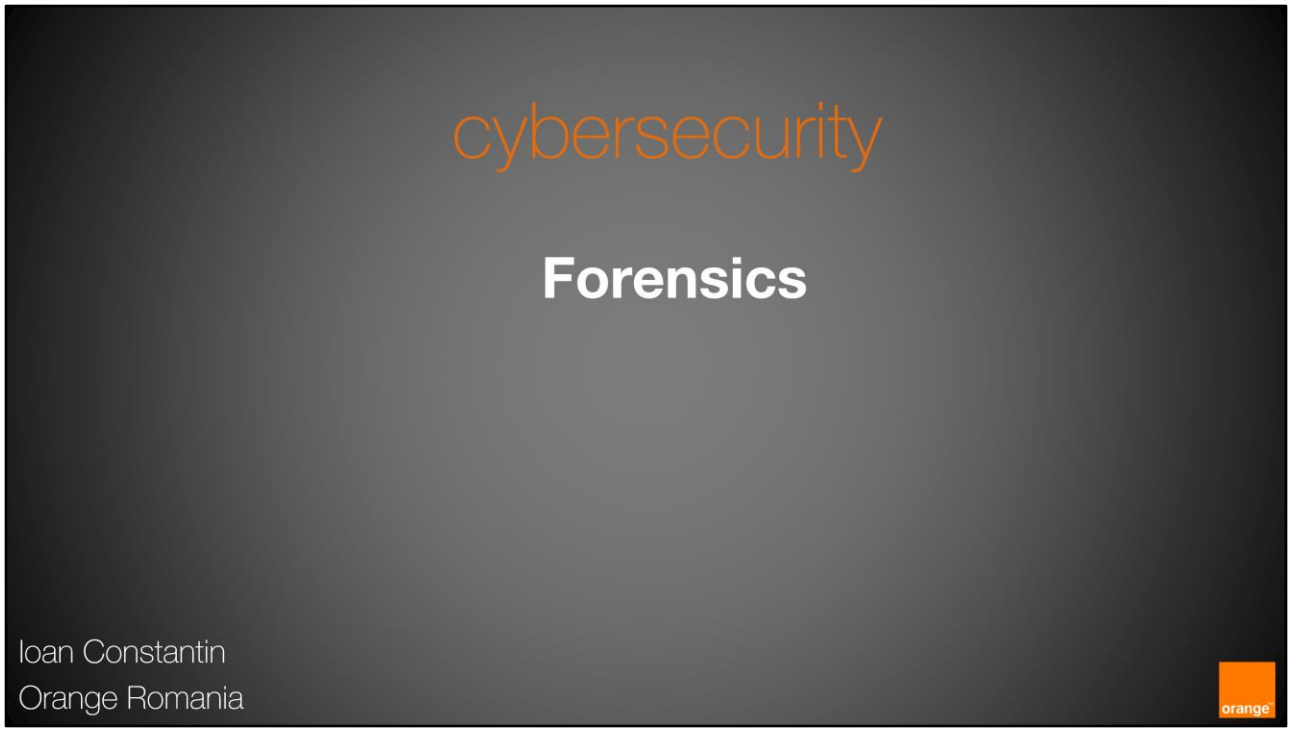


Computer forensics is the practice of collecting, analyzing and reporting on digital data in a way that is legally admissible. It can be used in the detection and prevention of crime and in any dispute where evidence is stored digitally. Computer forensics follows a similar process to other forensic disciplines, and faces similar issues.

There are few areas of crime or dispute where computer forensics cannot be applied. Law enforcement agencies have been among the earliest and heaviest users of computer forensics and consequently have often been at the forefront of developments in the field.

Computers may constitute a ‘scene of a crime’, for example with hacking or denial of service attacks or they may hold evidence in the form of emails, internet history, documents or other files relevant to crimes such as murder, kidnap, fraud and drug trafficking.

It is not just the content of emails, documents and other files which may be of interest to investigators but also the ‘metadata’ associated with those files. A computer forensic examination may reveal when a document first appeared on a computer, when it was last edited, when it was last saved or printed and which user carried out these actions.

More recently, commercial organisations have used computer forensics to their benefit in a variety of cases such as;

* Intellectual Property theft
* Industrial espionage
* Employment disputes
* Fraud investigations
* Forgeries
* Bankruptcy investigations
* Inappropriate email and internet use in the work place
* Regulatory compliance

agenda

- Concepts and definitions
  - What is computer forensics?
  - Why is it necessary?
  - Regulatory and legal stuff ☺
- Types of media for data storage
- Types of data
- Filesystems
- Metadata
- What's in a *cloud*
- Logs are your friend
- Caches, Swap Files, Temporary Storage
- Recovering information
  - What about encryption?

*Further reading in the slide notes

orange

Further reading:

http://labmice.techtarget.com/security/forensics.htm
http://symas6b5.gb-02.live-paas.net/wp-content/uploads/2017/03/ACPO-digital-evidence-v5.pdf
http://www.open.edu/openlearn/science-maths-technology/digital-forensics/content-section-0?active-tab=description-tab

Computer forensics is a branch of forensic science which deals with the application of investigative analysis techniques on computers in order to retrieve and preserve evidence in a way that is legally admissible. This means that a major aspect of the science of computer forensics lies in the ability of the forensics expert to present findings in a way that is acceptable and usable by a court of law.

