universidade
de aveiro

Computer Systems Forensic Analysis
AFSC

Introduction to digital forensics

Artur Varanda

School Year 2023-2024

Digital investigation focus:

- digital devices that has been involved in an incident or crime
- device used to:
  - ✓ commit a physical crime – *e. g.* a suspect used the Internet to conduct research about a physical crime
  - ✓ execute a digital event that violates a policy or law – *e. g.* an attacker gains unauthorized access to a computer, a user downloads contraband material, or a user sends a threatening e-mail, *etc*;
- When the violation is detected, an investigation is started to answer:
  - ✓ what, who, when, how
  - ✓ in some cases “where” and “why”

A digital investigation is

- a process where we develop and test hypotheses that answer questions about digital events
  - ✓ a scientific method
  - ✓ where we develop a hypothesis using evidence that we find
  - ✓ and then test the hypothesis by looking for additional evidence that shows the hypothesis is impossible

Digital evidence

Is a digital object that contains **reliable** information that supports or refutes a hypothesis. The digital evidence must be:

- admissible, authentic, accurate and complete

Digital evidence is:

- information stored or transmitted in digital formats or media, the content of which is evidence, whether material or merely indicative, of a particular incident or event;
- It is fragile and volatile, so the attention of a certified expert is required in order to ensure that the data of probative value are effectively isolated and extracted correctly and lawfully.

Digital evidence challenges:

- **hard to control** – it is very easy to create, modify, transmit or delete data in short amount of time
- **diversity and complexity** – some times is hard to identify the digital evidences because information systems evolve too fast

Forensic means

it has legal requirements to be accepted into a court of law

Note:

A **digital forensic investigation** is a more restricted form of digital investigation

Definition by Brian Carrier

Process that uses science and technology to analyze digital objects and that develops and tests theories, which **can be entered into a court of law**, to answer questions about events that occurred.

Another definition

The systematic and technological inspection of a computer system and its contents in order to obtain evidence of a crime or any other use that is under investigation.

Types of analysis to find evidences:

- *live analysis* – when the operating system or other resources of the system being investigated is used to find evidence
  - ✓ advantages: get data from RAM of a running process
  - ✓ disadvantages: risk of getting false information because the software could maliciously hide or falsify data
- *post-mortem analysis* – when trusted applications in a trusted operating system are used to find evidence (lab environment)
  - ✓ advantages: fully controlled environment
  - ✓ disadvantages: information from RAM is lost, *e. g.* key to decrypt a file, ...

A post-mortem analysis is more ideal, but not always possible.

A server has been compromised, how it occurred and who did it?

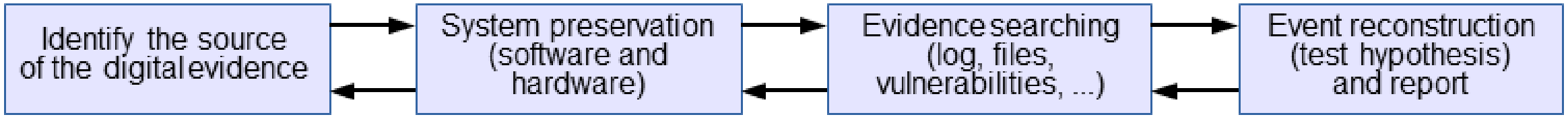
- find data that were created by events related to the incident recover deleted log entries from the server
- find attack tools
- find the vulnerabilities that existed on the server
- using this data, and more, we develop an hypotheses
  - ✓ which vulnerability the attacker used to gain access
  - ✓ what he/she did afterwards
- later, examine the firewall configuration and logs
  - ✓ determine that some of the scenarios in our hypotheses are impossible because that type of network traffic could not have existed
  - ✓ evidence was found that refutes one or more hypotheses

Digital Crime Scene Investigation Methodology

Investigation process

- There is no single way to conduct an investigation
- It does not matter which process is used,
  - As long as we find the right person and do not break any laws in the process
- However, some are more efficient than others

The four major phases – based on the physical crime scene investigation process



1 – **Preparation**

- physically identifying the origin of the digital evidence
- choose the best approach to analyze it
- equipment seizure

