Digital Forensics

Federico Conti

2024/25

Contents

Forensic Acquisition

Acquisition is the process of cloning or copying digital data evidence.

- forensically sound (integrity and non-repudiation)
  - the copies must be identical to the original
  - the procedures must be documented and implemented using known methods and technologies, so that they can be verified by the opposite party
- a critical step
  - proper handling of data ensures that all actions taken on it can be checked, repeated, and verified at any time
  - incomplete or incorrect handling of data has the potential to compromise the entire investigation

It is generally recommended to avoid conducting analysis on the original device (namely, best evidence).

Creating a forensic image is typically considered the most effective method for preserving digital evidence (One or more, usually two).

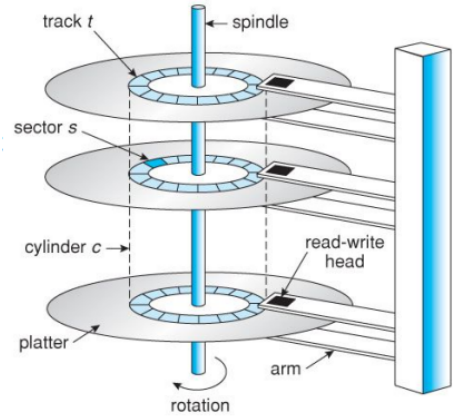
Accessing the original media only once during the acquisition phase can help minimize the risk of altering or damaging the evidence

HDD and SSD technologies

HDD

A **hard disk** is a sealed unit containing a number of platters in a stack:

- each platter has two wothey moverking surfaces
- each working surface is divided into a number of concentric rings called tracks
  - the collection of all tracks that are the same distance from the edge of the platter is called a cylinder.
- each track is further divided into sectors
  - a sector is the smallest unit that can be accessed on a storage device
  - traditionally containing 512 bytes of data each, but recent hard drives have switched to 4KB sectors (Advanced Format)
  - a cluster is a group of sectors (from 1 to 64 sectors) that make up the smallest unit of disk allocation for a file within a file system



The data on a hard drive is read by read-write hands. The standard configuration uses one head per surface and they moves simultaneously from one cylinder to another.

A **low level format** is performed on the blank platters to create data structures for tracks and sectors

- creates all of the headers and trailers marking the beginning and ends of each sector
- header and trailer also keep the linear sector numbers (cf. LBA later), and error-correcting codes (ECC)

All disks are shipped with a few bad sectors (additional ones can be expected to go bad slowly over time).

- disks keep spare sectors to replace bad ones
- ECC calculation is performed with every disk read or write: if an error is detected but the data is recoverable, then a *soft error* has occurred if the data on a bad sector cannot be recovered, then a *hard error* has occurred. A bad sector can be replace with a spare one (but any information written is usually lost)

Older hard drives used a system called **CHS (Cylinder-Head-Sector)** addressing to locate data on the disk. This method relied on:

- Cylinders (C) → The track number (a ring of data on a disk platter).
- Heads (H) → The read/write head that accesses a platter's surface.
- Sectors (S) → The smallest unit of storage on a track.

Eample: CHS (100, 2, 30) would mean: Cylinder 100; Head 2 (indicating the second platter side); Sector 30 (the exact location on that track).

CHS had physical limitations → It could only handle disks up to 504 MB (later ECHS extended this to 8 GB).

Instead of using three values (CHS), **Logical Block Addressing(LBA)** assigns a single number to each sector. LBA starts at 0 and counts up sequentially. The operating system and file system treat the disk as a continuous array of sectors, making it easier to manage.

Example: LBA 123456: The 123,456th sector on the disk.

SSD

A **Solid-State Drive (SSD)** is a non-volatile storage device, meaning it retains data even when the power is off. Unlike Hard Disk Drives (HDDs), which use spinning magnetic platters, SSDs rely on flash memory chips to store data.

