Digital Forensics and Cybercrime

Gianvito Caleca

2024

Contents

Part I

Cybercrime: threats, modus operandi, underground economy, financially-motivated malware

Chapter 1

Cybercrime: Threat Landscape

Working on security we work on **risk**, which is composed of different elements:

$$Risk = Asset \times Vulnerabilities \times Threats$$

The risk is the statistical and economical evaluation of the exposure to damage because of the presence of vulnerabilities and threats.
With no threats there are no risks, because it is the only thing that can nullify this equation. Assets and vulnerabilities are never absent.
With no *Threat model* we're completely misguided on the management of our system in terms of security.

1.1 Threat landscape

Threats can be roughly divided along three directions:

Internal vs External treaths an internal treath comes from the inside of the organization, it is part of the organization, while an external one don't.

Generic vs Targeted many security threats are generic:

- Generic threats: when you walk towards the underground, you're subjected to pickpocketers: they don't pick you because it's you, you're one of the others, it is a generic treath.
  Generic threats are not linked to us by definition, they are linked to us because we look alike somebody easy to be pickpocketed, not because we were the target.
- We can also consider also threats which are not so very generic: aggressions for sexual reasons are more targeted to one gender than one other, while still being generic.

- Targeted threats: specifically designed against us: criminals target **that specific company** *(e.g. the know how for racing cars, i want to steal from THAT company in particular).*

These directions also affect the kind of attacker: pickpocketers vs highly "professional" skilled stealer of information

Financially motivated vs Anything Else we divide:

- **financial attackers** *(which are the most of them)*,
- **other attackers** *(governments, secret services, hacktivists..)*.

Financially motivated attackers has 2 important positive characteristics:

- **easy to predict:** you look at valuable goods from companies and you can predict who the attackers can be *(e.g. ransomwares)*
- **they're relatively easy to handle:** just deny them the opportunity to take money, like making it too costly wrt what they want to earn.
  Notice that what is easy to understand is how to handle them, not to actually handle.

Non financially motivated ones are more difficult to handle:

- We cannot make them too costly or too risky *(e.g. Russians vs Ukraine's Donbass power plant: they have an entire state funding them, and they cannot be arrested for it)*, they have more money and they can take more risks
- If they're internal, they already know about the company and its security systems.
- The more motivated is the attack, the more determinated is the attacker

1.1.1 A gartner quadrant of threats

	Generic	Specific
Internal	**disgruntled**[1] employee	**socially-engineered or disonhest** employee *(financially motivated)*
External	Criminals, usually looking to make money *(financially motivated)*	A variety of advanced hackers *(mostly financially motivated)*

- **Generic internal threats:** a disgruntled employee who wants a raise, has something against his colleagues, was fired ...
  The treath doesn't need to be linked on what the organization does.
- **Specific internal attacks:** former employee stealing from old employer and bringing to the new one, it also exists a specific job which consists in being hired from companies just to steal from them.
  Or, employee can also be *social-engineered* into behaving like attackers.
- **Generic external attackers:** attackers which want to earn quick money, they aim to things which are common for all companies, like stealing money from bank accounts and so ....
  They are mostly motivated by financial reasons.
- **Specific external attackers:** industrial spies which aim for money, or governments or terrorists aiming for the destruction of a power plant near Donetsk.

The most important quadrant is than the one which comprends disonhest or social engineered people *(human factor)*: our instinct say us that we need to protect from the outside because we are clan-based animals, and we have our group of trustworthy animals *(our clan)* to which we are positively implicitly biased. **Think about the most famous security technologies:**

- Firewalls are meant to keep people out
- Antiviruses are meant for generic attacks, they use a generic list of malwares

Most of our security technologies protect against external attackers.

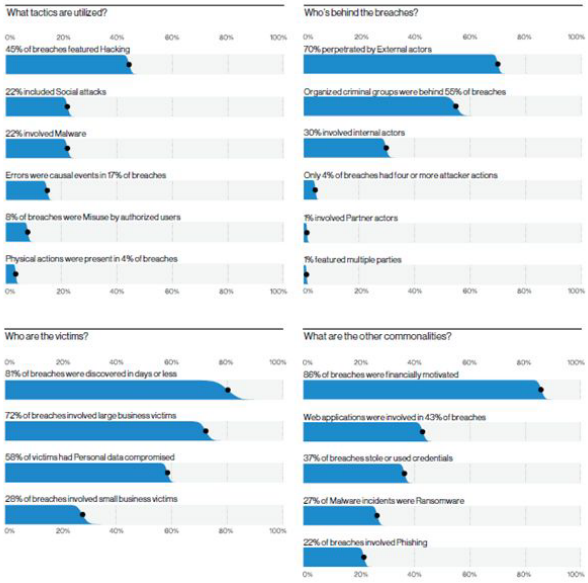
1.1.2 Examples of internal treaths

A bit more on:

- **Reckless or socially engineered employees:** in security we often have a wrong perception of the *human element*.
  When a company gets infected by a malware because a user clicked on a link or something similar, we blame the *stupid* user for it. But, actually it's not fault of the employee: it is fault of who had the responsibility to manage the fact that a threat was present and it had to be controlled.
  Since there will always be someone to click on links, our job is to make sure that they doing it doesn't cause damages to the organization.
  e.g. in aviation they study human factor to understand what to do to don't let people make mistakes: in airplanes cockpits there are different levers with different shapes, them are exactly shaped the same in different kinds of planes, it is a standard, and the reason is because people lost their lives because of similar levers.
  The solution wasn't to train pilots, but to make the threat difficult to happen.

- **Thirty-party users:** they're for example consultants that work in a company for a long time, where they actually do jobs very similar to the ones of the actual employees while not being true employees of the company.

1.1.3 Data breaches and targeted attacks



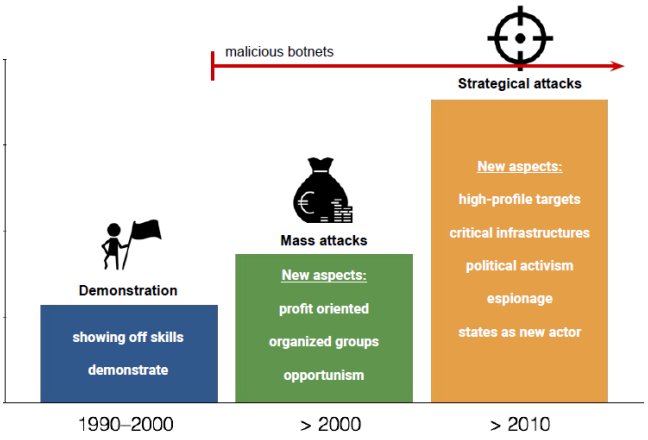
Verizon has a big sample of incidents made to their customers:

- The largest part of the attacks features hacking of some sort, the second larger chunk comprends social attacks.
- The 70% of breaches are perpetrated by external actors. This means that 30% of them starts from the inside, an incredibly high percentage.
  Consider that data are skewed[2], in the real world the internal attacks can be even more! *(remember the external threats reasoning, rationality (data) tells us that a lot of threats come from the inside, while our instinct guide us to be aware of the outside)*
- The 86% of them were financially motivated, 14% which are not, are a lot too!

These data are biased and are not representative of the true reality of threats, because simply they're not all collected. This collection can still help us to rationally think on how to manage the overall situation.
Keep in mind that some security incidents are specifically designed to be difficultly found, a lot of them has never been. This means that there is an entire set of breaches that no one ever investigated.
We measure attacks that we investigated/detected, and this is called the **observation bias**:
Other attacks by their nature be discovered, *(e.g. Dos, ransomware . . .)* that's the reason why they are so prevalent in our statistics: they are observed.

[2] biased

1.2 Brief history of malicious software



Over the last few decades attacks and attackers changed:

- **1990-2000:** most of the attacks were meant to show skills and explore
- **2000-2010:** with the birth of the Internet, massive, profit oriented attacks born. In this period were born also groups working as profitable enterprises of cybercrime.
- **2010-today:** attacks evolved: mass attacks kept happening, but now a lot of them are high profile financial attacks *(before: ransomwares, i ask small amount of money to single persons. now: ransomware attack specific companies to get a lot of money)*

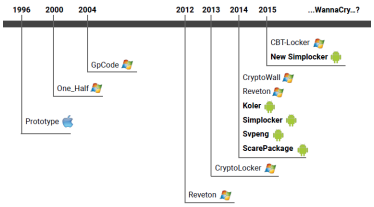
1.3 Financially motivated attackers

Today financially motivated attackers are the "mass", they're interested in monetizing their attacks in possibly two ways:

- **Direct Monetization:**
  - Credit card/bank account frauds
  - Ransomware attacks
  - Crypto-miners
  - Fake antiviruses: they show you warnings that say your computer is infected, and to pay to activate the premium version to get rid of the malware. In reality they are just fake programs asking you to pay for a license.
  - Premium calls: back in the days when you called to connect to the provider, someone managed to change the number to let you call a costly one.
- **Indirect Monetization:**
  - Abuse of computing resources
  - Information gathering (stealing of account infos to sell them)
  - Making the machines part of a botnet to rent or sell the botnet

1.4 Direct monetization: ransomware attacks

1.4.1 Brief history of ransomware



In 1996, a paper describing a crypto-malware that was exactly a ransomware was published about 15 years before CryptoLocker which was the first really successful one.
They just *"looked into the crystal ball"* of what in the future would be valuable: taking as hostage digital valuables could become a good way to steal money.
Then something happened:
the internet in 1996 was really different than the one of 2013, ransomwares needed a way for the user to pay a ransom in a way that was easy to perform. The preferred way to perform payments on internet is with credit cards, but this kind of payments have two important difects:

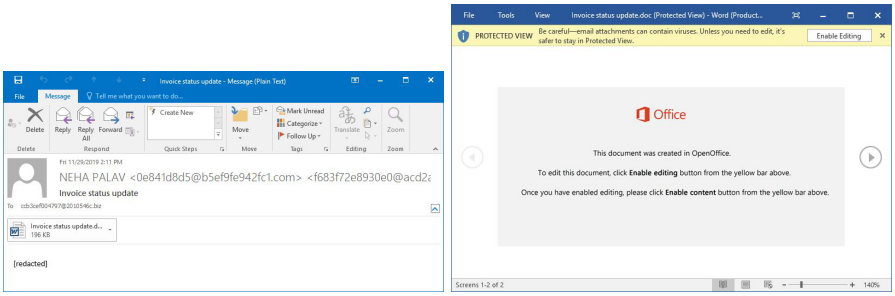
- they are trackable
- they are refundable when fraudolent

Then a new suitable way for payments was born: **cryptocurrencies**.
They perfectly fit this need: they are easily accessible, and payments are not reversible.

