COMPUTER FORENSICS AND CYBER CRIME ANALYSIS

Thats a lot of words in the title

Fabio Lorenzato

Contents

Part I

Legal part

Chapter 1

Introduction

Back in the days, it was really easy to analyze the communication between two people, because the technology was easier and know, meaning that the technology is *between us*. Nowadays the technology is more *about us*, for example we use AI for facial recognition. Another example are the social network algorithms that track our activity online. This means that we use the technology to be "analyzed" instead of using it only to communicate. Most likely, the technology will be *in us*, which not only means robotics, but also be more integrated with human beings.
In those scenarios, computer forensics became more complex, having to take in account no only the quality of data(ie: a voice call), but also the quantity, which makes mistakes more likely to happen in some instances. Another aspect of computer forensics will be to discern what is human from what is a machine(ie: deep fakes, generative ai), and in particular where the liability lies. Furthermore, it is necessary to handle the legal aspect of different countries, which is quite difficult because of natural language limitations. An ideal solution would be to translate laws in code, which is basically impossible, so we can approximate this solution with **legal design** This has still to deal with two issues, which are information overloading(too much information), and the fact that laws are written by lawyers for lawyers, which makes them difficult to understand.
Another aspect to consider is the GDPR, and the still to come AI act, for example the article 22 prevents automatic systems to make decisions having legal weight without the control of a human. An example of this is the Lex Machina framework, which was able to predict the decisions of the judges via ML methodologies, which was banned because it could influence the decisions of people and judges. After all, false positives in digital forensics **can change people lives**.

Chapter 2

Foundations of digital forensics

First of all, lets define what can be considered digital evidence. There are many definition but the one that is most commonly used,and the preferred one at the exam, is:

> **Digital evidence** is *any* information of evidential value whether memorized or sent in a **digital format**.

Digital evidence has some defining characteristics: it is invisible to the untrained eye (it usually not simple to find), it may need to be interpreted by a specialist, it may be altered or destroyed trough normal use and it can be copied without limits (you must always work on a copy to avoid tampering evidence).

Digital evidence, to be admissible in court, must have some properties:

- **Authenticity**: avoid any digital evidence **tampering**. Only certified formats should be used.
- **Reliable and believable**: the evidence must be redly **understandable** to the judge. Sometimes a report is not available to explain the evidence, so it must be self-explanatory.
- **Proportional**: **respect** fundamental **rights** of parties affected by the measure
- **Admissible**: **compliant with law** and best practices(admissible in court). One thing is to have evidence, the other is to have it admissible in court, and that's not always the case.

There are three main kinds of digital evidence:

- **Created by man**: any piece of digital data that is the result of a step or action taken by a human person. Of those, there are two types:
  - **Human to human**, such as an email
  - **Human to machine**, such as a file saved on a disk
- **Created independently by the computer**: any piece of digital data that is the result of the processing of data carried out by a software in accordance with a specific algorithm and without human intervention (e.g. telephone records or Internet Service Provider logs)

- **Created by both man and the computer**: an electronic spreadsheet where data is entered by the human, while the computer works out the result.

Lets now define what digital forensics is:

> **Digital forensics** is the process of getting hold of evidence without modifying the IT system in which the evidence is found, ensure that the evidence acquired in another medium is identical to the original and analyze the evidence without modifying it.

The "Big Five" of digital forensics

There are some principle that are to be followed during digital forensics:

- **Data Integrity**: No action taken should change electronic devices or media, which may subsequently be relied upon in court.
- **Chain of Custody**: An audit trail of all actions taken when handling electronic evidence should be created and preserved.
- **Specialist Support** If investigations involving search and seizure of electronic evidence it may be necessary to consult external specialists.
- **Appropriate Training**: First responders must be appropriately trained to be able to search for and seize electronic evidence if no experts are available at the scene.
- **Legality**: The person and agency in charge of the case are responsible for ensuring that the law and the above listed principles are adhered to.