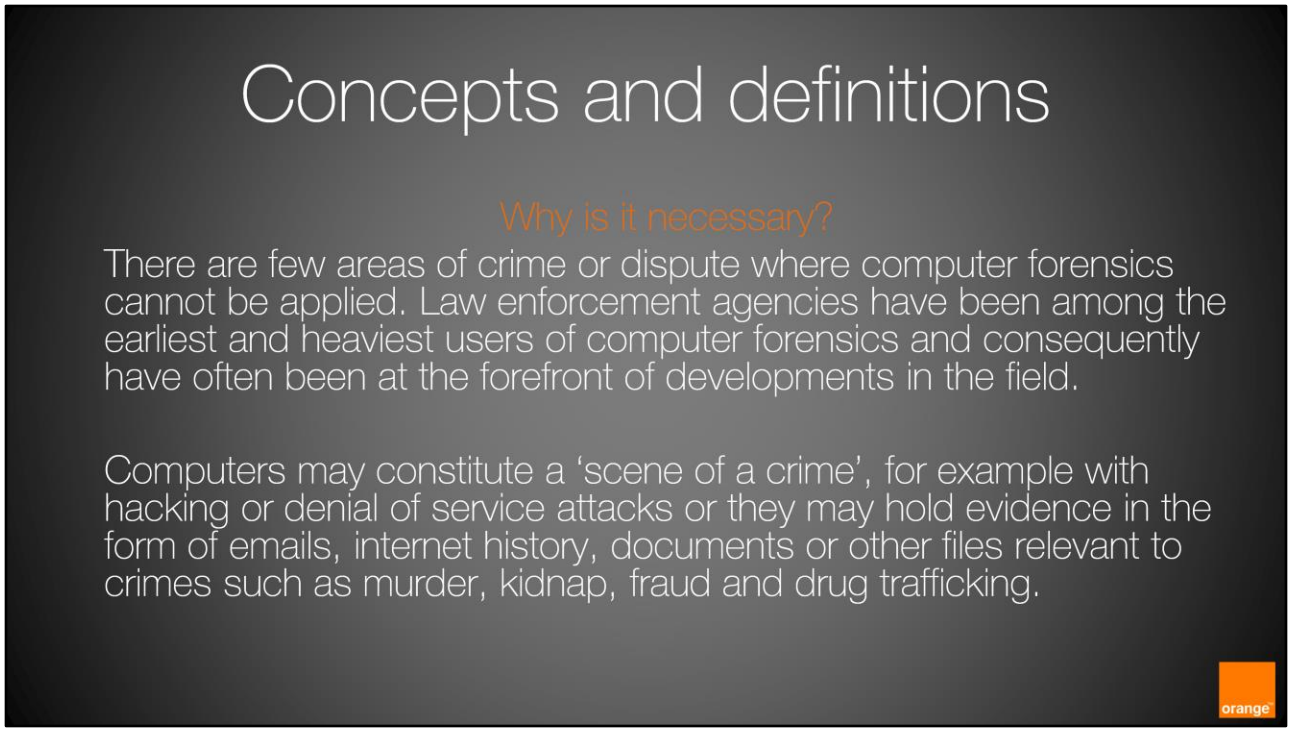
The goal of computer forensics is the performance of a structured investigation on a computing device to find out either what happened or who was responsible for what happened, while at the same time maintaining a properly documented chain of evidence in a formal report.

Computer forensics is an integral and necessary tool in the fight against cybercrime.

According to the U.S. Department of Justice (the "DoJ"), the term cybercrime refers to any illegal activity for which a computer is used as its primary means of commission, transmission, or storage and the term has rapidly gained acceptance in New Zealand. The list of criminal activities made possible by the widespread use of computers has grown exponentially in recent decades, and includes such acts as dissemination of computer viruses, network intrusion, identity theft, and even cyberbullying, stalking and terrorism.

While computer forensics may have been used traditionally by law enforcement

organizations like the police in the fight against crime, there are presently many different areas of its application, as private and commercial organizations have adopted its use for a multitude of purposes.



Computer forensic methods started to be used for collecting digital evidence for courts in the mid 1980s with the emergence and rapid growth in the use of personal computers by individuals and firms. Over the years, as the use of personal computers increased and became even more widespread, cybercrime or computer related crimes have also increased and become even more diverse.

The uses for computer forensics are varied. They range from helping law enforcement officials in the investigation of child pornography, to investigating fraud, murder, espionage, rape and cyber-stalking. In the private sector, computer forensics has been used by commercial organizations to investigate a wide range of cases including industrial espionage, fraud, intellectual property theft, forgeries, disputes with employees, regulatory compliance, bankruptcies and for the inappropriate use of a computer, Internet and email in the work place.

The discipline of computer forensics is very much concerned with the presentation of legally acceptable evidence, reports and conclusions. This has made it necessary that computer forensic investigators must follow certain rules and guidelines in order to preserve the integrity of their work. Work is not done, for example, on the physical device in question, rather after it has been physically isolated, the forensic analyst must make a digital copy of the data. To ensure that correct, Court-accepted procedures are followed, the professional investigator should be using a suite of tools such as EnCase©, which is used by Law Enforcement authorities in Romania and

internationally. This is particularly important, as the evidence discovered can, if appropriate, be handed to authorities such as NZ Police in a form with which they are completely familiar.

It is the forensic analyst's responsibility to avoid any change of data on a device that may be used as evidence in court. The audit trail created by the analyst must also be clearly understandable and a third party should be able to achieve the same results using the same processes.

As in many other professions, there are also issues that limit or adversely affect the performance of computer forensics experts. The number one hurdle a forensic analyst faces is encryption mechanisms. Although most encryption can be cracked using very powerful computers, there are still certain encryption keys that are either extremely difficult or nearly impossible to crack. In such cases, the analyst will be unable to proceed with that particular task.