1.4.2 Ransomware screenshot and social meanings

Look at the structure of the timer: it is made to let you to percieve urgency. The stress of the urgency makes human take bad decisions, with the timer running you're more prone to pay because you feel that you have limited time. In social engineering attacks, urgency is always an important part. *(think about ultimatums in diplomacy, job offers with time limits, poltrone e sofà...)*

1.4.3 How to get infected



(a) (b)

Looking at the images we can see an interesting example:

- User receives an e-mail with a short text that contains some sensitive information and an attachment which is social engineered in a way that many people would click
- Microsoft Office shows a message to try to prevent code execution, but the user is prone to click "Enable" because the text contained in the document makes him think that's the right thing to do, even if a lot of grammar errors are present and the whole situation looks suspicious

One could have said that the user must had to be trained to not click, but the reality is that he is not understanding what's happening on his computer.
We can't train a generic user to not click: if your job is to open invoices received by e-mail, there is no way you can be trained.
Even if we train him to click only on what is coming from attendible e-mail addresses, if the *"attendible"* part was hijacked?
What we really should do is to figure out a way that when someone clicks enable, this won't cause any damage. Of course, to train the user to let this happen the less frequent possible is good, but not the main thing to do.

1.4.4 Encryption mechanism: how do ransomware work

Once the malware starts running on target computer, it generates a random symmetric key[3], usually one per file, and it encrypts each file with it.
The symmetric key itself it's encrypted with a public key, of an asymmetric key pair generated on the server of the group that runs the malware. The private key is stored on the server, and it will released only when the ransom is payed.
Most of this process is automated: when cryptocurrencies are transferred to the criminals' address, the server releases the key associated to it. Different keys to allow the attackers to decrypt only a part of the samples to demonstrate that they can do it.

[3] used both to encrypt and decrypt information

1.5 Indirect monetization: Botnets

The word botnet comes from the words *robot* and *network*. A botnet is composed by few hundreds to millions of infected devices which run some sort of malware they got by opening an infected email attachment, by plugging an infected USB drive, visiting an infected website *(these are examples of ways to get infected)*
Each botnet has a **botmaster** who controls and rents it out to perform tasks *(denial of service, spamming, phishing campaigns, crypto mining ...)*, and his only objective is to earn money by renting them.
The significant challenge with this type of crime is that each bot per-se is not dangerous for the machine itself, but for other machines, and since the cost of cleaning up a machine falls on its owner, some can decide to not do anything about. This cost can be seen as a small fee to be payed by *the community* to get everybody safe. Like a vaccination, some cases of infected machines are not a big trouble for the community, while a lot of them can be really dangerous for everybody.

1.5.1 Rise of the bots

Back in the days botnets were used to control IRC chats *(see IRC wars)*. Then instead of using the compromised machines to control the chat, botmasters started to use the chat to control the bots.
In 1999, there was one of the first DDoS attacks, which was against University of Minnesota and used at least 227 bots.
In the 2000s DDoS attacks against high profile websites *(Amazon, CNN, eBay ...)* got huge media coverage.

1.5.2 Geolocalization of botnets command and control

Rank	Country	Q2 2020	% Change Q on Q
#1	United States	896	7%
#2	Russia	812	32%
#3	Netherlands	337	61%
#4	Germany	185	7%
#5	Singapore	131	157%
#6	France	108	35%
#7	Great Britain	89	37%
#8	China	74	-15%
#9	Bulgaria	72	38%
#10	Hungary	70	New Entry

This chart shows the number of new botnet C&Cs detected by *Spamhaus* in the second quarter of 2020, and the increase wrt the first one.
By keeping in mind the *observation bias*, we can also introduce another kind of bias which is called **heatmap effect**: we're biased on the density of the population and on how much data we collect from a certain country.
In this kind of charts we'll always have U.S.A. on top, because there's more penetration in the internet, and also this kind of things is tracked. While we'll have less data from less technological advanced states or *perfectly democratic countries* like Russia or the P.R.C. which can result in lower positions while in reality being the first ones.
Here we see an important increase of C&C from The Netherlands:

- it is possible that **something is on**
- or simply that some Netherlands organization decided to participate in this data collection feeding a lot of *new* data

Funny thing about Russia and P.R.C: Russian people hack russian computers too, while P.R.C citizens don't hack into their fellows machines.

1.5.3 Type of (botnet) malware families

Rank	Q2 2020	% Change Q on Q	Malware Family	Description
#1	436	772%	AgentTesla	Credential Stealer
#2	290	-46%	Lokibot	Credential Stealer
#3	169	New entry	RedLineStealer	Credential Stealer
#4	156	51%	NanoCore RAT	Remote Access Tool (RAT)
#5	112	-27%	Gozi	e-banking Trojan
#6	98	-5%	AZORult	Credential Stealer
#7	92	1%	RemcosRAT	Remote Access Tool (RAT)
#8	74	23%	njrat	Remote Access Tool (RAT)
#9	74	New entry	DanaBot	Credential Stealer
#10	69	17%	ArkeiStealer	Credential Stealer
#11	67	63%	KPOTStealer	Credential Stealer
#12	62	New entry	IcedID	e-banking Trojan
#13	61	New entry	AveMaria	Remote Access Tool (RAT)
#14	55	-18%	Adwind	Remote Access Tool (RAT)
#15	51	21%	NetWire	Remote Access Tool (RAT)
#16	47	New entry	QNodeService	Remote Access Tool (RAT)
#17	47	57%	RaccoonStealer	Credential Stealer
#18	46	-4%	Pony	Credential Stealer
#19	45	105%	AsyncRAT	Remote Access Tool (RAT)
#20	43	New entry	Zloader	Loader

Every botnet in general can be used to do any sort of things, but they're used to do only one of them:

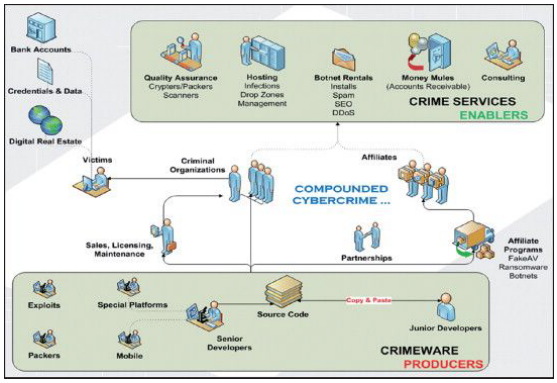
- **Credential Stealers:** used to steal sensitive credentials.
- **Banking Trojans:** credential stealers specifically designed to perform stealing of bank accounts information. *(Gozi is the most common one, Zeus the most important one)*
- **Remote Access Tools:** basic botnet oriented malware, which allows to control a computer in order to perform whatever.
- **Loaders:** specifically designed to allow a botmaster to load a program of whatever sort on a computer. Maybe a client ask you to install a certain malware on a lot of computers, and you can do it with a loader.

We talk about families because there are criminal groups that only develop their source code, and who performs the attacks buys the code and personalizes it for the specific purpose they need.
So we can consider three businesses: how to develop them, how to configure them, how to use them.
There is a market for services related to malwares and cyber attacks *(cybercrime as a service, underground market...)* structured around the needs of cyber criminals. This market is fueled by the money that these schemes make. Some of these money pays for the tools used to make that money.

1.6 The cybercrime ecosystem

The status quo consists in **organized groups** performing **various activities:**

- Development and procurement of exploits
- Site infection
- Victim monitoring
- Selling *exploit kits*



These ecosystems can exists because some of the activites done are not illegal: developing exploits per-se is not illegal, selling one is not illegal, the usage of them may be or may be not illegal. Even the services related to the configuration of malware are not illegal, while execution and operation may be. If you're sufficiently shielded and you run these *businesses* in countries that cannot persecute you, we can even know your name and surname but you cannot be arrested.

- Developers write the source code for malwares with the help of packers and exploits developers
- Crime service enablers:
  - Quality assurance makes sure that antiviruses don't detect their software
  - Bulletproof hosting is a kind of *close-an-eye* hosting, *(e.g. russina business networks)* with permit certain suspect activities over their infrastructure, they're sort of borderline organization with lot of regular customers, and some bad ones

1.6.1 Identity sales

People's identities are sold on the black market. They are worth because they can be used for fraud, to open bank accounts used for money laudering for example. The more they're useful, the more they are worth. Worth 50 dollars for you that sell it, the one who pays is going to use them for fraud which is worth more. The black market is fueled from an enormous amount of money that people stole.

1.6.2 Drive by download

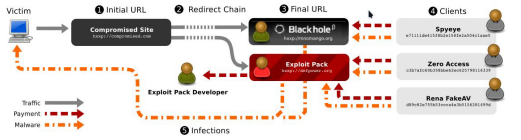


Figure 1: The drive-by-download infection chain. Within the exploit-as-a-service ecosystem, two roles have appeared: *exploit kits* that aid miscreants in compromising browsers (❸), and *Traffic-PPI* markets that sell installs to clients (❹) while managing all aspects of a successful exploit (❶,❷,❸).