2.1 Digital Investigation Procedure

Digital investigation is carried out in many steps.

2.1.1 Identify the Suspect

The first phase is also the most difficult one, because there are many tools to be anonymous in the internet. The general approach to this phase is the following:

1. An investigator receives a **complaint** by a victim of **cybercrime** or detect an illegal content on line. (OSINT can be used to detect illegal content)
2. The investigator uses the Court System to compel the ISP to **reveal** a **physical location** that corresponds likely to the source of Network (**IP Address**)
3. Under a **warrant** (depending from the Jurisdiction) the location is searched and any computer or other device is seized

But why the ISP has to cooperate and give away the IP address? In the EU it is because **Data Retention Directive** 2006/24/EC, that requires that all the data of the communication must be stored for a certain amount of time, from 6 to 24 months. This directive was unconstitutional in many countries,

> **Data retention** (or data preservation) generally refers to the **storage** of call detail records (CDRs) of telephony and internet traffic and transaction data (IPDRs) by governments and commercial organizations

In any case data retention is usually a problem because there's no a homogeneous law for data retention, mainly for privacy reasons.

Even if there's no data retention law, the request of user data disclosure have been a lot in the past years have steadily increased, because the processes have been automated, as you can also see from figure 2.1.

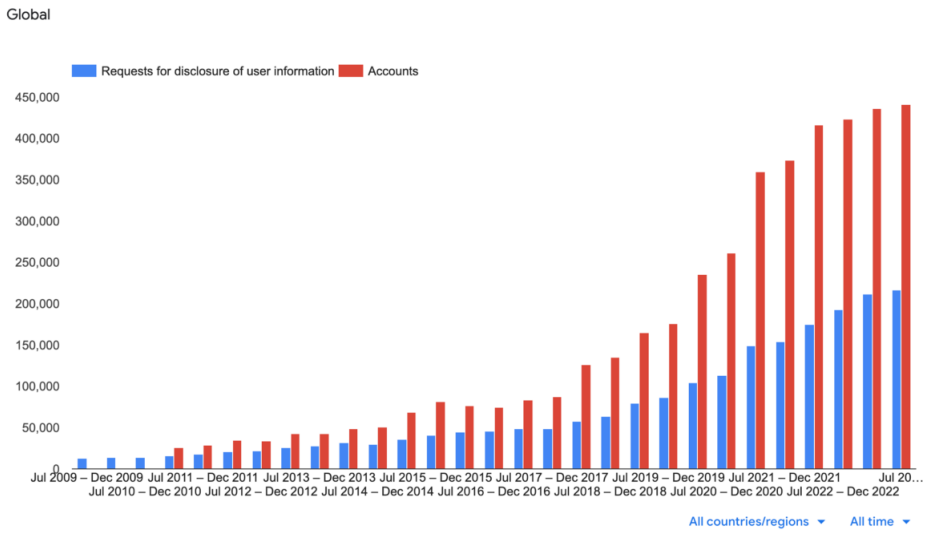

Figure 2.1: Requests for user data disclosure during the years

Another instrument that could be used to identify the suspect is the **facial recognition**, that can be used only for terrorism situations, but not in a systematic way.

2.1.2 Detecting and Seizing Digital Evidence

Anyone wanting to seize and validate digital/electronic evidences (content of an e-mail or an entire hard-disk) has to respect two fundamental **rules**: Bit-Stream Copy and Hash Function.

Bit-Stream Copy

> A **bit-stream copy** can **clone** the **entire drive**.

It is a particular form of duplication in which the content of the physical unit is read sequentially loading the minimum quantity of data that can from time to time be directed, then recording it in the same sequence on a standard binary file, generating a physical image of the original medium.

Hash Functions

During the forensic analysis of modifiable media, the Hash guarantees the intangible nature of the data that it contains.