- The smallest unit of an SSD is a page, which is composed of several memory cells: the page sizes are 2KB, 4KB, 8KB, 16KB or larger (usually they are 4 KB in size).
- Several pages on the SSD are summarized to a block: the block size typically varies between 256KB (128 pages * 2KB per page) and 4MB (256 pages * 16KB per page)

Operation	Description
Read and Write	An SSD can read and write data at the page level.
Write	Writing is only possible if other pages in the block are empty. Otherwise, it leads to write amplification.
Erase Data	An SSD can only erase an entire block at once due to the physical and electrical characteristics of the memory cells.

Modifying data requires a **Program/Erase (P/E) cycle**.

- During a P/E cycle, an entire block containing the targeted pages is written to memory
- The block is then marked for deletion, and the updated data is rewritten to another block

The erase operation does not happen immediately after data is marked for deletion. Instead, the SSD performs it asynchronously when necessary to optimize performance.

Garbage Collection helps free up space efficiently while minimizing interruptions to read/write operations.

Every flash memory cell has a limited number of P/E cycles before it wears out Wear leveling is a technique used by SSDs to distribute writes evenly across all memory blocks.

Unlike HDDs, SSDs cannot simply overwrite deleted files. If deleted data is not managed properly, unnecessary data copies can cause write amplification, increasing wear on the SSD. TRIM command allows the OS to inform the SSD which pages are no longer needed, enabling it to manage space more efficiently.

Forensic Investigators Face a Big Problem with SSDs|

TRIM permanently deletes data by triggering the garbage collector, making data recovery impossible. The garbage collector operates independently within the SSD controller, meaning:

- Even a hardware write blocker (a forensic tool to prevent data changes) cannot stop it.
- Data can change midway or between acquisitions, complicating forensic investig

Interface standards and protocols

Disks are accessed using standard interfaces that define how they physically and logically connect to a computer system. Each standard improves data transfer speeds and fixes limitations from older technologies. A disk interface consists of two key components:

- Physical Connection → Defines the type of cable and connector used to attach the disk to the system.
- Logical Connection → Defines the protocol that controls how data is transferred between the disk and the computer.

ATA (Advanced Technology Attachment):

a widely used interface standard with multiple versions:

- ATA-1, ATA-2, . . . , ATA-8 → Each new version improves speed and capacity.
- ATAPI (ATA Packet Interface) → Allows removable media (CD/DVD drives) to be connected using ATA but still uses SCSI commands for data transfer.

Type	Description
PATA (Parallel ATA)	Older, also known as IDE (Integrated Drive Electronics). Uses ribbon cables with multiple pins.
SATA (Serial ATA)	Modern standard introduced in ATA/ATAPI-7. Uses thin serial cables, improving speed and efficiency.
SATA 3.0	Supports speeds up to 6 Gbit/s (SATA-600). Common in modern HDDs and SSDs.
ATA-8	Introduced optimizations for SSDs, including TRIM support.

SCSI (Small Computer System Interface):

now replaced by SAS (Serial Attached SCSI) that is based on the SCSI standard, but uses a serial interface to connect storage. Better scalable and faster (supports data transfer rates of up to 24 Gbit/s).

NVMe (Non-Volatile Memory Express):

Uses a PCIe interface to connect storage devices to a computer. Commonly used for high-performance solid-state drives since it supports data transfer rates of up to 32 GB/s.

USB (Universal Serial Bus):

- uses a serial interface to connect storage devices to a computer
- mass storage is the standard protocol used for storage devices
- commonly used for external hard drives, flash drives, and other portable storage devices
- USB 3.1 Gen 1 standard supports speeds up to 5 Gbit/s, while the USB 3.1 Gen 2 standard, supports speeds up to 20 Gbit/s.

ATA specification

Introduced in ATA-3, **hard disk passwords** are an optional security feature designed to restrict unauthorized access to a hard drive. However, it is a lock mechanism, not encryption, meaning data remains unencrypted and could be accessed by other means.

There are two passwords:

1. User Password → Set by the owner.
2. Master Password → was designed so an administrator can gain access in case the user password was lost (every hard disk is initially supplied with an undocumented master password).