This is another way in which malwares are installed:
an exploit breaks into your browser and executes code on your machine. This requires your browser to be vulnerable and for you to visit a compromised website with the exploit running *(streaming, cracks of programs websites,...)*.
The most compromised are *factually illegal websites*: you're not scared of the strange url if you are in need to see the last movie in streaming, download the expensive program's crack, this is in contrast with a normal situation like going to amazon to buy shoes.
But, *traps* can be also present in legitimate websites.
What actually happens is that the users endup in a series of redirects, called **redirect chain**, which will endup in a visit to a website running what is called **exploitation pack** for your browser.

1.6.3 Exploitation kits sales

This kind of kits is sold in the dark web by well known organizations, take *Blackhole* as example:
they were buying exploits from developers to earn 10s of millions per month renting them.
Their boss, surprisingly called Dmitry "Paunch" Fedotov, was arrested in 2013 after years of running the website.

1.6.4 Monetization on the dark web

Monetization takes a lot of forms, one of them is stealing credit card numbers to sell them for a relatively small price.
The hackers get *easy and fast* money, while the buyers need to use the real

money on the cards to buy expensive objects or to start frauds because they can be easily refunded. *(apple computers, ethnic travel example..)*.

1.6.5 Cybercrime and perception

Online, you detach people from the picture. It's actually easier to commit crimes, it's easier to crack a program instead of stealing a car, lot of cyber criminals would never be criminals in the real world. But when you do cybercrime, your perspective is different: you have another perception of what you're doing, that's the reason because crime happens more online.
Ethical people understand the consequences of their actions, and may decide to not do them, but someone can not care about what can happen as a consequence of their actions and perform them anyways:
in the Jamal Khashoggi case, the exploiter which wrote the code for the spyware that ended up to make him killed, would never kill someone. But the consequences of his actions did.

1.6.6 Money mules and money laundering

Most cybercrimes end up with a digital form of money, which crimials want to bring in a physiscal form, completely disconnected from what they did in principle.

- They can pay invoices to companies somewhere in the world, which in a certain way makes sure that the money come back clean. *(traditional way)*
- They can make use of **money mules**

Money mules are intermediary people which by purpose or not make earnings coming from cybercrime clean money.
Accomplished ones may be people that have nothing to lose *(prejudiced, poor people ...)*, they just open bank accounts by their names and get the money transferred and then withdrawn and handed to criminals.
Some of them may also open accounts by somebody else's name using identities that may be stolen with an attack and bought online. They're most difficult to catch because the Police needs to be actually there while they are withdrawing money.
Unaccomplished ones may be people which are fooled by things like the Nigerian prince scam: they get the money on their bank account, send 70% of them to criminals in packages or by moneytransfer, and keep 30%. Most of the cases they get arrested and have to pay back all the money, also the ones they don't own no more.
Buying and selling of goods *(ricettazione)*, videogames curriencies, ..., are other ways to perform money laundering.

Chapter 2

Cryptocurrencies: abuses and forensics

In this chapter we will talk about the cryptocurrencies role in cybercrime.

2.1 Bitcoin and blockchain

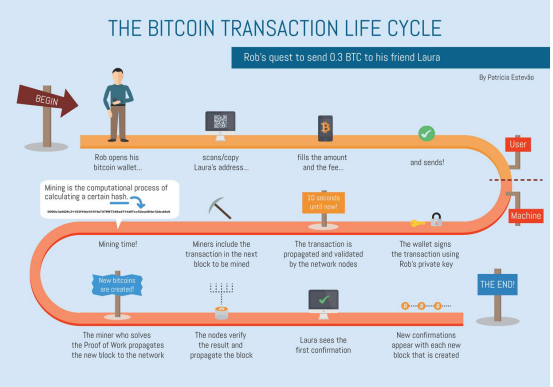
Bitcoin is an attempt to create **electronic cash**: a digital currency that has lots of properties that physical money have, one of them is **no central authority**. What Satoshi Nakamoto wanted to have was an electronic distributed ledger[1]. Having no central authority is a really hard problem to solve *(byzantine consensus)*, he did it with **the blockchain**:
The blockchain is a **shared, append-only, trustable** ledger of all coins transactions. The limits of distributed consensus defined in the Byzantine Problem and CAP Theorem are solved using the technique of **proof-of-work**.

2.2 Wallet and addresses

A wallet is the software that allows to manage and store the **public** and **private** keys for each of the user's bitcoin addresses, to create and sign transactions, track the balance.
A bitcoin address is an alphanumeric string which identifies a *point* where you can send bitcoin to, and where they can be sent from.
In order to know how many Bitcoin are *stored* in that key you need to go to the origin of all the transactions and track their flow. This computation is not really efficient but **it can run without central authority**, so the only reason to use a blockchain is that is really so important to get a way out of a central authority.

[1] ledger=registro

2.3 The bitcoin transaction life cycle

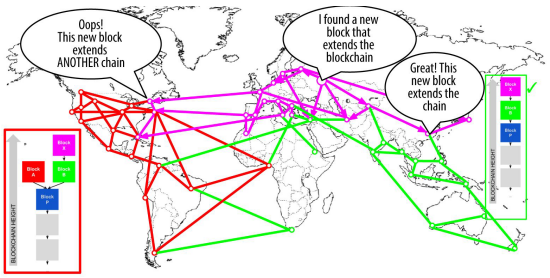


As soon as the transaction is written in the immutable ledger, the money are property of the destination address.
To be sure that this is the only transaction done with those Bitcoins, the immutable ledger must be consolidated: that would have been easy with a central authority, but to agree without one, we need a history that **cannot be modified**, and to do this we use mining.

2.4 Mining

It is the process to generate a block containing a set of new transactions, and to append it to the blockchain.
Miners compete to generate a new valid block that solves a complex mathematical problem by bruteforce: the solver is then rewarded with a fixed number of BTC.
This is a computational demanding activity: miners put a set of transactions in the new block, put there a destination for the reward and then start bruteforcing the block to find a sha256 hash of it which starts **with a certain number of zeros**: the more the zeros, the more difficult is the process.
When a block is added to the blockchain, the other miners stop bruteforcing their blocks and start to compute the hash of another one, to link it to the last found block.
The difficulty increases more and more, the reward got smaller and smaller but BTC value gets larger and larger.
Besides of the reward you get, you can also put a fee you want to receive from the transactions you included in the block.
Since rewards are becoming smaller, fees are becoming larger, making small transactions unfeasible.

2.5 Fork events



Sooner or later two miners will get a solution at more or less the same time: this causes a **fork** where the situation is that for some miners the blockchain ends with block A and for some others with block B.
Now the blockchains are two, and sooner or later some miner will find a solution and append a new block to one of them.
The rule says that **the longest blockchain is the true one**, as soon as one blockchain becomes the longest, all the miners move to that leaving the other one sadly alone.
This is done because in order to revert a payment you'd need to go back on your block, go ahead mining until your blockchain is the longest one: this means that you would compete with all the other miners.
To be sure that a certain transaction is there and cannot be canceled, you must wait that the block in which it finds itself is a couple blocks behind the last one.

2.6 Bitcoin and black market

Bitcoin immediately started to be used in black markets: there was a way to pay without an infrastructure.

2.6.1 Pseudo-anonimity

The first reason to use Bitcoin for black markets is **anonimity**.
In reality Bitcoin is actually **pseudonymous**: the identifier for the transactions is not directly connected to an identity, but to an address which can be seen as a username.
Every entity generates multiple keys for themselves, so more pseudonyms.
This pseudonimity is robust because addresses don't refer in any way to their owners.
Transactions are public by the way, and this is a terrible property if you're interested in anonimity; while addresses are not connected to people, there is something that connects addresses related to the same person:

- A transaction can have multiple keys as input, if a transaction has multiple inputs, it is very likely that the inputs are all owned by the same entity
- Bitcoin works in a way in which you cannot spend less than the whole amount of bitcoin you have in your address: to send a smaller amount to somebody, you'd generate a transaction with two outputs
  - The destination address
  - A new address called Shadow Address which is owned by the same person who started the transaction, where the rest of the currency is sent.

Until 2013, a bug made very easy to know which of the two outputs in a transaction were the shadow address, hence making easier to track their correlations. It was then possible to track addresses in a way in which keys belonging to the same entity were collapsed in large sets.
When you have one of those addresses connected to somebody's name you can actually connect all the others in the set to him.

The Silk Road Example

Silk Road was a black market which used BTC as currency, it was ran by someone under the name of *"Dead Pirate Roberts"*.
It was found out that someone using the nickname *"altoid"* were advertising Silk Road on different websites, and with the same nickname earlier in the past was also hiring developers for a PHP project. *(which then became Silk Road)*
In that specific post, altoid was asking to contact him at *"rossulbricht at gmail dot com"* if interested.
He also posted a request for help which included PHP code with his address in it: 1LDNLreKJ6GawBHPgB5yfVLBERi8g3SbQS[2], which then was found to be associated with Silk Road
Ross Ulbricht was arrested in 2013. *(Hope he was arrested for using PHP and not for selling drugs)*

2.6.2 Why do people use Bitcoin for ransoms?

For these two characteristics:

- BTC transactions, in contrast with common electronic payment systems, are irreversible.
- To use other kinds of digital payments an infrastructure is needed *(bank account...)*, with cryptocurrency you just need an address.