Concepts and definitions

Regulatory and legal stuff ☺

Let's focus on two use cases:

a) Forensics as support for cyber security analysts
b) Forensics as support in judicial cases – search and seizure of digital information

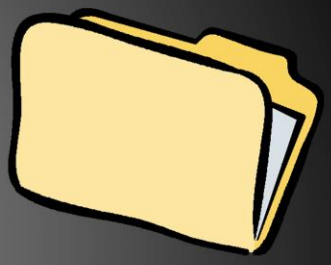
Types of data storage

By persistence	By connection type	By physical location	By consistency
Permanent* (i.e. – HDDs)	Internal Storage	Local Storage	“Hard-iron” storage
Temporary (i.e – RAM)	External Storage	Remote Storage	Virtualized Storage
	Removable Storage		

Storage media

- The tools are specific;
- The interface to the media is important as to determine the correct tools for investigation;
- Some media is prone to fail if not handled properly
- Legal requirements sometimes forbid altering the data AND media under any form, during the forensics process.

Data: The basics



Data in structured and organized in **files**.

Files are managed by a **File System**

The **File System** is a component of the **Operating System***

For each of these layers, there is a form of abstraction in place:

-The User recognizes the file per it's name and/or content;

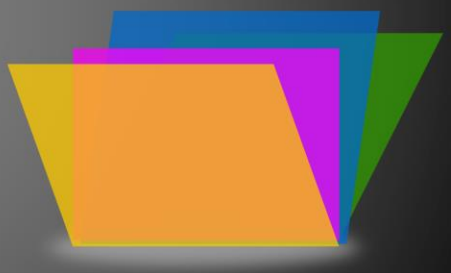
-The Operating System recognizes de file by the ID* fetched from the File System

-The File System recognizes the file by it's location in memory, start point, length and end-point

-The memory device will read or write the/to the file as commanded by the storage controller. The memory device is structure-agnostic.

Data: File Structures

The File System: Controls how data is stored and retrieved.
A file system usually consists of three layers:
Logical Layer: interacts with the OS and apps via APIs
Physical Layer: this manages the physical operation of the storage device(s). It tells memory (disks, ROM, RAM etc.) to write and read blocks, it handles buffering and memory management
Virtual Layer: This allows multiple concurrent instances of Physical Layers to be connected to the same Logical Layer



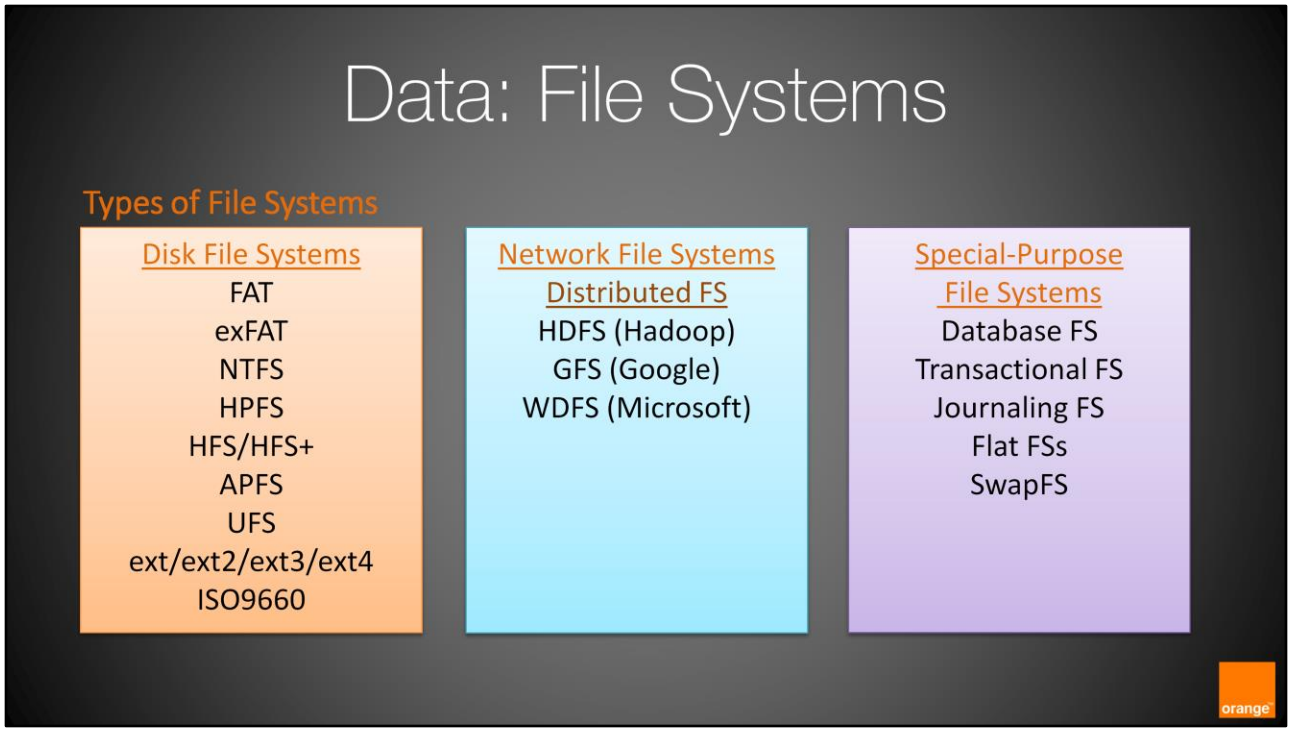
Data: File Formats

A file format is a standard way that information is encoded for storage in a computer file. It specifies how bits are used to encode information in a digital storage medium. File formats may be either proprietary or free and may be either unpublished or open.

Most open and free file formats have their specifications published

Some proprietary or unpublished file formats do not. This is where reverse engineering comes in handy in forensics.

orange



A disk file system takes advantages of the ability of disk storage media to randomly address data in a short amount of time. Additional considerations include the speed of accessing data following that initially requested and the anticipation that the following data may also be requested. This permits multiple users (or processes) access to various data on the disk without regard to the sequential location of the data. Examples include FAT (FAT12, FAT16, FAT32), exFAT, NTFS, HFS and HFS+, HPFS, APFS, UFS, ext2, ext3, ext4, XFS, btrfs, ISO 9660, Files-11, Veritas File System, VMFS, ZFS, ReiserFS and UDF. Some disk file systems are journaling file systems or versioning file systems.