If passwords are being used, there are two modes that the disk can operate:

1. high security mode: the user and master password can unlock the disk
2. maximum-security mode: the user password can unlock the disk but the master password can unlock the disk after the disk content have been wiped

A protected HD will require the `SECURITY_UNLOCK` command to be executed with the correct password before any other ATA command. After the password has been entered, the disk works normally until the disk is powered on

Some ATA commands are still enabled on the HD when it is locked (so it may show as a valid disk when it is connected to a computer); however, trying to read data from a locked disk will produce an error

Setting the Password:

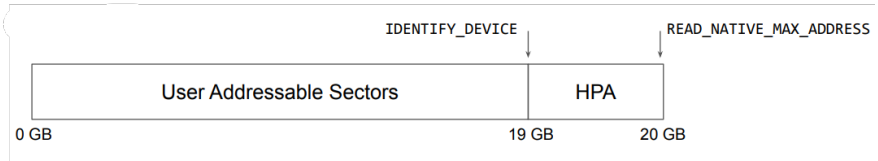
- Can be configured via BIOS settings.
- Linux users can manage HDD passwords using tools like hdparm.

Introduced in ATA-4, the **Host Protected Area (HPA)** is a hidden storage section on a hard disk that is not accessible to the operating system. It is used mainly by hardware vendors for system recovery files, diagnostics, or factory reset tools. A HPA is at the end of the disk and when used, it can be accessed by reconfiguring the hard disk.

Two ATA commands that return maximum physical addressable sectors

- `READ_NATIVE_MAX_ADDRESS`: return the maximum physical address
- `IDENTIFY_DEVICE`: return only the number of sectors that a user can access

To create an HPA, the `SET_MAX_ADRESS` command is used to set the maximum address to which the user should have access (to remove it, use `SET_MAX_ADDRESS = READ_NATIVE_MAX_ADDRESS`).



the `SET_MAX_ADRESS` command support different settings, e.g.,

- volatility bit: the HPA exist after the hard disk is reset or power cycled (otherwise the effect is permanent)
- locking command: prevents modification to the maximum address until next reset

when the BIOS requires to read/write some data in the HPA it uses `SET_MAX_ADRESS` with volatility bit and locking.

It is possible to protect settings with a password (different from HD passwords).

The **Device Configuration Overlay (DCO)** was introduced in ATA-6 and allows manufacturers to limit the apparent capabilities of a hard disk. This feature enables backward compatibility with older systems but can also be exploited to hide data.

A computer uses the `IDENTIFY_DEVICE` command to check an HDD's specifications (size, features, supported commands). If a DCO is applied, the `IDENTIFY_DEVICE` command will not show the actual full disk size or features.

This is achieved using two special ATA commands:

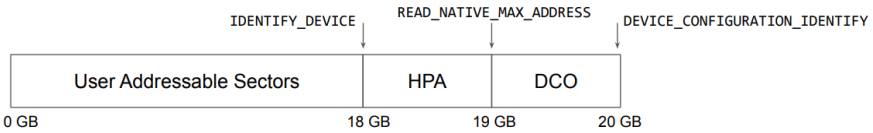
- `DEVICE_CONFIGURATION_SET` → Modifies the DCO settings to restrict visible disk space or disable features.
- `DEVICE_CONFIGURATION_RESET` → Restores the original settings, removing the DCO restrictions.

Example: A 2 TB hard drive can be made to appear as a 500 GB drive, hiding the remaining 1.5 TB from the operating system.

The DCO and HPA can co-exist on the same HDD (but DCO must be set first).

The `DEVICE_CONFIGURATION_IDENTIFY` command return the actual features and size of a disk:

- we can detect DCO if `DEVICE_CONFIGURATION_IDENTIFY` $\neq$ `IDENTIFY_DEVICE`



At least three different methods for detecting HPA on Linux: `dmesg`, `hdparm`, and `disk_stat` (https://wiki.sleuthkit.org/index.php?title=Disk_stat)