[2]today there are 20 dollars of Bitcoin

Part II

Fraud detection and analysis

Chapter 3

Frauds: definition and types

3.1 What is a fraud?

Wrongful or criminal deception intended to result in financial or personal gain
Fraud is an uncommon, well-considered, imperceptibly concealed, time-evolving and often carefully organized crime which appears in many types of forms
The two definitions toghether explain well what a fraud is. The second expecially gives us the fraud's characteristics.
A fraud is:

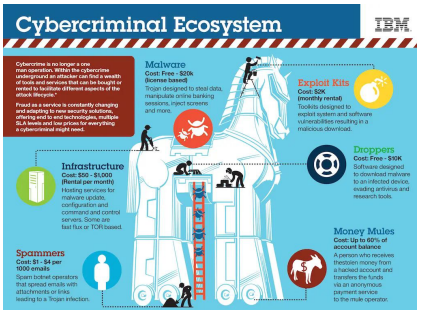
- a **social phenomenon**
- **Uncommon:** usually the percentace of frauds is very small. Only a minority of cases concerns frauds, and only a limited number of them will be **known** to concern fraud.
  This makes them difficult to **be detected**, because fraudulent cases are covered by legitimate ones, and makes difficult to **learn from historical cases**, because of the small number of available samples.
- **Well considered and imperceptibly concealed:** fraudsters try to remain *unnoticed and covered*, they blend in[1] frauds by not-behaving differently from non-fraudsters. Fraudsters **hide** very well by well-considering and planning how to precisely commit fraud
- **Time-evolving:** since fraud detection systems improve and learn by example, fraudsters adapt and refine their methods to remain undetected. Fraudsters techniques evolve in time along with or better ahead of fraud detection mechanisms. This is an adversarial situation in which usually the fraudster is always ahead. Each security solution has a cost which can be a waste if the fraudster is able to break it in days.
  Not only frauds change, but also the legitimate behavior of the users.

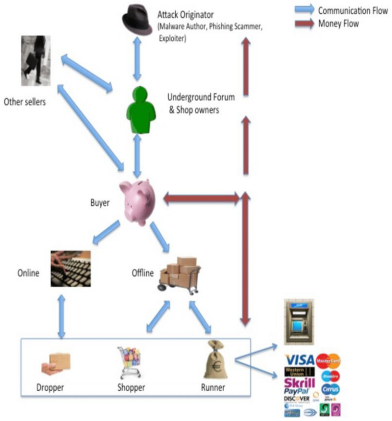
[1] mimetizzano

Also this must be followed in time, this is called concept drift: legitimate behavior of models changes in time.

- **Carefully organized crime:** they're not usually composed of a single operation, and fraudsters belong to a complex and organized structure. The entire cybercrime group (context) after a fraud must be investigated.

3.2 Cybercriminal ecosystem



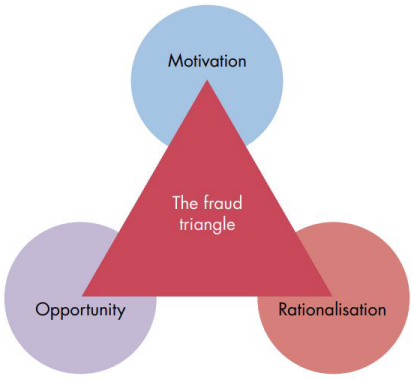


- Some groups are specialized in creating malwares
- Some in providing infrastructures to perform money laundering
- Others to perform financial crimes
- Others spamming, phishing
- Others exploiting

Every single time we think about a financial fraud, what we see is just the last part of an entire interaction between criminals in the entire ecosystem.
A buyer usually isn't able to perform cyber-crimes, he just interacts with criminals trough forums to pay for their services, which put toghether form a crime.

3.3 Why do people commit frauds?

The *basic driver* for committing fraud is the **potential monetary gain or benefit**. The fraud triangle is a general model that tries to define which are the drivers for fraudsters:



- Motivation: usually the need to escape from a tragic situation, or just greed
- Opportunity: the fraudsters know that a system has low security or they know someone from the inside that knows about some vulnerability to exploit
- Rationalisation: usually they perform frauds when they think that what they're doing is *not so bad*

Fraud detection and prevention systems aim to reduce as much as possible the opportunities for attackers, making attacks non-convenient.

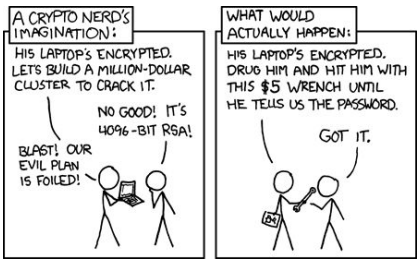
3.4 Fraud categories

- **Banking and credit card frauds:** Unauthorized taking of another's credit. They happen in two different ways:
  - **Application fraud:** the target is the company, which is convinced to release a credit card under a fake identity. Then, fraudsters spend as much as possible in a short space of time.
  - **Behavioral fraud:** the target is a user, details of legitimate cards are obtained fraudulently by stealing credentials.
- **Insurance Frauds:** Related to any type of insurance
  - Selling nonexistent policies
  - Buying policies to get paid from fake incidents
- **Corruption:** The misuse of entrusted power for private gain
- **Counterfeit:** An imitation intended to be passed off deceptively as genuine. Tipically concerns valuable objects like credit cards, IDs, popular products, money ...

- **Product warranty fraud:** A fraud that tries to exploit the warranty of a product
- **Healthcare fraud:** Filling of dishonest healthcare claims to make profit. Common in the US.
- **Telecommunications fraud:** Theft or use of telecommunication services to commit other forms of fraud
- **Money laundering:** Taking the proceeds of criminal activity and making them appear legal. Usually one of the most effective and efficient crimes
- **Click fraud:** Illegal clicks on a website's advertisements to increase the payable number of clicks
- **Identity theft:** Obtaining the personal or financial information of another person for the purpose of assuming that person's identity to make transactions or purchases
- **Tax Evasion:** Illegal act or practice of failing to pay taxes that are owed
- **Plagiarism:** Steal and pass off the ideas or words of another as own

3.4.1 Social Engineering

It is one of the main vectors to perform frauds. Attackers use human interaction to obtain or compromise information by psychologically manipulating a person into knowlingly or unknowingly giving up information.
It essentially means *hacking* into a person to steal valuable information from many sources. It consistently works, while technology vulnerabilities are hardened and solved.
The attackers usually address the weakest element of the system, which is most of the times the user. There is no patch for human stupidity.



3.4.2 Famous social engineering cases

- **Kevin Mitnick**
- **Frank Abagnale**

Phishing

Is the fraudulent process of attempting to acquire sensitive information by masquerading as a trustworthy entity in an electronic communication.

The Nigerian Prince Scam

Email requesting the victim to help a Nigerian Prince into retrieving the money from an account. Part of money laundering where the victim is the only one persecutable.

Vishing

To steal or change someone credentials by faking to be someone else and convincing their keeper

Getting dressed

People dressed in certain ways are able to convince other people surrounding them to be someone which is allowed to do something like asking money from the cashier faking to be a bank guard.

3.4.3 Frauds Impact

Frauds is often mistakenly considered to be a victimless crime, but actually it has an impact:

- A typical organizations loses 5% of its revenues to fraud each year
- The total cost of insurance fraud in the USA is estimated to be more than $40 billion per year
- Fraud costs the United Kingdom £73 billion a year
- Credit card companies lose approximately 7 cents per every hundred dollars of transactions due to fraud

There is a need to invest into **up-to-date defense infrastructures**

3.5 Anti-Fraud Strategy

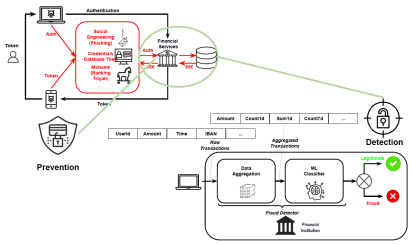
Advanced anti fraud mechanisms:

- Reduce losses due to fraud
  - Prevented & Detected
  - Fraudsters tend to look for the easy way and will look for easier opportunities

We call:

- **Fraud detection:** the process to recognize or discover frauds once they happen (ex-post approach)
- **Fraud prevention:** the set of techniques used to avoid or reduce frauds (ex-ante approach)

3.5.1 Example: Banking Fraud



The system must be defended on the financial services side, because phone and laptop are compromised.
The system tries to prevent frauds before they are executed and detect frauds after. These kind of systems are based on Machine Learning usually, automatically analizing data in order to understand what is legitimate and what is not.

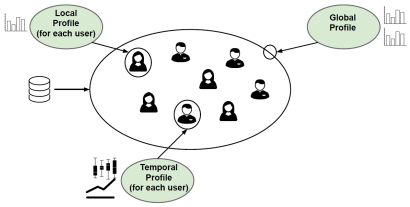
3.5.2 Fraud Detection and Prevention

They are **complementary** but not independent:

- Fraudsters adapt to the prevention mechanism changing their behavior, this impacts the detection mechanism.
- This also happens for the detection mechanism, impacting the prevention mechanism.

So what happens is that both are updated in order to impact in the good sense the other mechanism.

3.5.3 Fraud Detection approach



Data are analised by:

- A local perspective: local profile characterizes each user's spending pattern to evaluate the anomaly of new transaction
- A global perspective: global profile characterizes *classes* of spending patterns and mitigate the undertraining problem
- Temporal Profile deals with frauds that exploit the repetition of legitimate looking transactions over time

In the end a score is computed for each single transaction, scoring the risk of them to be frauds. At the end of the day, all of the fraudulent-checked transactions are personally verified by real people.
Consider that this has a cost. False positive can also make the organization to lose clients.

3.5.4 Fraud prevention mechanism

All mechanism put in place to implement strong customer authentication: a European regulatory requirement to reduce fraud and make online payments more secure.
Strong Customer Authentication requires to use at least two out of the following elements:

- Something the customer knows: such as a password
- Something the customer has: phone or hardware token
- Something the customer is: such as fingerprint

Classic technologies

One time password generators used to generate part of the code required to online payments by the use of the other part, smart cards with private keys were able to generate a password based on the user's card.
Static OTP lists were a list of codes usable to perform some operations.

Modern Technology

Nowadays software implement the same functionalities of password generators.