2 – System Preservation

- goals
  - ✓ preserve the state of the digital crime scene
  - ✓ reduce the amount of evidence that may be lost
- actions vary depending on the legal, business, or operational requirements of the investigation
  - ✓ legal requirements may cause you to unplug the system and make a full copy of all data or,
  - ✓ could be a case involving a spyware infection or a honeypot and no preservation is performed
  - ✓ if it's not going to court, techniques in between can be used

Preservation Techniques

post-mortem analysis

- ✓ pull the plug to reduce the amount of evidence that is overwritten
- ✓ make duplicate copies of all data
- ✓ use write blockers to prevent evidence from being overwritten

live analysis

- ✓ kill or suspend suspect processes unplug or limit network connection
  - ✓ use an empty hub or switch to prevent log messages about a dead link
  - ✓ use network filters to avoid a remote connection from perpetrator to delete data
- ✓ backup important data (logs, files, *etc*)

Data integrity

- when important data are saved during a *post-mortem* or live analysis, a cryptographic hash should be calculated to later show that the data have not changed

Cryptographic hash algorithms

Data integrity – MD5 cryptographic hash

- this algorithm is broken since 2004
- use only for retro compatibility purposes
- it is possible to create collisions – different files with the same hash value examples:

  http://www.mscs.dal.ca/~selinger/md5collision/

Demonstration

Data integrity – Hash values by itself are not enough

- given a message *M*, its hash value is $H(M) = h$
- someone can change both *M* and *h*, because *h* doesn't depend on a secret

Possible solution:

Digital Signatures

- depends on a private key
- better if done with a secure device,

e. g. the Portuguese Citizen Card (Cartão de Cidadão)

3 – Evidence searching

- goal: find data that support or refute hypotheses about an incident
- typically starts with a survey of common locations based on the type of incident:
  - ✓ Web-browsing habits: look at the Web browser cache, history file, and bookmarks
  - ✓ Linux intrusion: look for signs of a rootkit or new user accounts

It is important to look also for evidence that **refutes** your hypothesis instead of only looking for evidence that only supports your hypothesis.

The searching process:

1. define the general characteristics of the object for which we are searching
2. look for that object in a collection of data
3. two key steps:
   - determining for what we are looking
   - where we expect to find it

Example:

search all files with pictures

Search techniques:

- most searching for evidence is done in a file system and inside files

- search for files based on:
  - ✓ their names, or patterns in their names
  - ✓ a keyword in their content
  - ✓ temporal data, such as the last accessed or written time
  - ✓ hash values and compare them against a database
    - allows to find all files of a given type even if someone has changed their name
    - National Software Reference Library (NSRL) database https://www.nsrl.nist.gov

- analyzing network data based on:
  - ✓ packet headers, such as IP addresses, port number, protocol, . . .
  - ✓ keywords inside packets content

4 – Event reconstruction and report

- goal: try to answer questions about digital events in the system
- during the Evidence Searching Phase we might find several files that violate a law

    but it doesn't answer questions about events

    the file may have been the effect of an event, but what application downloaded it?

        a web browser?

        a malicious software? — **several cases have used malware as a defense**

it may be possible to correlate the digital events with physical events

Event reconstruction requires knowledge about the applications and the OS that are installed on the system so that you can create hypotheses based on their capabilities

Examples:

- different versions of Windows OS (XP, 7, 8, 10) can cause different events
- different versions of Firefox, or Chrome Web browsers can cause different events

Exercises

Automate comparison of hash values

1. calculate the SHA256 values of all files inside a directory, *e. g.* `C:\Windows` or `/etc` and store the result in a file:

```
sha256sum * > SHA256.txt    # works on Linux
```

2. verify the values:

```
sha256sum -c SHA256.txt                # works on Linux
```

Tip for Windows OS:

hash calculation tool https://www.slavasoft.com/hashcalc/

Crack hash values

1. calculate the SHA256 of a common word, *e. g.* "Aveiro":

```
echo -n "Aveiro" | sha256sum    # this is a Linux command
```