Network (Distributed) file systems do not share block level access to the same storage but use a network protocol. These are commonly known as network file systems, even though they are not the only file systems that use the network to send data .Distributed file systems can restrict access to the file system depending on access lists or capabilities on both the servers and the clients, depending on how the protocol is designed.

The difference between a distributed file system and a distributed data store is that a distributed file system allows files to be accessed using the same interfaces and semantics as local files – for example, mounting/unmounting, listing directories, read/write at byte boundaries, system's native permission model. Distributed data

stores, by contrast, require using a different API or library and have different semantics (most often those of a database).
A distributed file system may also be created by software implementing IBM's Distributed Data Management Architecture (DDM), in which programs running on one computer use local interfaces and semantics to create, manage and access files located on other networked computers. All such client requests are trapped and converted to equivalent messages defined by the DDM. Using protocols also defined by the DDM, these messages are transmitted to the specified remote computer on which a DDM server program interprets the messages and uses the file system interfaces of that computer to locate and interact with the specified file.

Special Purpose FS:
Another concept for file management is the idea of a database-based file system. Instead of, or in addition to, hierarchical structured management, files are identified by their characteristics, like type of file, topic, author, or similar rich metadata.

Transaction processing introduces the isolation guarantee[clarification needed], which states that operations within a transaction are hidden from other threads on the system until the transaction commits, and that interfering operations on the system will be properly serialized with the transaction. Transactions also provide the atomicity guarantee, ensuring that operations inside of a transaction are either all committed or the transaction can be aborted and the system discards all of its partial results. This means that if there is a crash or power failure, after recovery, the stored state will be consistent. Either the software will be completely installed or the failed installation will be completely rolled back, but an unusable partial install will not be left on the system.

Journaling file systems are one technique used to introduce transaction-level consistency to file system structures. Journal transactions are not exposed to programs as part of the OS API; they are only used internally to ensure consistency at the granularity of a single system call.

In a flat file system, there are no subdirectories; directory entries for all files are stored in a single directory.

Data: File Formats

Identifying file types: by file extension / by metadata

File Extensions: it may be considered metadata as it implyies information about how data might be stored in the file.

DOS and Windows use separate namespaces for extensions

UNIX and derived use the same namespace for the entire filename, including extension(s).

File Extensions: most Operating Systems have special requirements for interpreting files as executables, reg. to file extensions such as DOS (.exe, .com) while others do not use extensions for executables at all (such as UNIX and derived – MacOS, Linux)

Metadata

- Is data that provides information about other data
- Information stored in the files itself
- It does not require a special namespace
- Most metadata formats are open and/or made public by the developers
- Metadata is usually stored in the beginning of the file for easy identification
- The part of the file that contains metadata is usually called the file header

What information can it provide? File size, file type, associated programs with the filetype, author's name, file creation date, file modification date, previews (excerpts) of the content etc.

How is it relevant to forensics? – **identification, classification and indexing of assets**

orange

Metadata is data that describes other data. Meta is a prefix that in most information technology usages means "an underlying definition or description."

Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier. For example, author, date created and date modified and file size are examples of very basic document metadata. Having the ability to filter through that metadata makes it much easier for someone to locate a specific document.

In addition to document files, metadata is used for images, videos, spreadsheets and web pages. The use of metadata on web pages can be very important. Metadata for web pages contain descriptions of the page's contents, as well as keywords linked to the content. These are usually expressed in the form of metatags. The metadata containing the web page's description and summary is often displayed in search results by search engines, making its accuracy and details very important since it can determine whether a user decides to visit the site or not. Metatags are often evaluated by search engines to help decide a web page's relevance, and were used as the key factor in determining position in a search until the late 1990s. The increase in search engine optimization (SEO) towards the end of the 1990s led to many websites "keyword stuffing" their metadata to trick search engines, making their websites seem more relevant than others. Since then search engines have reduced their reliance on metatags, though they are still factored in when indexing pages. Many

search engines also try to halt web pages' ability to thwart their system by regularly changing their criteria for rankings, with Google being notorious for frequently changing their highly-undisclosed ranking algorithms.

Metadata can be created manually, or by automated information processing. Manual creation tends to be more accurate, allowing the user to input any information they feel is relevant or needed to help describe the file. Automated metadata creation can be much more elementary, usually only displaying information such as file size, file extension, when the file was created and who created the file.

Metadata (2)

Types of Metadata:

- File System Metadata: Access Control Lists, Modified /Access/Creation Date, User Name Authors etc.
- Digital media metadata (Image, Video, Sound)
- Document Metadata (Last Print Time, Number of Pages, Distribution Lists etc)

Common types and examples:

- EXIF (JPEG, TIFF,): stores information about the camera make and model used to take the picture, the geolocation of the camera at that time, the time and date, resolution, color mode etc.
- ID3 (MPEG 2-Layer 3):
- MSOLE Compund File: stores a large amount of metadata about MS Office files, is stored in the property set streams of each file (doc, xls, ppt etc)

Cloud Data

Challenges:

- You don't have physical access to media
- You don't own the filesystem or the operating system
- Abstraction imposes a large overhead
- Access is restricted to one ore more users, you're not one of them

Cloud Data (2)