> The Hash is a **one-way function**, by means of which a document of random length is converted into a limited and fixed length string.

This string represents a sort of 'digital fingerprint' of the non- encrypted text, and is called the Hash Value or the Message Digest. If the document is modified, even to the slightest extent, then the fingerprint changes as well. In other words, by calculating and recording the fingerprint, and then recalculating it, it can be shown beyond all doubt whether the contents of the file, or the medium, have been altered, even accidentally.

In any case we have two main problems while acquiring data: encryption and jurisdiction. After all, the ISP, TELECO or even a bank does have to cooperate and give out the data. Another big issue it the cloud computing aspect, because the location of data is another big problem, because it can be either:

- **at rest**: does not reside on the device.
- **in transit**: cannot be easily analysed because of encryption.
- **in execution**: will be present only in the cloud instance

Validation of digital evidence is a very important step, because it is the only way to prove that the evidence is authentic. This is especially important for proof found on the internet. There are some tools that can be used to validate digital evidence, such as Web Forensics. Another issue that has to be accouted for is the **chain of custody**, that is the process of maintaining and documenting the location and handling of evidence. After all, the bit is eternal, but the storage medium is not.

2.1.3 Validating Digital Evidence

2.1.4 Chain of Custody

2.1.5 Analysis of Digital Evidence

2.1.6 Presentation in Court

Chapter 3

Cybercrime Convention

Part II

Technical part

Chapter 4

Introduction to digital forensics

First of all, we need to understand some basic concepts.
Forensics analysis is the process of investigating and analyzing information to gather evidence to solve legal problems. In order to do that, the collections, presevation and analysis or presentation of digital evidence is required to support the investigation. **Computer forensics** does just that.
Forensic science is not a new field, it has been around since the ancient times. The first recorded use of forensic science was around 1900 BC in Babylon, where fingerprints were used to identify the author of a clay tablet. Digital forensics is a relatively new field, it started in the 1980s with the advent of personal computers, for example the first convicted person for digital crimes was Robert Tappan Morris, who created the Morris Worm in 1988, and found out by analyzing computer logs and network activity.

> **Computer Forensics** is the **discipline** that combines elements of law and computer science to collect and analyze data from computer systems, networks, wireless communications, and storage devices in a way that is admissible as evidence in a court of law.

In general, when carrying out a forensic investigation, the following questions are important to keep in mind to fully understand the crime scene:

- **What** happened? What is the timeline of events?
- **Who** was involved?
- **When** did it happen?
- **Where** did it happen?
- **How** did it happen?
- **Why** did it happen?
- **How** did the incident occur?

All this questions allow to **support legal proceedings**, the mitigation of damages and mature future prevention strategies.

4.1 Computer forensics goals

The main goals of computer forensics are:

1. **retrieve** the **input** data(ie: what has been typed)
2. **determine** the **actions** performed by the user (ie: what programs have been run)
3. **analyze** the **used files** (ie: what files have been modified)
4. **identify** the **damage** done to the system (ie: what files have been deleted)

> In essence, the goal of computer forensics is to **gain conprehansion** of **what happened**, at least technically speaking.

4.2 CF terminology & relevant concepts

Before going any further, we need to understand some basic concepts and terminology used in the field of computer forensics.

For a definition of **digital evidence**, we can refer to definition 2. In general, we can expect to deal with different types of digital evidence, because they can use different level of **abstraction**. Most of the time that are fragile, because they can be easily modified or destroyed, which is undesirable during forensics investigations. Its also difficult to correlate connection between data and real events, and to prove that correlation in court.

Another important concept is the **chain of custody**.

> The **chain of custody** is the **documented** and **unbroken process** of handling evidence, from the moment it is collected to the moment it is presented in court.

This is important because it ensures that the evidence is not tampered with, or even accessible by unauthorized personnel, and its most important to ensure that the evidence is admissible in court. For those reasons it requires knowledge about logging procedures, secure storage and legal protocols, because many security measures are required to ensure the integrity and confidentiality of the evidence.