2. copy the hash value and paste it to crackstation.net or hashes.com

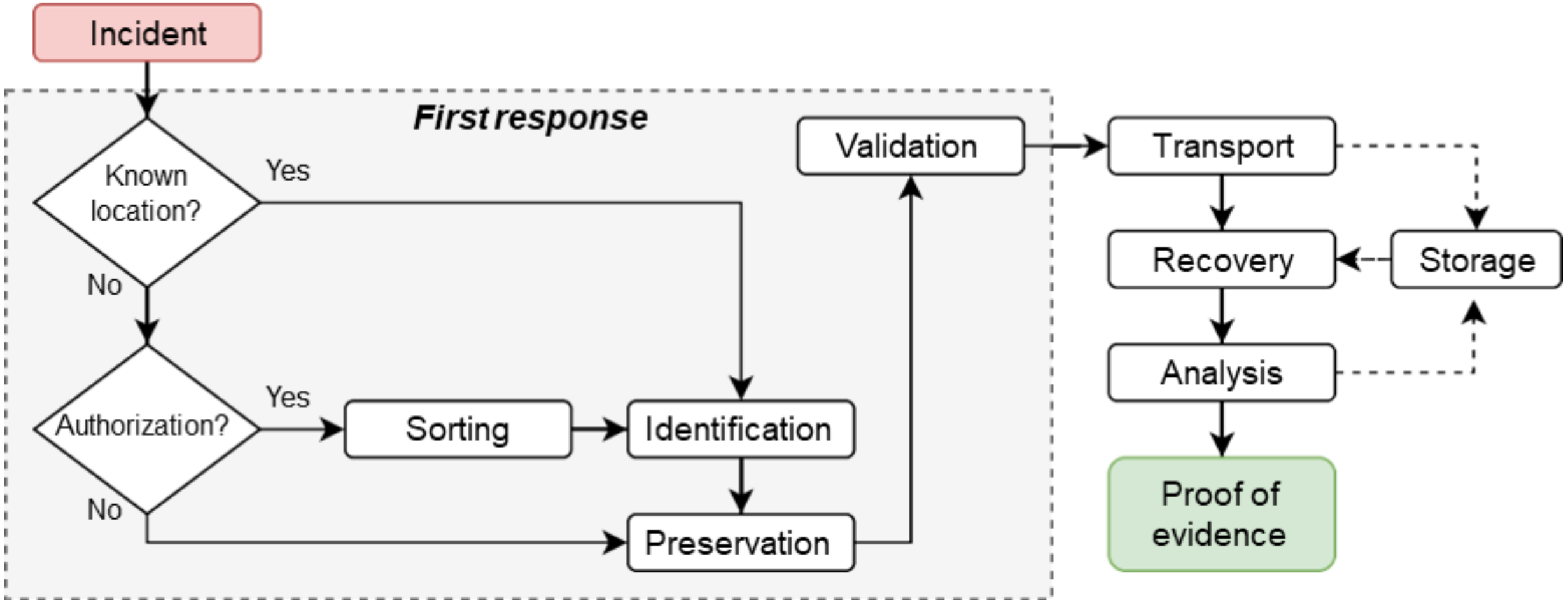
was the site able to find your word?

Digital Evidence Handling

Principles By EU-OLAF recommendation

1. The actions triggered by the first responders should not alter the data maintained on a computer or in a storage device that may be submitted to a court as evidence;
2. In exceptional circumstances, if it is considered necessary to access the original data stored on a computer or in a storage device, those who do so must have skills to be able to provide evidence, explaining the relevance and implications of his actions;
3. A chain of custody, or other record of all processes applied to digital evidence must be created and preserved. An independent third party should be able to examine these processes and obtain the same result;
4. The person responsible for the investigation must assume overall responsibility for compliance with the law and the present principles.

ROUTE OF THE DIGITAL EVIDENCE



1 – Identification

Data states

- stored – data permanently stored in a storage device, *e. g.* an hard drive;
- in transit – data being sent through a local network, or Internet, to a reception device;
- in reception – data being received in a device, but not yet available to the user;
- in creation – data being locally produced and only partially available to the user;

Data sources

- storage devices – hard drives, SSD, USB drive, tapes, . . .
- temporary location – RAM, page files, swap partitions, cache files, . . .
- peripheral devices – printers, plotters, memory card readers, . . .
- active network devices – switches, routers, modems, print servers, . . .