Attackers adapt to it

- Sim swapping attack: to obtain codes usually the user receives a PIN over sms. Attackers contact phone carriers to make them give a new SIM card by pretending to be you.
- Another attack was a social engineering attack to steal banking credentials, perform a transaction without having the code, calling you by faking to be your bank to tell you to provide the token to block a transaction while actually performing it.

3.6 Expert-based knowledge

3.6.1 Expert-based approach

The expert-based approach is based on the domain knowledge of the fraud analyst: he is able to gather and process the right information in the right manner. Performing detailed and manual investigation of suspicious cases, may indicate a new fraud mechanism, an analyst can understand and address the new mechanism.
Comprehension of the fraud mechanism allows extending the fraud detection and prevention mechanism.
Every time a new transaction is flagged suspicious, is going to be investigated, if it's labeled as fraud, the specific case will be added to the fraud detection mechanism.

3.6.2 Fraud investigation and management

There are two possible types of measures that can be taken when fraudulent activities are detected and confirmed:

- Corrective Measures
- Preventive Measures

Corrective Measures

Aim to resolve the fraud and correct the consequences. We call **retrospective screening** the set of actions to retrospectively detect frauds and subsequently address similar cases.
The sooner corrective measures are taken and fraud is detected, the more effective such measures are and the more losses can be avoided.

Preventive Measures

Aim to prevent future frauds, making the organization more robust and less vulnerable.
The typical process:

- Investigate the fraud case to understand the underlying mechanisms
- Extend the available expert knowledge with the discovered mechanisms
- Adjust the detection and prevention system

3.6.3 Rule-based engine

If-then rules are applied to future transactions and trigger an alert when a fraud may be committed.
They are based on previously detected fraud patterns.
Disadvantages:

- **Expensive to build:** each fraud case must be detected, and rules for the specific case must be written. Rules' complexity increases with fraud scheme complexity.
- **Difficult to mantain and manage:** the rule base must be kept lean and effective, every signaled case requires human follow-up and investigation.
- **New fraud patterns are not automatically signaled:** rules are based on past history, but fraud is a dynamic phenomenon and fraudsters can learn and circumvent the rules.

Rule based engine must be continously monitored, improved and updated to remain effective.

3.6.4 Fraud becomes easier to detect the more time has passed

A fraud mechanism is first successifully used:

- it will increase in the usage, fraudsters appear to be repeated offenders
- it will increase the number of this particular type of fraud, because fraudsters share knowledge

The more the scheme is used, the more it will become **evident**, becoming **statistically easier to detect**, then also similar frauds committed in the past will be discovered.
Big data can be explored and exploited for fraud detection at a lower cost $\rightarrow$ data-driven fraud detection.

3.6.5 Expert based vs Automated fraud-detection systems

Expert based systems relies on human expert input, evaluation and monitoring. They are a starting point and complementary tool to develop an effective fraud-detection and prevention system.
Automated data-driven systems require less human involvement and can lead to a more efficient and effective system, but, **expert knowledge remains crucial to build effective systems**

3.6.6 Data-driven fraud detection

- **Precision:**
  - Organizations have a limited investigation capacity
  - These approaches have increased detection power wrt classic approaches
  - They process massive volumes of information to discover frauds that are not apparent to the human eye
  - The goals of a fraud-detection system are to make optimal use of the limited investigation capacity and maximize the fraction of fraudulent cases among the inspected ones, high precision means high fraction of inspected frauds
- **Operational efficiency:**
  - operational requirements impose time constraints on the processing of a case
- **Cost efficiency:** expert based fraud detection systems are challenging and labor intensive, automated data-driven approaches are compliant with stringent operational requirements
- **Growing amount of interest:**
  - frauds have a negative social and financial impacts, which increases awareness and attention for them.
  - This makes grow investments and research from academia, industry and government.

3.7 Fraud-Detection Techniques

Frauds are hard to detect, and they remain hard and complex to detect even if fraud-detection approaches have evolved and gained significant power over the past years.

3.7.1 Fraud detection is challenging

- Where's wally? Scan of the picture to seek for the things which have particular signs which resemble the ones of Willy.
- Grid of numbers: average behavior, whatever deviates from the norm is stange and so fraudulent

3.7.2 Fraud detection techniques

Fraudsters develop advanced strategies to cover their tracks to avoid being detected. There is the need for new techniques that are able to detect and address stealthy patterns.

- **Unsupervised Learning** or **descriptive** analytics techniques
- **Supervised Learning** or **predictive** analytics techniques

3.7.3 Unsupervised learning techniques

Or descriptive analytics.

- **Unsupervised:**
  - They do not require labeled observations
  - They learn the norm from hystorical observation, **detecting anomalies means to find behaviors that deviates from the norm**
- Allow detecting **novel** fraud patterns
  - Which are different in nature from historical frauds
  - And make use of new, unknown mechanisms

This is a **complementary tool** to improve its expert rule-based fraud-detection system

Example: Telecommunications Fraud

Date (m/d)	Time	Day	Duration	Origin	Destination	Fraud
1/01	10:05:01	Mon	13 mins	Brooklyn, NY	Stamford, CT	
1/05	14:53:27	Fri	5 mins	Brooklyn, NY	Greenwich, CT	
1/08	09:42:01	Mon	3 mins	Bronx, NY	White Plains, NY	
1/08	15:01:24	Mon	9 mins	Brooklyn, NY	Brooklyn, NY	
1/09	15:06:09	Tue	5 mins	Manhattan, NY	Stamford, CT	
1/09	16:28:50	Tue	53 sec	Brooklyn, NY	Brooklyn, NY	
1/10	01:45:36	Wed	35 sec	Boston, MA	Chelsea, MA	Bandit
1/10	01:46:29	Wed	34 sec	Boston, MA	Yonkers, MA	Bandit
1/10	01:50:54	Wed	39 sec	Boston, MA	Chelsea, MA	Bandit
1/10	11:23:28	Wed	24 sec	White Plains, NY	Congers, NY	
1/11	22:00:28	Thu	37 sec	Boston, MA	East Boston, MA	Bandit
1/11	22:04:01	Thu	37 sec	Boston, MA	East Boston, MA	Bandit

Limitations of unsupervised techniques

They are prone to **deception**[2]: camouflage-like fraud strategies, because it detect new frauds only if them lead to detectable deviations from normality. These techniques need to be improved by complementing other tools.

3.7.4 Supervised learning techniques

Or predictive analytics.
Learn from historical observations to retrieve patterns that allow differentiating normal and fraudulent behavior
Aim at finding tracks that fraudsters cannot hide: *"known alarms"*. These techniques can be applied to:

- Predict frauds
- Detect frauds
- Estimate the amount of fraud

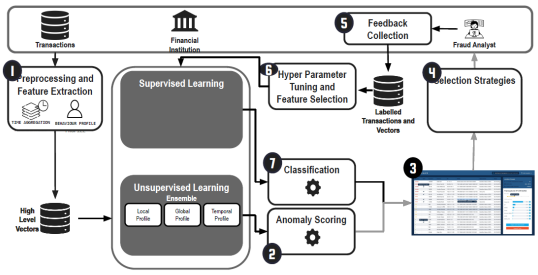
Limitations of supervised techniques

- They need historical examples to learn from *(i.e. a labeled dataset)*
- They have low detection power against different and new fraud types

[2] inganno

3.7.5 Complementarity of supervised and unsupervised methods

Use both methods in developing a powerful fraud-detection and prevention system. Each method focuses on different aspects of fraud. The expert will get a



list of transactions ranked by their anomaly scoring. He will analyse, label and use them to tune the supervised learning model.
The unsupervised learning model is used to optimise the detection capacity of the model.

3.7.6 Graph/Network Analysis

It is possible to combine the before-told supervised+unsupervised model with techinques based on graphs/network analysis.
In the financial fraud detection domain, maybe graphs are not useful, while for anti-money laundering it can be useful to have graphs of the transactions between different bank istitutions.
It extends the abilities of the fraud-detection system by learning and detecting charactersistics of fraudulent behavior in a network of linked entities.

3.7.7 Developing a fraud-detection system

It is composed by:

- Expert-based rule engine
- Unsupervised learning systems
- Supervised learning systems

The exact order of adopting the different techniques depend on the characteristics of the type of fraud:
If the company has no labeled dataset start from rules + unsupervised, for example.

3.7.8 Challenges of developing fraud-detection models

Frauds are a dynamic phenomenon

Fraudsters adapt their approaches to commit fraud without being exposed: they study fraud-detection and prevention systems to understand their functioning and discover their weaknesses.
Models must be able to adapt to it.

Good detection power

Detect fraud as accurately as possible, trying to have:

- **Low false negative rate:** this can have a financial impact for the institution
- **Low false alarm rate:** customers can even change institution if annoyed

Skewness of the data

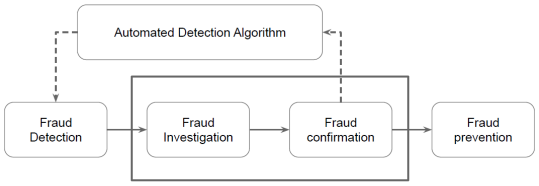
The number of available examples of fraudulent cases is usually small.

Operational Efficiency

Since you usually have a limited amount of time to detect and investigate suspicious case, the sistem must also allow to keep with the stringent operational constraints.