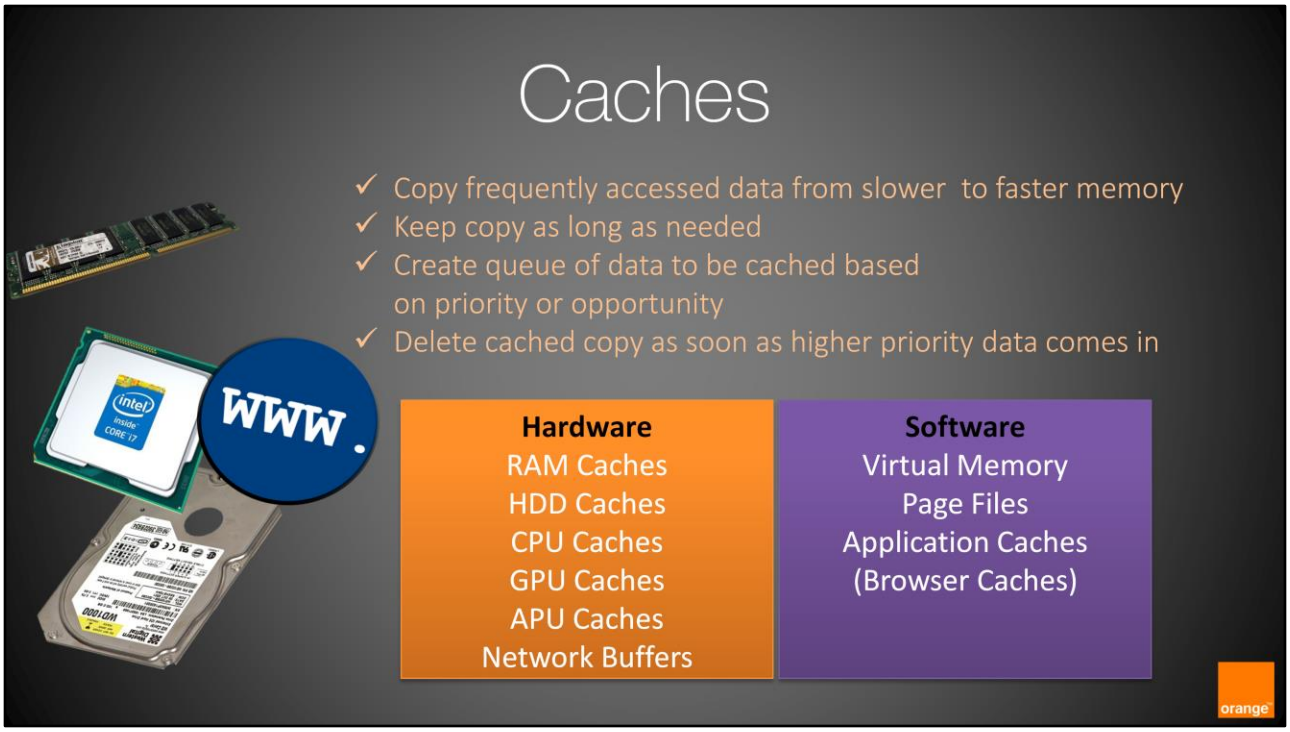
Solutions:

- You rely on the client / web browser.
- You should investigate logs
- You should dump caches / swap partitions
- You should investigate backups

Logs

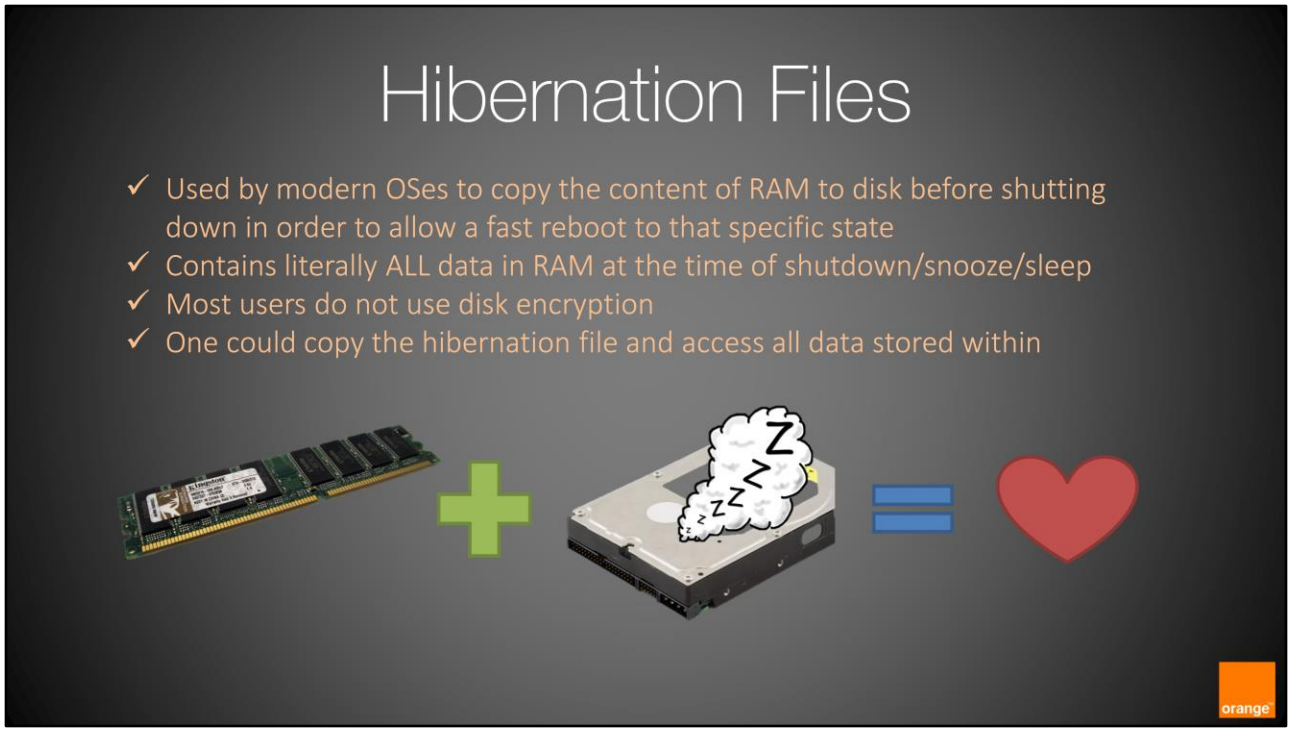
Benefits:

- Every device, operating system and application generates a huge amount of data in the form of metrics and logs.
- Logs are great resources for figuring out **who** did **what** and **when**?
- Logs are usually stored for long periods of time



Caches can hold a variety of important data, albeit in small amounts as opposed to primary storage.
Hardware caches are usually limited in size, up to several Gigabytes (SLC Caches in consumer-grade SSDs, 'Vapor' Cache in disk controllers etc.)
Most CPU/GPU/APU Caches will only hold data snippets and instructions. Most of the information held in these will not be readily important from a forensics standpoint but can be useful for providing context for other data.
Software caches like Virtual Memory, Page Files, Hibernation Files or Application Caches will usually hold a larger amount of data compared to hardware caches. Most OSes will compress these files.

Both Linux and Microsoft Windows systems expand RAM by using disk. In this virtual memory model, the OS moves data in memory to a special location on disk in order to free RAM for additional operations. When the data on disk is needed again, it's moved back into RAM. The area on disk used for this purpose is called the *swap file* or *swap space*. In Linux environments, the swap area is an actual disk partition. On a Windows XP machine, the swap space is a file called Pagefile.sys.
Since everything in RAM is subject to being swapped to disk, some very interesting information can be found in a swap file. In addition to plain-text data that might be encrypted in a disk file, encryption keys might also be present. This is due to weaknesses in some applications that allow unencrypted keys to reside in memory. Further, information contained in e-mails or stored at remote locations might still reside in swap space. Any standard disk maintenance utility can access this information.



Hibernation (or suspend to disk) in computing is powering down a computer while retaining its state. Upon hibernation, the computer saves the contents of its random access memory (RAM) to a hard disk or other non-volatile storage. Upon resumption, the computer is exactly as it was before entering hibernation.
hiberfil.sys is the file used by default by Microsoft Windows to save the machine's state as part of the hibernation process. The operating system also keeps an open file handle to this file, so no user, including the Administrator, can read the file while the system is running.
Although often presumed, the size of the hiberfil.sys is not one-to-one in size to the available, or total RAM of the machine.
The data structures required to parse the file format are available in the Microsoft Windows debug symbols, including some of the various compression methods used.

Hibernation Files

```
user@ubuntu:# strings _pagefile.sys | grep "http://" | more
http://pki.google.com/GIAG2.crt0+
http://clients1.google.com/ocsp0
http://pki.google.com/GIAG2.crl0
)http://crl.geotrust.com/crls/gtglobal.crl0=
!http://gtglobal-ocsp.geotrust.com0
)http://crl.geotrust.com/crls/secureca.crl0N
http://pki.google.com/GIAG2.crt0+
http://clients1.google.com/ocsp0
http://pki.google.com/GIAG2.crl0
)http://crl.geotrust.com/crls/gtglobal.crl0=
!http://gtglobal-ocsp.geotrust.com0
)http://crl.ge
http://pki.google.com/GIAG2.crt0+
http://clients1.google.com/ocsp0
http://pki.google.com/GIAG2.crl0
)http://crl.geotrust.com/crls/gtglobal.crl0=
!http://gtglobal-ocsp.geotrust.com0
)http://crl.geotrust.com/crls/secureca.crl0N
http://pki.google.com/GIAG2.crt0+
...snip...
```

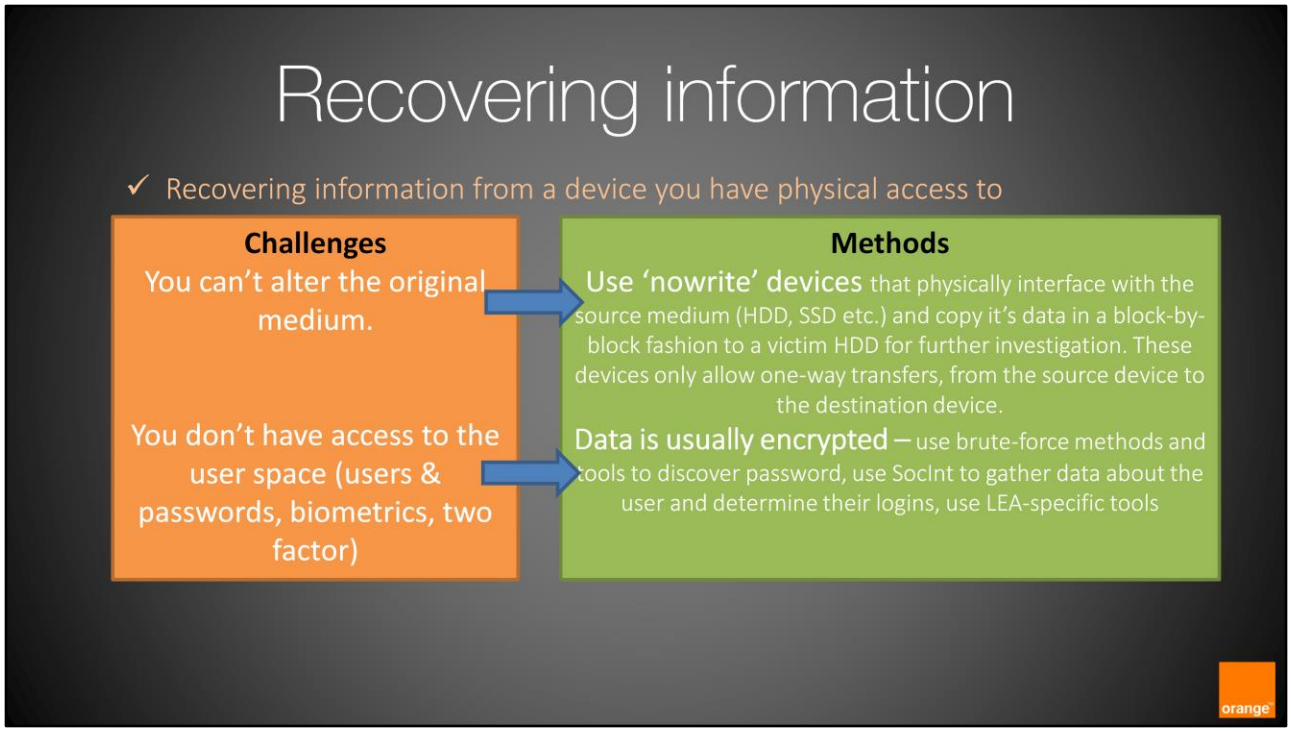
```
user@ubuntu:# pagebrute -r special.yara -f _pagefile.sys
[+] - PAGE_BRUTE processing file: _pagefile.sys
[+] - YARA rule of File type provided for compilation: special.yara
..... Ruleset Compilation Successful.
[+] - PAGE_BRUTE running with the following options:
     [-] - FILE: _pagefile.sys
     [-] - PAGE_SIZE: 4096
     [-] - RULES TYPE: FILE
     [-] - RULE LOCATION: special.yara
     [-] - INVERSION SCAN: False
     [-] - WORKING DIR: PAGE_BRUTE-2016-04-26-16-06-30-RESULTS
     ================

        [!] FLAGGED BLOCK 9243: special
        [!] FLAGGED BLOCK 27881: special
        [!] FLAGGED BLOCK 28009: special
        [!] FLAGGED BLOCK 28011: special
        [!] FLAGGED BLOCK 28475: special
        [!] FLAGGED BLOCK 32190: special
        [!] FLAGGED BLOCK 32214: special
        [!] FLAGGED BLOCK 32222: special
        [!] FLAGGED BLOCK 32242: special
...snip...
```