1 – Identification

logical and physical location of the data

- local or remote devices
- dedicated storage systems (usually on data centers)
- computational systems (*e. g.* PC, laptop) has at least one storage device

main data types

- simple and human readable, *e. g.* photos, text documents, spread sheets
- complex and/or structured data, *e. g.* data bases, file system
- raw data, streams of data

2 – Preservation

do not modify any data that could have been evidence

- Copy important data, put the original in a safe place, and analyze the copy so that you can restore the original if the data is modified;
- Calculate hash values of important data so that you can later prove that the data has not changed – better yet if you do a digital signature;
- Use a write-blocking device during procedures that could write to the suspect data;
- Minimize the number of files created during a live analysis because they could overwrite evidence in unallocated space;
- Be careful when opening files on a live analysis because you could be modifying data, such as the last access time;

2 – Preservation: Prioritize the evidence to be collected

Order of volatility:

1. CPU, cache, and register content;
2. routing table, ARP cache, process table, kernel statistics
3. RAM
4. temporary file system, swap space
5. data on local storage media
6. remotely logged data
7. data contained on archival media (backups)

2 – Preservation: Sorting

Some times it is not possible to bring all devices due to several constrains: legal, time, technically unfeasible, . . .

Data sorting

In those situations a pre-analysis is required, but can only be performed if authorized accordingly to the country's laws.

3 – Isolate

isolate yourself from the suspect data because you do not know what it might do

- an executable from the suspect system could delete all files on your computer, or it could communicate with a remote system;
- opening an HTML file from the suspect system could cause your Web browser to execute scripts and download files from a remote server;

Create an isolated environment

- use virtual machines (VMware, VirtualBox, Xen, . . .)
- use an analysis network that is not connected to the outside world, or that is connected using a firewall that allows only limited connectivity
- isolation is very difficult, or impossible, with live analysis

4 – Correlate

correlate data with other independent sources

- helps reduce the risk of forged data
- timestamps can be easily changed in most systems
- find log entries, network traffic, or other events that can confirm the file activity times

This task is time consuming, specially if done without the help of software

5 – Log

log and document **all** of your actions

- helps identify what you have already done and what your results
- helps identify what searches you have not done yet
- in a live analysis, or performing techniques that will modify data, it is important to document what you do so that you can later document what changes in the system were made because of your actions

5 – Log: Identify devices:

- create tags to uniquely identify devices *e. g.* PC01, PC01.1, PC01.2, . . .
- take photographs
  - ✓ after placing tags
  - ✓ should be easy to read any relevant information, if needed take an overview photo and then a close up shot
    - computer brand, model, serial number, . . .
    - network cable connections, etc

Examples

Are there any problems with these photos?

5 – **Log**: Templates

Create templates with the required info to identify devices and services

Tag ID	PC01.HD03
Device	Hard disc drive, 2.5"
Brand	Seagate
Model	Momentus 5400.6 ST9250315AS
Serial number	5VC9CWTT
Interface	SATA
Capacity	250,0 GB
Type of intervention	Forensic copy
Working condition	Normal
Pictures	Yes, see Fig. 1 and 2
Observations	None

Tag ID	DNS01
Domain name	`lotreur.com`
Type of information	DNS history
Registrar	Center of Ukrainian Internet Names (UKR-NAMES)
Creation date	01-Mar-2013
Current state	expired
Registrant email	`c152136@rmqkr.net`
Annexes	Yes, see Annex A

5 – Log: Reception and tagging

- verify the list of devices delivered for analysis
- tag the devices, taking into account that one process may have:
  - ✓ one suspect with several devices,
  - ✓ or several suspects and many devices
  - ✓ there are processes with 30+ suspects and 200+ devices!