> **Data Acquisition** is the process of **collecting evidence** from devices without altering or damaging the original data.

This is one of the most subtle one, because it requires a deep knowledge of how memory works, because it may be required to do disk imaging when the data is at rest or even live data acquisition when the data is in use. In any case, understanding of whats going on in memory is required to avoid data corruption and tampering.

> **Write Blockers** are hardware devices, or software tools, used to prevent any data from being written to a storage device during analysis, preserving the original data content.

Hardware write blockers are the most secure, because we don't have to trust that the software is behaving correctly, but are also more expensive. In any case, they are fundamental for legally defensible acquisitions.

> **Forensic Image** is a **bit-by-bit copy** of digital media, including deleted files and data in slack space, which is an exact replica of the original device.

This is a strict requirement for digital forensics, because it allows to preserve the original data, for example if the data is not exactly copied the same way, the result of an hash function will be different.

4.3 Forensics scenarios

When doing a forensic investigation, there can be different scenarios that can be encountered. Some common scenarios are:

- internet abuse from employee
- computer-aided frauds
- data unauthorized manipulation, like data theft or disclosure
- computer/network damage assessment
- ... and any time digital evidences may be involved in an incident

4.4 Investigation phases

The investigation process can be divided into several phases, at least from a technical point of view. Those depends on the standards that are followed in a given country where the investigation is taking place(fore example in America the NIST standards are followed).

4.4.1 Identification

Its the first step of the investigation, an it take place when the crime scene has been accessed. The goal is to identify which are the potential sources of relevant that, which will be used as digital evidence if relevant.
Most of the time we have an overwhelming amount of data, but most of it is not relevant to the investigation. Reducing the amount of data is also a goal of this phase. On the other hand, it is also possible to miss some important data, so it is important to be careful.
It is important to recognize (all relevant) data sources before any acquisition begins, like:

- hard drives (HDD/SSD)
- memory (RAM)
- mobile devices (smartphones, tablets)
- cloud storage

- network traffic
- removable media (USB drives, DVDs)
- ...

The steps to follow in this phase are:

- if possible, acquire data before reaching the crime scene(pre-analysis), instructing the staff to identify possible sources of evidence
- perform an **initial survey** of the scene (physical or network environment)
- **identify key devices** and data locations (local storage, remote servers, cloud services)
- **check** for **connected devices**, including peripherals like printers, removable media, or network-attached devices
- **map** all potential **data sources** using network topology

Pay much attention to the ephemeral storage of data, with reference to the order of volatility. The order of volatility is a concept that refers to the order in which data should be collected, based on how long it will be available.

4.4.2 Collection

After the identification phase, one must physically or remotely taking possession of the evidence (e.g. a computer) and its connection (e.g. Network or physical, like USB disk). The timing is most crucial in this phase, while also maintaining the integrity of the evidence, to minimize the risk of evidence tampering or data loss.
The steps to follow in this phase are:

- isolate devices to prevent them from being tampered with remotely (e.g., disconnect them from the network)
- use devices to block external communication for mobile or wireless devices (e.g. faraday bags)
- use network isolation tools for virtual and cloud environments to prevent remote access (e.g security groups, virtual private cloud, firewall rule)

Its important to ensure the integrity of the evidence while maintaining the system running, because shutting it down can cause the loss of volatile data(e.g. RAM).

4.4.3 Acquisition

electronically retrieving data by running various CF tools and software suite. Its a separate phase from the collection phase.

4.4.4 Evaluation

Now that all the data has been collected, it is time to evaluate it. In this phase the data is analyzed to substantiate claims and to determine how they could be used against the suspect.

4.4.5 Presentation

At last, the evidence is presented in a clear and understandable way to the court, in a manner which is suitable for lawyers, non-technical staff and the law.