Various tools such as Volatility can be used to process page files and hibernation files. One important difference between the two is that the hibernation file is compressed. One has to extract the raw memory dump from the compressed file(s) and search its content for clear-text data.

Recovering information

- ✓ We now know what are the most common data types and storage medium
- ✓ The challenge for a cyber forensic investigator is, usually, to recover information that was somehow made unavailable by a user (trough deletion, encryption, by moving it in cloud storage etc.)
- ✓ We'll have a look at the most common scenario:
  - ✓ Recovering data from the device/storage medium (i.e – with physical access to the device)
- ✓ We'll discuss the challenges, the tools and the methods for each scenario



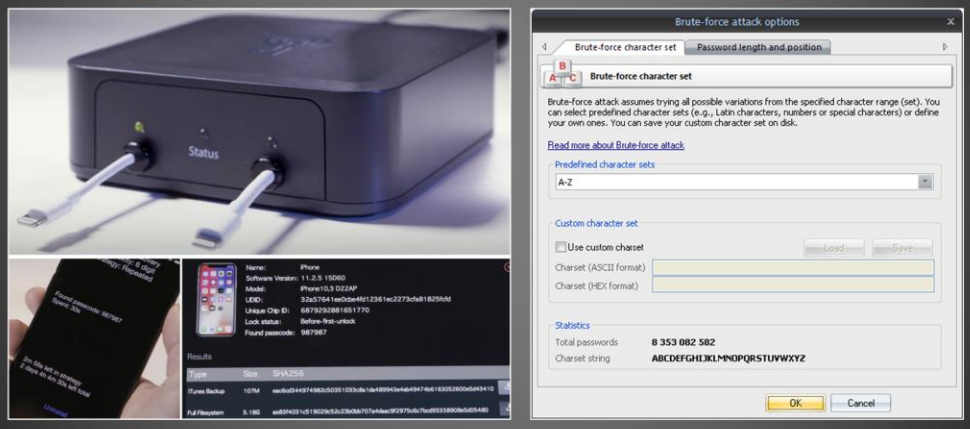
While there are several challenges one can encounter while trying to run forensics on device or storage medium, two of the most common and important challenges are:
-the need to copy all available data without altering the original medium (due to regulatory, legal and chain of custody requirements) and
-the need to access the user space on that device and/or decrypt data

For the first challenge, one can use 'one-way copy' or 'no write' devices that allow an investigator to image the processed storage medium without ever writing to it. Such devices are usually in the form of hardware boxes that interface both source and destination medium and copy the data from the source by the lowest logical unit such as blocks for HDDs/SSDs/Optical Drives. The investigator can further inspect the data on the 'victim' –destination drive and apply specific tools for data recovery and investigation (such as password crackers, decryptors, etc.) without having to worry about the original source being altered.

The second challenge usually requires the investigator to use very specific, limited distribution tools that are available only to LEAs. Such tools will use some sort of vulnerability, be it in software or firmware of the target device to gain access to its resources.

Recovering information

- ✓ Recovering information from a device you have physical access to

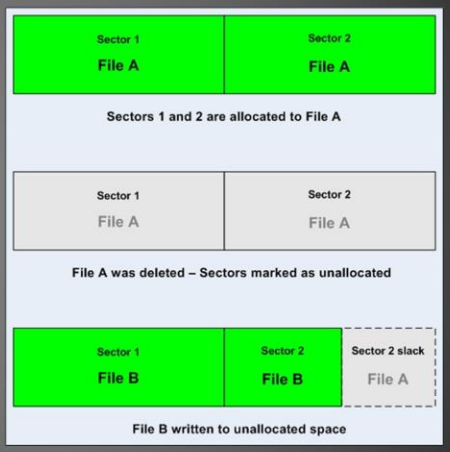


Bruteforce boxes leverages vulnerabilities in mobile OSes allowing an investigator to try multiple combinations of passcodes until they stumble upon the correct one.
Such boxes are sold exclusively to LEAs (Law Enforcement Agencies) and there is a ongoing dispute on whether such use is legal, even by LEAs.
Such devices exists for HDD/SSD decryption. The distribution is controlled and limited to LEAs and most 'boxes' are meant to be used with just one model and make from a single vendor as they leverage specific vulnerabilities, either zero-day or disclosed that affect only that product.