5 – Log: Tagging rules (as an example)

- letters to identify the owners of the devices in a given process
  - ✓ `A`, `B`, `C`, …
- set of acronyms for each device type, followed by a 2 digits number
  - ✓ `FPHxx`, `SPHxx`, `SIMxx`, `MCxx`, `GPSxx`, `CAMxx`, `PCxx`, `LAPxx`, `HDDxx`, `SSDxx`, `PENxx`, …
- a tag is composed by

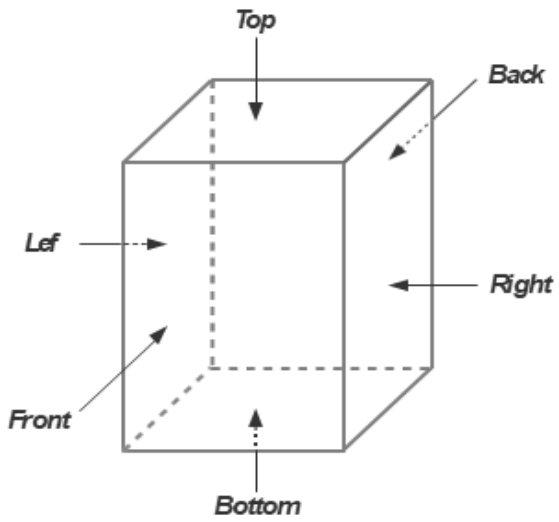
`ownerID.deviceID[.inside deviceID]`

example of a samartphone with 2 SIM cards and a memory card:

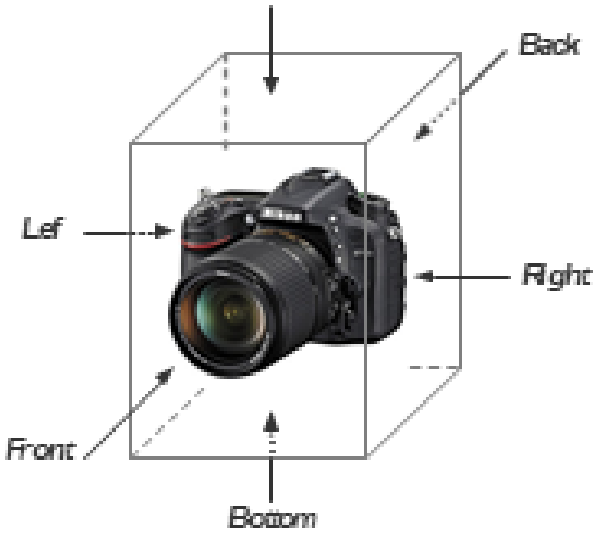
- ✓ `A.SPH01` – the handset from owner `A`
- ✓ `A.SPH01.SIM01` – first SIM card
- ✓ `A.SPH01.SIM02` – second SIM card
- ✓ `A.SPH01.MC01` – memory card

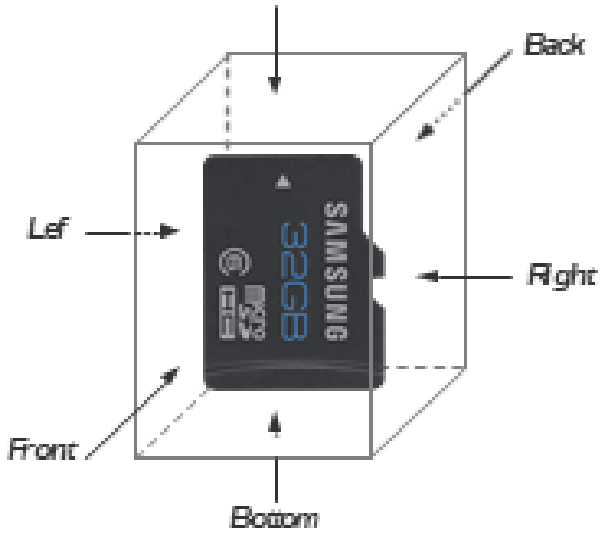
5 – Log: photograph rules (as an example)