3.8 Fraud management cycle

This reasoning comes directly from every single step we analised in the previous section. The fraud management cycle is composed by the following parts:



- **Fraud Detection:** Applying detection models on new, unseen observations and assigning a fraud risk to every observation
- **Fraud Investigation:** A human expert investigates suspicious, flagged cases given the involved subtlety and complexity

- **Fraud Confirmation:** Determining true fraud label, possibly involving field research
- **Fraud Prevention:** Preventing fraud to be committed in the future
- **Automated Detection Algorithm:** Feedback loop: newly detected cases should be added to the database of historical fraud cases, used to learn or induce the detection model

3.8.1 Regular update of the model

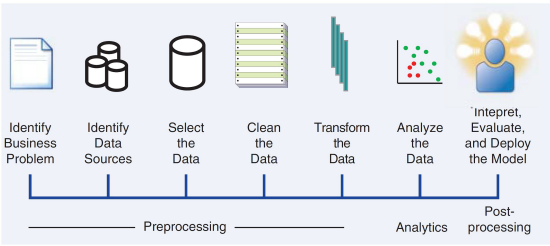
A regular update is recommendable given the dynamic nature of fraud.
With which frequency should the model be retrained or updated? It depends on several factors:

- The volatility of the fraud behavior
- The detection power of the current model
- The amount of confirmed cases already available in the database
- The rate at which new cases are being confirmed
- The required effort to retrain the model

The model must be monitored to verify its performances day by day.

3.9 Fraud Analytical Process

If you want to build a data driven model you have to follow a sequence of steps that are part of an iterative model: **The preprocessing part is always an**



important and crytical part:

- **Identify business problem**
- **Identify data sources:** data are the key ingredient to any analytical exercise
- **Select the data:** data selection has an impact on the analytical models

- **Clean the data:** get rid of inconsistencies, such as missing values and duplicate data
- **Transform the data:** additional transformations

Analytics: the analytical model is estimated on the preprocessed and transformed data. The actual fraud-detection model is built
Post-processing: the model is interpreted and evaluated by the fraud experts

3.9.1 Possible analysis output

When we build a model we want it to be able to:

- Be able to confirm and detect the same fraud present in the dataset: **trivial fraudulent patterns**
- Be able to detect novel knowledge, it must be able to generalise: **unknown patterns**

Once the analytical model has been appropriately validated and approved, it can be put into production

3.9.2 Additional consideration

Each alerted fraud must be investigated: the output of the model must be user-friendly.
To be in *the pipeline*, the model has to be integrated with other applications, part of the company system.
The model must be also monitored and updated on a regular basis

3.9.3 Key characteristics of a successful fraud analytics model

Statistical accuracy

We need to make sure that the model besides being very good at detecting frauds in the train environment, it must be able to generalise well and not overfitting.

Interpretability

A fraud-detection model must be interpretable. Model's interpretability depends on the technique used:

- **Open-box models:** it is possible to understand the underlying reasons why the model signals a case to be suspicious
- **Closed-box models:** non interpretable models

Operational efficiency

When cases need to be evaluated in real time, operational efficiency is crucial and is a main concern during model performance assesment.

Fraud Detection Costs

Developing and implementing a fraud-detection model involves a significant cost to an organization.
Cost-benefit analysis to gain insight of the returns on investment of building an advanced fraud-detection system:

- **Direct costs:**
  - Management
  - Operational
  - Equipment
- **Indirect costs (more relevant):**
  - Less usability
  - Slower performance
  - Less privacy (due to security controls)
  - Reduced productivity (slower users)

More money doesn't always mean more security.

Regulatory compliance

Depending on the context there may be internal or organization specifica and external regulation that applies to the development and application of a model, **a fraud-detection model should be in line and comply with all applicable regulation and legislation**

Chapter 4

Machine Learning for Fraud Detection

We will discuss a selection of techniques with a particular focus on the fraud practitioner's perspective

4.1 Data Preprocessing Step

Data are tipically dirty. **Data-filter mechanisms** must be applied to clean up and reduce the data, because even the littlemost mistake can make the data totally unusable and lead to a faulty model.

4.1.1 Types of data sources

When building a data driven model you may use data from different sources that provide different types of information. We will consider transactional data.

4.1.2 Transactional data

Structured and detailed information capturing the key characteristics of a customer transaction Usually summarized over long time horizons by **aggregating**:

- averages
- trends
- maximum or minimum values

On these data, the features you want to keep are:

- **Amount of money**
- **Frequency**

- **Recency**

These three features are meaningful when interpreted individually, and their interaction is very useful for fraud detection and anti-money laundering.
All these features come from expert driven analysis of the data.

4.1.3 Merging data sources

Once the correct sources are selected, them must be merged in a structured manner (i.e. a table): The rows represent the entities to analyze, while the columns contain the information about the entities.

4.1.4 Types of data elements

- **Continous:** data elements defined on an interval which can be limited or unlimited
- **Categorical:**
  - *Nominal:* can only take on a limited set of values with no meaningful ordering in between
  - *Ordinal:* can only take on a limited set of values, but with a meaningful ordering in between
  - *Binary:* can only take one out of two values

In the fraud detection domain data usually are a mixture of the three kinds of categorical types.
Preprocessing is also needed to transform all the sources in a way to make them understandable by the model.

4.1.5 Sampling

It is the process to take a subset of historical data to build an analytical model. Usually models are built on samples of the entire dataset, much smaller than the whole dataset. Why don't we use the whole dataset?

- Data must be representative for the future entities, so maybe lots of data which are not recent can cause lack in representation.
- Some features never change, maybe a small dataset would represent the same information of the whole one.

Sampling timing and bias

Trade off between:

- Lots of data (more robust model)
- Recent data (more representative model)

Sampling can introduce bias:
Which month is the more representative? (different behavior in december or february). Every month may deviate from the norm (average), two possible solutions:

- **Build separate models:** for different months or homogeneous time-frames, complex solution, multiple models to manage.
- **Build a single model:** Sample observations over a period covering a full business cycle, reduced power since less tailoring on particular time frames, but lower complexity and costs.

Stratified sampling:

In a fraud detection context data sets are very skew:

- Stratify according to the target fraud indicator to have samples containing exactly the same percentage of fraudulent/non-fraudulent transaction as in the original data
- Strafity according to predictor variables: resemble the real product transaction distribution

Notice that sampling may also introduce bias according to the considered subset of data.

4.1.6 Missing Values

Non applicable, undisclosed or generated by errors information, we must deal with them (and we see some techniques).
Some of the techniques can directly deal with missing values while others instead need some additional preprocessing.
Dealing with missing values:

- **Replace:** with known values like the mean or median
- **Delete:** if the feature is uncorrelated with the label, we can delete it
- **Keep:** some missing values can be meaningful and may have a relation with frauds, so they must be considered as a separate category

Perform a statistical test to understand if the missing value is correlated or not with the target label, this will impact the choice made to cope with these values.

Example:

- Here for example is possible to delete row 6 because of the number of missing values.

ID	Age	Income	Marital Status	Credit Bureau Score	Fraud
1	34	1,800	?	620	Yes
2	28	1,200	Single	?	No
3	22	1,000	Single	?	No
4	60	2,200	Widowed	700	Yes
5	58	2,000	Married	?	No
6	44	?	?	?	No
7	22	1,200	Single	?	No
8	26	1,500	Married	350	No
9	34	?	Single	?	Yes
10	50	2,100	Divorced	?	No

- Many missing values are present also in the credit bureau score, but in this case we see that the rows where it is inserted are labeled as frauds, it is better to keep the missing values or to replace them with an average or something.
- Marital status for entry 1 will be missing: if we use our knowledge to put a value, we risk to introduce bias, an alternative solution can be to use the mode of the dataset.

4.1.7 Outliers

Outliers are extreme observations that are very dissimilar from the rest of the population.

- Valid: boss salary is $ 1,000,000
- Invalid: age is 300

In this specific part of the processing step we want to identify the outliers that may influence too much our model, we call:

- **Univariate Outliers** the ones which outly in one dimension
- **Multivariate Outliers** the ones which outly in multiple dimensions

Univariate Outlier Detection and Treatment

Minimum and maximum values for each of the data elements.

- Graphical Tools:
  - Histograms (grafico a torre): check for values less frequent from the average, if included they may introduce bias
  - Box Plot: It is a plot which represents three key quartiles of the data to find values far from them

- Z-scores: measures how many standard deviations and observation lies away from the mean, computes the difference of the sample from the mean, and then divides for the std deviation

Multivariate Outlier Detection and Treatment

- Fit regression lines and inspect the observations with large errors
- Make some statistical test to understand what to do with them

How to deal with outliers

Various schemes:

- Invalid observations can be treated as missing values
- Valid observation can be capped to a maximum or minimum value

Expert-based outlier detection

Not all invalid values are outlying and may go unnoticed if not explicitly looked into.
Apply some rules based on expert knowledge to data to check and alert for issues, as existing relations between the different variables

Example:

- Birthdate: 1/01/1980
- Category: Child

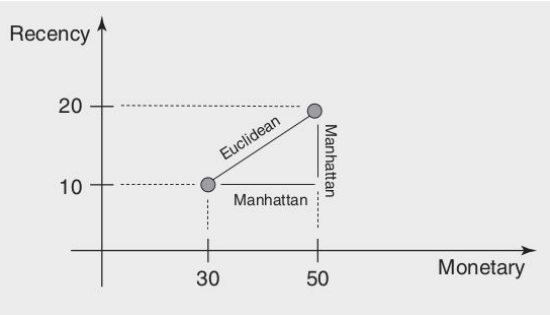
We don't know which of the two is the invalid value, but we know that the pair is. Both values are not outlying and therefore explicit precautions must be taken to notice.

4.1.8 Standardizing Data

Euclidean vs Manhattan distance

- Euclidean:
$$\sqrt{(1500-1000)^2+(10-5)^2}=500$$
- Manhattan:
$$|1500-1000|+|10-5|=505$$

Warning: the distance between monetary is more determining than the one for recency, there is the need to standardize.



Standardization

Means to scale variables to a similar range, multiple techniques are available:

- **Min/Max standardization:** impose a new max and a new min and scale accordingly
- **Z-score standardization:** calculate the z-scores
- **Decimal scaling**

4.1.9 Categorization

For categorical variables, it is needed to reduce the number of categories.
Basic methods:

- **Equal interval binning:** create bins with the same range
- **Equal frequency binning:** create bins with the same number of observations
- **Chi-squared analysis**
- **Pivot table**

4.1.10 Variable Selection

Typically, only few variables contribute to the prediction. We need to select only the helpful ones.
Usually, in fraud detection domain, the number of variables is in the range of 10-15, how do we find them?