Most password crackers / bruteforce attacks are widely available as free software. These programs will work with earlier versions of Microsoft Windows and various MacOS and Linux distributions but will fail with newer versions, such as Windows 10 and MacOS 10(or newer) as it stores user credentials encrypted and it limits the number of password entry retries.

Recovering information

- ✓ What happens when you delete a file?
- ✓ The vast majority of File System implementations will only mark the storage space (unit) occupied by that file as available, without actually overwriting it with data.
- ✓ Most "Data Recovery" tools will work if used immediately after deletion
- ✓ Some will be able to recover only parts of the deleted files



orange

Fortunately for the forensics investigator, most users aren't very good at covering their tracks. Ignorance of how computers manage memory and disks results in incriminating file or memory content stored in various locations invisible to the subject of an investigation. In this post, we'll look at three potential locations for this information -- deleted files and slack space, swap space, and hibernation files.

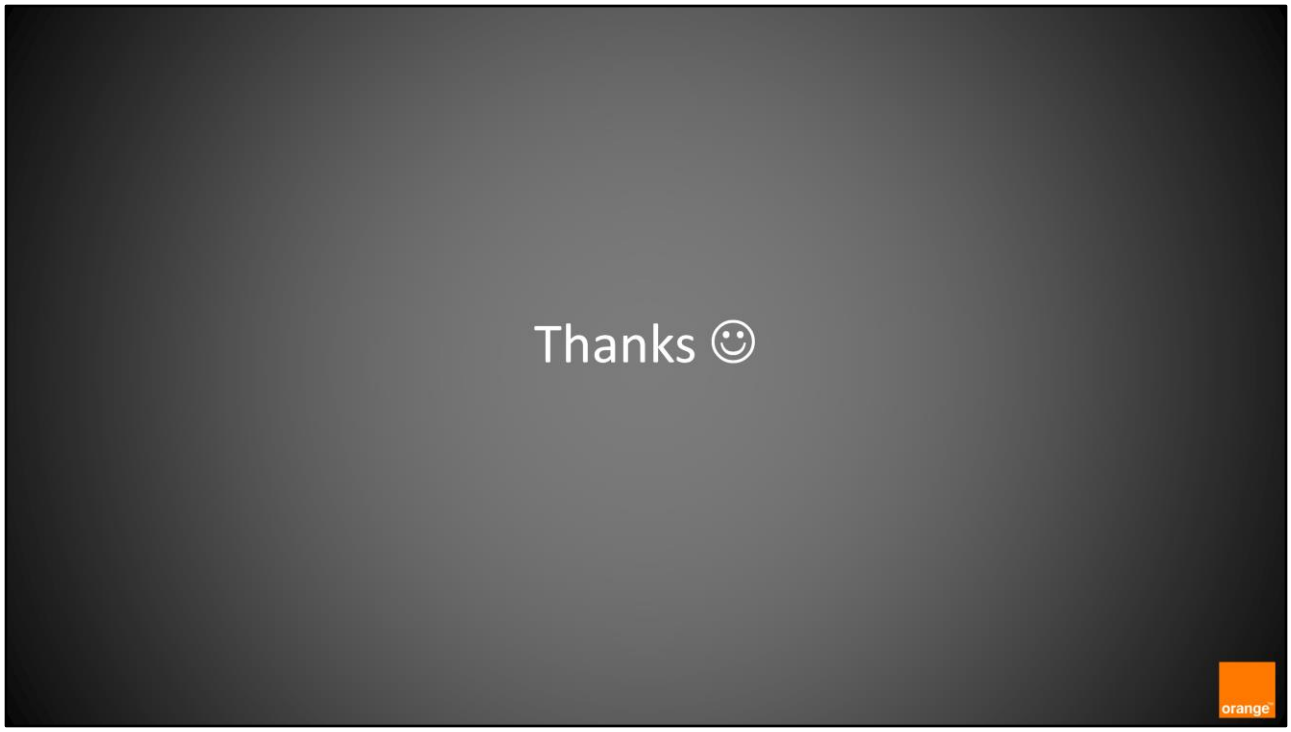
Deleted files and slack space When an operating system writes a file to disk, it allocates a certain number of sectors. The number of sectors allocated depends on the limitations of the operating system and configuration decisions made by the system administrator. The sectors allocated and their location on the disk are recorded in a directory table for later access. When the file is deleted, the space originally allocated to it is simply marked as unallocated. The actual data remains on the disk. Deleted files in this state are easily recoverable by many disk utilities, but what happens if a new file is written to this same space? **Figure A** shows what might happen to the original data.

At some point in the past, File A was written to sectors 1 and 2. The sectors were completely filled by the file's content. When the user decides to delete the file, the sectors are marked as unallocated. However, the file content remains.

Sometime after File A is deleted, the user requests the OS to save File B. The OS once again allocates sectors 1 and 2, but notice that the file content doesn't completely fill sector 2. The unwritten portion of sector 2 is known as *slack space*, and it still contains content from File A. Slack space data can be read and analyzed by any of the popular forensics toolkits.

Quick Recap

1. What is computer forensics
2. What are the principal types of media used to store data?
3. What are the most common File Systems?
4. What 's Metadata?
5. What are Logs? Why are they useful?
6. What are the challenges in recovering data?



Got questions? Drop me a line: ioan.constantin@orange.com