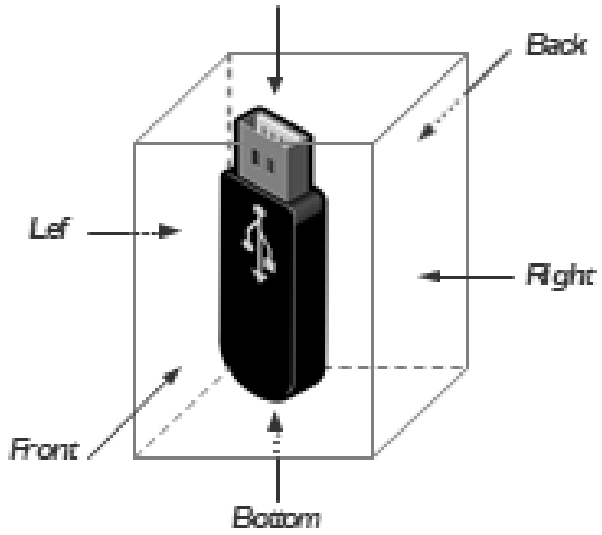
- **include the scale of a ruler** in the photo
- photograph all important views
  - there are mandatory views by device type and some optional
  - attention to details, like serial numbers, IMEI, *etc*
- apply the views names accordingly to the 3D box
  - position of the device is important
  - for some devices might be ambiguous which is the front, *e. g.* SIM card
- photos filenames: `tagID-view name[-detail].jpg`
  - examples:
    `A.SPH01-front.jpg`,
    `A.SPH01-back-serial.jpg`,
    `A.SPH01.MC01-front.jpg`,