Filters

It is a selection mechanism based on statistical tests. Allow a quick screening of which variables should be retained for further analysis. They allow reduction in the number of dimensions of the data set early in the analysis. Drawbacks:

- They do not consider correlation betwen dimensions

Principal Component Analysis

Dimension reduction by forming new linearly independent variables, which are linear combinations of the original set, which have the property to better describe the distribution.
The new variables are called principal components. The variance contained in the original data can be summarized by a limited number of principal components. Left out the ones which account for a very small fraction of variance.
Limitations:

- **Reduced interpretability:** a transaction may be labeled as fraudulent whithout us knowing why because of the strange form of features.

4.2 Unsupervised Learning for Fraud Detection

Unsupervised learning aims at finding anomalous behavior deviating from the **norm**.
The challenge is to find a way to correctly model the norm, like the behavior of the average customer at a snapshot in time or the average behavior of a given customer across a time period.

4.2.1 Unsupervised learning = anomaly detection

It aims at finding anomalies, useful when:

- Organizations are starting doing fraud detection
- No labeled historical dataset available
- Fraudsters are continously adapting

4.2.2 Unsupervised Learning Challenge

Define the average behavior or norm:

- it depends on the application field
- Boundary between norm and outliers is not clear-cut (fraudsters try to blend into norm)
- The norm may change overtime

Anomalies do not necessarily represent frauds, unsupervised learning for fraud detection requires extensive validation of the identified suspicious observations

4.2.3 Basic tasks to find anomalies

- **Graphical outlier detection:** find outliers with histograms or box plots for the one-dimensional, with scatter plots for the multi-dimensional
  - Disadvantages: less formal and limited to few dimensions, requires active end-user involvement, very difficult to perform for a large dimensional dataset
- **Statistical outlier detection:**
  - Use z-score: if the z-score is bigger than 3 consider outliers
  - Fit a distribution: outliers are the observations with small values for the probability density distribution
  - **Break Point Analysis**
  - **Peer Group Analysis**

Break Point Analysis

It is an **intra**-account fraud detection method. A **break point** is a sudden change in account behavior.

- Define a fixed time window
- Split it into an *old* and *new* part
- Compare the new part with the old part
- The old part represents a local profile against which new observations are compared

Peer Group Analysis

It is an **inter**-account fraud detection method. A **peer group** is a group of accounts that behave similarly to the target account.
When the behavior of the target account deviates substantially from the peers, an anomaly can be signaled.

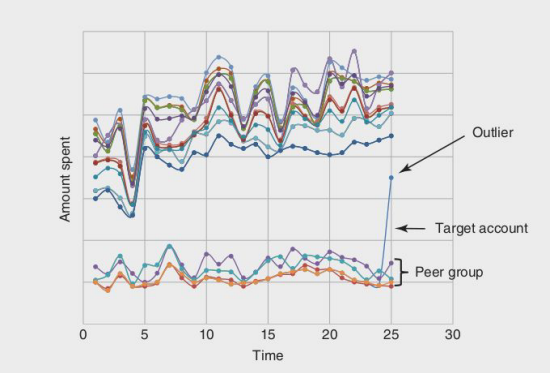
- Identify the peer group of a particular account
  - use prior business knowledge
  - or statistical similarity metrics
  - define the number of peers (too small means sensitive to noise, too large means insensitive to local important irregularities)
- Target account behavior is contrasted with its peers
  - With statistical tests
  - or distance metrics

Example: Credit card fraud

Weekly amount time series:

$$y_1, y_2, \ldots, y_{n-1}, y_n$$

. Verify whether the amount spent at time n is anomalous.



- **1:** identify the k peers of the target account
- **2:** compare the behaviors

Break Point vs Peer Group analysis

Both will detect local anomalies rather than global anomalies. They both have issues with seasonality.

4.2.4 Clustering

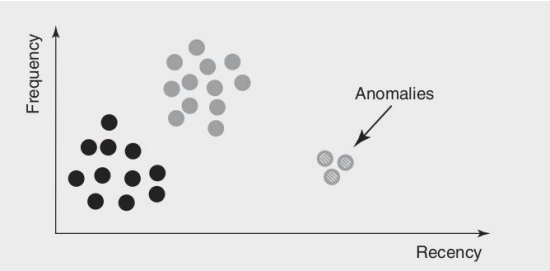
Technique that tries to divide the dataset into clusters (groups) that maximize:

- Cohesion: homogeneity within a cluster
- Separation: heterogeneity between clusters

Each cluster should contain elements very similar between them and very different from other groups' elements.

Clustering for fraud detection

We want to detect frauds as deviation from the norm. Frauds can be seen as elements inside the corpus which do not belong to any cluster, or that belong to very small clusters far away from the others.
Important: we're not 100% sure that them are frauds, they're just outliers, so they need to be investigated to understand what they are.
An efficient unsupervised algorithm tries to minimize as much as possible the number of false positives



Distance metric

To group elements toghether we need to define a similarity measure:
at the end of the day each element of the cluster must be transformed in vectors of numbers, to understand if two entities are similar we need to compute distance between them.
Possible distance measures can be: euclidean, minkowski or manhattan, mahalanobis *(when high variance in dataset)*

Distance metric: Continous vs Categorical

- **Continous Variables:**
  - Euclidean metric
  - Pearson correlation or cosine measure
- **Categorical Variables:** (binary variables 01010101)
  - Simple matching coefficient:
    compute the number of identical matches between the variable values
  - Jaccard index: measure the similarity between both claims across those red flags raised at least once

4.2.5 Clustering Techniques: Hierarchical clustering

Two main strategies: **agglomerative vs divisive**.

- **Agglomerative strategies:** assign each single element to a cluster, then agglomerate small clusters in greater ones
- **Divisive strategies:** split a great cluster in smaller clusters until each element is assigned to a single cluster

The final result is usually in between of the two extremes. But to agglomerate/divide, how to compute the distance between clusters?

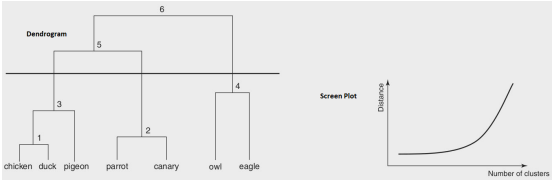
- **Single Linkage:** consider the distance between the closest elements in the two clusters

- **Complete Linkage:** consider the distance between the furthest elements in the two clusters
- **Average Linkage:** consider the average distance between all elements of the two clusters
- **Centroid Method:** elect an element as the representative of the cluster (centroid), compute the distances between centroids

Number of Clusters

To decide the optimal number of clusters:

- **Dendrogram:** tree-like diagram that records the sequences of merges. Cut the dendrogram at the desired level to find the optimal clustering
- **Screen Plot:** plot the distance at which clusters are merged, the optimal clustering is indicated by the elbow point



Hierarchical clustering: conclusions

- **Advantages:**
  - The number of cluster doesn't need to be specified prior to the analysis
- **Disadvantages:**
  - Do not scale well to large datasets
  - The interpretation of the clusters is often subjective and depends on the business expert

4.2.6 Clustering Techniques: Non-hierarchical clustering

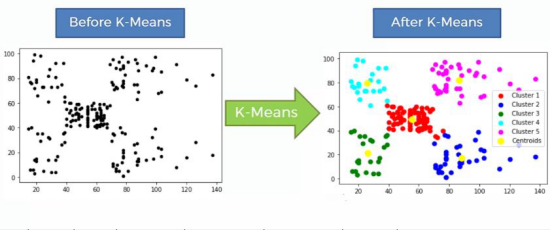
K-means

- Select the number of clusters k
- Select k observations as initial cluster centroids
- Assign each observation to the cluster with the closest centroid according to some distance

- When all observations have been assigned, recalculate the centroids
- Repeat until the cluster centroids no longer change

Limitations:

- The number of cluster must be specified before the start of the analysis, it can be choose by experts or computed as result of other clustering techniques
- It is sensitive to outliers, relevant in fraud detection setting



SOM: Self Organizing Maps

It is an unsupervised learning algorithm that allows to visualize and cluster high-dimensional data on low dimensional grids of neurons:
It is a feedforward neural network with an input and an output layer.
Each input is connected to all neurons in output layer with weights, all randomly initialized.
Functioning:
When a training vector x is presented, the weight vector of each neuron is compared with x. The most similar neuron in euclidean sense is called the *BMU: best matching unit.*
Then update all the weights based on BMU: move them towards the point, trying so to match the distribution of the original points.

- **Advantages:** it is able to automatically cluster data, results can be visualized and interpreted

4.3 Supervised Learning for Fraud Detection

Technique that assumes the availability of historical labelled data. With this technique is possible to identify **known fraudulent patterns**.
The aim is to build an analytical model predicting a target measure of interest, for example classifying a new instance as fraudulent or not.
The target fraud indicator is usually hard to obtain:

- Typically not provided by the collaborating entity
- They are not noise-free, it is crucial to avoid overfitting

4.3.1 Regression vs Classification: Target variables

- **Regression**
  - Continous target variable
  - Varies along a predefined limited or unlimited interval
- **Classification**
  - Categorical target variable
  - It can only take on a limited set of predefined values

4.3.2 Linear Regression

Technique to model a continous target variable.
The goal is to find the **best fit line** that can accurately predict the output of a continous variable.
General formulation:

$$y(x, w) = w_0 + \sum_{i=1}^{N} w_i x_i$$

The weights w_i measure the impact on the target variable y of each of the individual explanatory variables.
The weights can be estimated by minimizing a cost function, for example the **squared error function** (OLS: Ordinary Least Squares).

