that will bear the scrutiny of a court of law.”

(<https://www.dpconline.org/docs/technology-watch-reports/810-dpctw12-03-pdf/file>)

You can see in this definition a focus on forensics in the legal sense. While the term digital forensics is used more broadly now to include much of the work we do in cultural heritage, digital forensics is very much tied to the world of law enforcement, as we’ll discuss in the workshop.

The thing I want to emphasize right now, in terms of a definition of the concept of digital forensics, is the emphasis on chain of custody, regardless of the context of the practice (either investigative or archival). The aim of digital forensics best practices is to be able to demonstrate exactly how data has been moved from one place to another, and during the process, no changes have occurred. You can see how this would suit the needs of both a legal investigation or preservation efforts. Similarly, there is a preference among many in digital forensics for open source tools and technology, just as there is in the digital preservation community, for many of the same reasons: open source tools and formats offer a far higher level of documentation, and can be reverse engineered if necessary.

Forensic bridges, also known as write-blockers, are employed to allow the safe transfer of media from a carrier to your workstation without compromising or altering the integrity of the data. The USB bridges are the hardware I encounter and use most

frequently, and I imagine several of you have used those before, too.

Digital Forensics

The collage features three distinct digital forensics resources:

- NIST Computer Forensics Tools & Techniques Catalog:** A screenshot of the National Institute of Standards and Technology's catalog. It includes a search bar, a sidebar with 'Forensic Tool Functionalities' (e.g., Cloud Services, Data Analytics, Deleted File Recovery), and a search results page for 'Hardware Write Block' tools. The search parameters are set to 'any' for OS/runtime environment, 'any' for host access interface (PCIe, eSATA, SATA), and 'any' for write-blocked interfaces (CompactFlash, SD card, PCIe).
- WiebeTech Website:** A screenshot of the WiebeTech homepage, featuring a navigation menu and a prominent banner for 'Introducing Ditto x86', a high-performance USB 3.1 Gen 2 device for forensic investigation.
- Digital Intelligence Website:** A screenshot of the Digital Intelligence website, showing a navigation menu and a banner for an NMAAHC Workshop presentation by Caroline Gil and Eddy Colloton in 2021.

We're going to do a deep dive on write blockers and digital forensics hardware tomorrow, so for now I'm just going to briefly point out the companies whose hardware I encounter most often in the cultural heritage sector - WiebeTech and Digital Intelligence. Digital Intelligence sells products created by Tableau, who are in turn owned by the corporation that makes our DAMS, opentext.

Digital forensics tools are evaluated by NIST, the National Institute of Standards and Technology. Their website features a catalog of software and hardware tools, that they are calling their Computer Forensics Tools & Techniques Catalog.

Definitions

- Digital Carrier
 - Hard Drive
 - USB “thumb drive”
 - DVD
 - CD-ROM
 - Floppy disk
- Volume
 - Partition of a disk
 - file system of a CD-ROM
- Virtual Machine
 - Stored as disk images (.vdi, for example)
 - Can be made from disk images
 - To mount and explore disk images

NMAAHC Workshop, Caroline Gil and Eddy Colloton
2021

When Caroline and I use the term “digital carrier” throughout the workshop, we’re using that as a “catch all” term to refer to any physical object that can be disk imaged. We’ll also sometimes say “volume” which is an even further removed “catch-all” to refer to any digital object that can be disk imaged, even those that are not “media bound.” So a volume may be something non-physical such as an individual partition of a disk. The device being disk imaged is often referred to as the “target” drive by disk imaging software

I imagine everyone is familiar with the term virtual machines, as you downloaded and installed BitCurator, as a VM before the workshop. But I do want to emphasize the relationship between disk images and VMs. Virtualization software store the virtual hard drive of a VM as a disk image, in the case of Virtual Box, this is the vdi file. When creating disk images of computer hard drives, one possible way of QCing and interrogating the completeness of the disk image is to create a VM from the disk image. Also VMs can be a helpful tool for mounting and working with disk images given the lack of interoperability of software and certain file systems between operating systems. Having Linux, macos and Windows at your discretion can be very helpful when you have a collection of digital carriers with a wide variety of file systems.

Definitions

- Offset
 - the number of bytes “in” to a volume or digital carrier a particular piece of information begins
- CRCs
 - Like checksums, only smaller

```
Disk /home/yjwong/disk.img: 250.1 GB, 250058268160 b
255 heads, 63 sectors/track, 30401 cylinders, total
Units = sectors of 1 * 512 = 512 bytes
```

Device	Boot	Start	End
/home/yjwong/disk.img1		2048	3905535
/home/yjwong/disk.img2	*	3905536	488394751

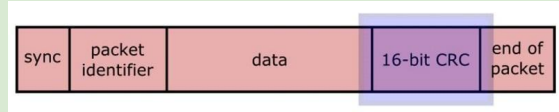


Image source: The Cyclic Redundancy Check (CRC): Finding—and Even Correcting—Errors in Digital Data by Robert Keim

NMAAHC Workshop, Caroline Gil and Eddy Colloton
2021

Disk imaging literature and how tos will often reference “offsets.” They’re referring to “byte offsets,” which are literally the address or location of a particular piece of information. Almost always this is used to describe the beginning of a file system. If you are mounting a particular file system or partition that is stored within a disk image, some command line applications will require you to list the offset. More commonly you will see the offset listed in metadata output of a disk image, sometimes helpfully labeled only “start” and a big list of numbers.

CRCs have been used in computing for a long time, for example, they’re built in to floppy disk encoding specs. CRC stands for cyclic redundancy check. They’re for error correction. It’s a hash that is built into encoding standards so that when a device reads a disk, it hits a mini checksum, to ensure it’s read the disk correctly. We’ll talk about how data is laid out on different formats, so this point will be re-iterated, but it’s good to keep in mind that digital carriers are often being read incredibly fast, and so the assumption has always been that read errors will happen, and systems are designed to ensure that risk of those errors is mitigated.