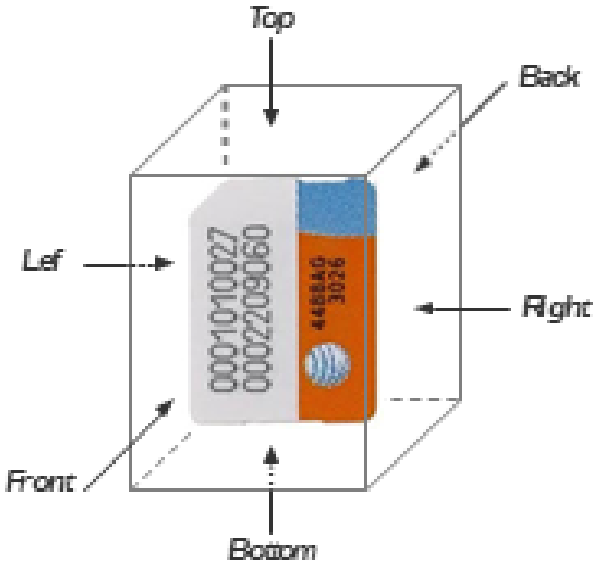


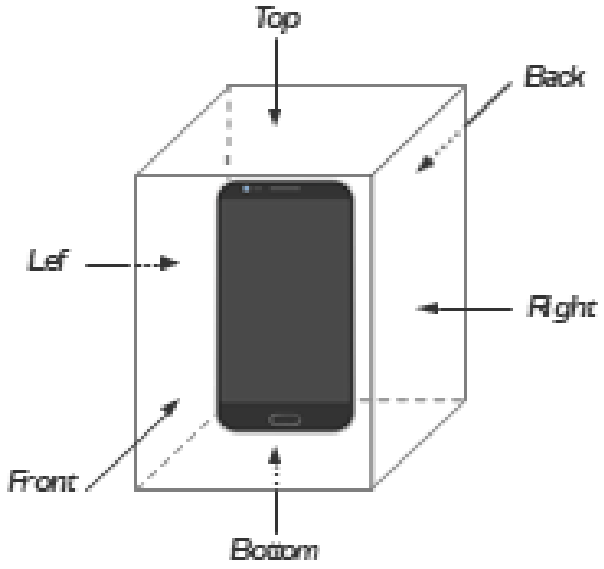
DIGITAL EVIDENCE HANDLING GUIDELINES

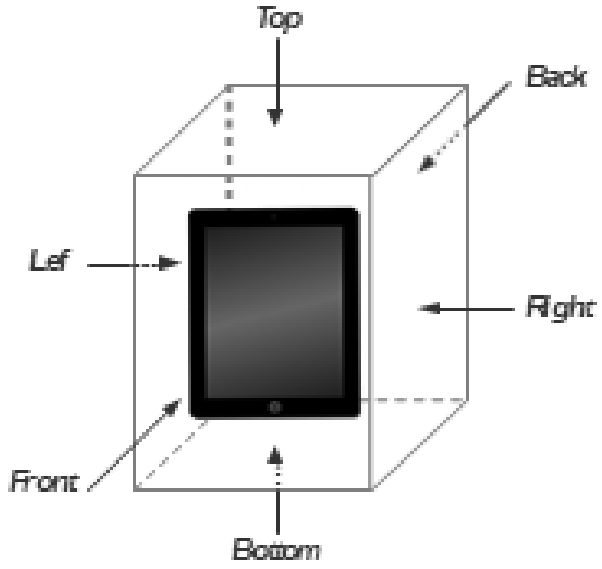












5 – Log: Catalog the device to **uniquely** identify it,
e. g.:

- tag ID device type
- brand and model number
- serial number, IMEI, UICCID, . . .
- type of intervention (logical or physical acquisition)
- device's condition (working / non-working)
- contents, *e. g.* has SIM or memory card
- worthy observations
- etc

Tag ID	A.SPH01
Device type	Smartphone
Brand	Samsung GSM
Model	GT-S5310
Serial n.	RV1D737C8AT
IMEI	356 431 051 982 186
SIM card	2: A.SPH01.SIM01 and A.SPH01.SIM02
Memory card	1: A.SPH01.MC01
Photos	Fig. 1, 2 and 3
Condition	Working
Observations	Battery not working
Intervention	Logic acquisition

ETHICAL CODE

Intent of the Ethical Code

- necessary to protect the integrity of the digital investigation process
- there are several codes

  Example: International Society of Forensic Computer Examiners (ISFCE)

  https://www.isfce.com/policy.html

A computer examiner will **always**:

- Demonstrate commitment and diligence in performance of assigned duties
- Demonstrate integrity in completing professional assignments
- Maintain the utmost objectivity in all forensic examinations and accurately present findings
- Conduct examinations based on established, validated procedures
- Abide by the highest moral and ethical standards and abide by this Code
- Testify truthfully in all matters before any board, court or proceeding
- Avoid any action that would knowingly present a conflict of interest
- Comply with all legal orders of the courts
- Thoroughly examine all evidence within the scope of the engagement

A computer Examiner will **never**:

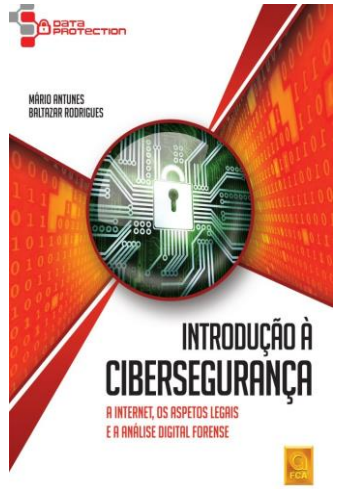
- Withhold any relevant evidence
- Reveal any confidential matters or knowledge learned in an examination without an order from a court of competent jurisdiction or with the express permission of the client
- Express an opinion on the guilt or innocence of any party
- Engage in any unethical or illegal conduct
- Knowingly undertake an assignment beyond his or her ability
- Misrepresent education, training or credentials
- Show bias or prejudice in findings or examinations
- Exceed authorization in conducting examinations

Exercises

- Make a list of the information to be registered to uniquely identify these devices or services:
  - memory cards, computers, hard drives, solid state drives, phones, smartphones, GPS devices, routers, switches, modems
  - services: web, DNS, POP3, IMAP, SMTP, SSH
- Create a tagging system
  - easy to memorize, that reflects hierarchy if needed (1 PC with several hard drives)
    - *e. g.* computer: PC01, HD inside computer: PC01.1, ...
- Write your findings on a document
  - create a table template for each device or service
  - add real photos for the devices: use your phones, your own PCs, ... (don't forget the ruler)

ANY QUESTIONS?

Antunes, M. & Rodrigues, B. (2018) Introdução à Cibersegurança: A Internet, os Aspetos Legais e a Análise Digital Forense. FCA (ISBN-13: 978-9727228614)



Carrier, B. (2005). File system forensic analysis. Addison-Wesley Professional (ISBN-13: 978-0321268174)