4.3.3 Logistic Regression (classification)

When the target variable is categorical, linear regression gives us no guarantees to have the target in between 0 and 1.
With an S-curved function like the sigmoid, we can bound the values of the target variable.
Logistic regression results can be interpreted as probabilities, the optimization method is Maximum Likelihood Estimation.

4.3.4 Variable Selection

Different methods to perform variable selection:

- **Statistical tests:** to verify if the coefficient of a variable is significantly different from zero, a low p-value represents a significant variable
  - Linear Regression: Student's t-distribution
  - Logistic Regression: Chi-squared distribution
- **Forward regression:** start from the empty model and always add variables

- **Backward regression:** start from the full model and always remove variables
- **Stepwise regression:** mix the two

Evaluation criteria for variable selection

- Statistical significance (p-value)
- Interpretability
- Operational Efficiency
- Legal Issues

4.3.5 Decision Trees

They are recursive-partitioning algorithms, based on a tree-like structure which can represent patterns in an underlying dataset.
Each top node (root node) specifies a testing condition, of which the outcome corresponds to a branch leading up to an internal node. The terminal nodes (leaves) assign fraud labels.
Implement three main functionalities:

- **Splitting decision**
- **Stopping decision**
- **Assignment decision**

Splitting Decision

Which variable to split at what value *(e.g. transaction amount is*

$$> \$100.000$$

or not).
To understand the concept, **impurity or chaos:**

- **Minimal impurity:** occurs when all customers are either good or bad
- **Maximal impurity:** occurs when one has the same number of good and bad customers

Kinds of impurity measures:

- entropy: bounded in 0 and 1
- gini: bounded in 0 and 0.5

They can be both used with no great differences. The aim of decision tree is to minimize one of these measures (impurity).
By the end of the day the algortim selects a sequence of candidates for a splitting decision and then selects the one that minimizes the impurity.
This approach considers different candidate splits for its root node. It is a greedy and recursive strategy by picking the one with the biggest gain *(gain = decrease in impurity)*.
It can be perfectly parallelized.

Stopping Decision

When to stop adding nodes to the tree. If the tree continues to split we'll reach the situation where there is one leaf node per observation (overfitting).
To avoid overfitting train-validation. Stop adding nodes when the validation test curve reaches its minimum.

Assignment Decision

What class *(fraud/not fraud)* to assign to a leaf node.

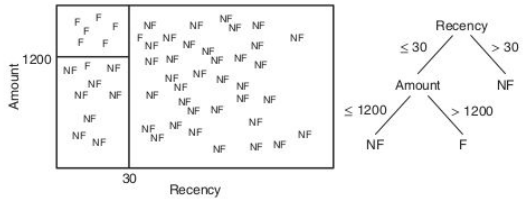
4.3.6 Decision Tree Properties

Rule set

Decision trees are interpretable models, they can be represented as a rule set: every path from a root node to a leaf makes up a simple if-then rule.
These trees can automatically extract rules, they're a tool which can help experts in build rules: no more manual update.

Decision Boundaries

Decision tree essentially model decision boundaries orthogonal to the axes



4.3.7 Using Decision Trees in Fraud Analytics

These mechanisms can be used to interpret the result of other machine learning techniques, for example to interpret clustering solutions:

given the clustering solution, build a decision tree with the cluster output interpreted as label, trying to explain it.
Advantages:

- White-box models with clear explanation: interpretable
- Operationally efficient
- Powerful techniques that allow for more complex decision boundaries than logistic regression
- Non-parametric

Disadvantage:

- Highly dependent on the sample that was used for tree construction

4.3.8 Neural Networks

Mathematical representations inspired by the functioning of the human brain.
Can model very complex patterns and decision boundaries in the data. They are a generalization of logistic regression, each neuron uses sigmoid as activation function.

MultiLayer Perceptron

Sequence of neurons organized in different layers:

- Input layer
- Hidden layers
- Output layer

Since in fraud analytics there are no complex patterns, neural networks with just one hidden layer are already good.

Weights Learning

The optimization is more complex: iterative algorithm that optimizes a cost-function

- For regression: MSE
- For classification: Maximum Likelihood

Start from a set of random weights, and iteratively adjust them to the patterns in the data using an optimization algorithm (backpropagation). The curvature of the objective function is not convex, to avoid local minima start with different randomizations of the weights and select the one with the best performances. Stop after a certain number of epochs without *progess*.

Overfitting

Options to mitigate overfitting:

- train test and stop training when you notice an increase of classificaiton error
- weight regularization: put some bound on the value of each weight is going to assume, because neural network work updating weights associated to hidden neurons, bigger weights give importance to something that don't.

Variable Selection and Interpretability

Neural networks are black box: relate input to output in a mathematically complex, non-transparent and opaque way.
In fraud domain, we need to perform variable selection without removing what can be important to our model: **Hinton Diagram**. The Hinton diagram visualizes the weights between the inputs and the hidden neurons as squares, where the size of the square is proportional to the size of the weight. It is possible to inspect the diagram and remove the variable whose weights are closest to zero. Another way is to use backward variable selection, by removing features until we start to see a decrease of performance.

Rule Extraction Procedure

To extract if-then classification rules, mimicking the behavior of the neural network.

- **Decompositional technique:** decompose the network's internal workings by inspecting weights and/or activation rules.
- **Pedagogical technique:** consider the neural network as a black box and use its predictions as input to a white-box analytical technique such as decision trees.

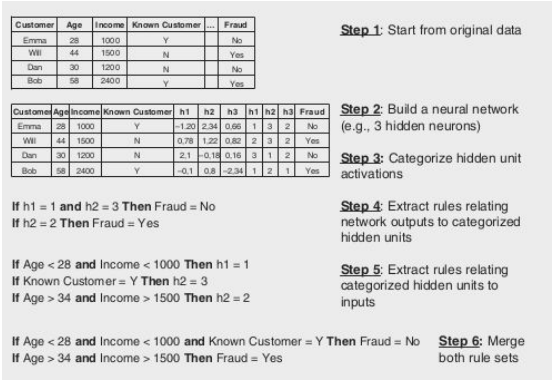
Decompositional Rule Extraction Example

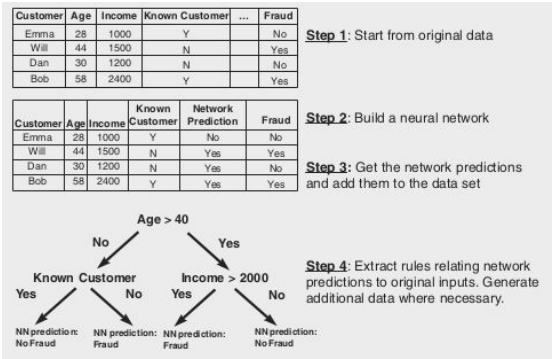
Pedagogical Rule Extraction Example

Rule Extraction Approaches Evaluation

Rule-sets must be evaluated in terms of:

- **Accuracy**
- **Conciseness**
- **Fidelity:** measures to what extent the extracted rule set succeeds in mimicking the neural network and is calculated as follows.





4.3.9 Two-stage model setup

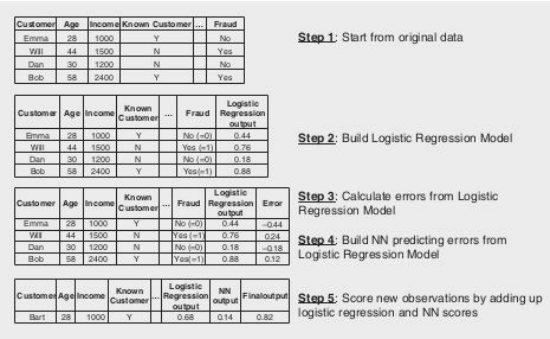
- Estimate an easy-to-understand model first *(e.g. linear regression, logistic regression)*, this will give us the interpretability part
- Use a neural network to predict the errors made by the simple model using the same set of predictors
- Both models are then combined in additive way

4.3.10 Support Vector Machines

Method that originates from linear programming:

- Objective function
- Constraints

SVMs aim at maximizing the margin to pull both classes as far apart as possible.
We call **support vectors** the training points that lie on the separating hyperplanes.



The optimization problem has a quadratic cost function, no local minima and only one global minimum.
In the real world it exists the non-linear separable case, add an error term to consider how much are we misclassifying, use kernel function to change feature space.

Interpretability and Variable Selection

They are black box methods, complex in settings where interpretability is important.
Variable selection can be performed using backward variable selection (this reduces the variables but not provide any additional insight into the workings of the SVM).

Rule Extraction

- **Decompositional:** represent SVM as a neural network
- **Pedagogical:**
  - Use SVM first to construct a dataset with SVM predictions for each one of the observations
  - Give the dataset to a decision tree algorithm to build a decision tree
- **Two stage models:** A simple model is estimated first, followed by a SVM to correct the errors of the latter

4.3.11 Ensemble Methods

The idea is to provide an explanation of the target label by putting toghether different models combining their results, based on the assumption that multiple diverse models capture different trends of the dataset.
To be successful, ensemble must be done with models sensitive to changes.

- **Bagging**

- **Boosting**
- **Random Forests**

Bagging

- Start by taking B bootstraps from the underlying sample. A bootstrap is a sample with replacement.
- Build a model for every bootstrap
- Use majority voting for classification, average for regression

The key element for bagging is the instability of the analytical technique. For models that are robust with respect to the underlying dataset, bagging will not give much added value.

Boosting

Estimate multiple models using a weighted data sample *(uniform at the beginning)*.
Iteratively re-weight data according to classification error.
The idea is that difficult observations should get more attention. The final ensemble is a weighted combination of all the individual models.

- Key Advantage: easy to implement
- Potential Drawback: risk to overfitting to the hard *(potentially noisy)* examples in the data, which will get higher weights as the algorithm proceeds. Relevant in fraud detection setting because the target labels are tipically quite noisy.

Random Forests

Creates a forest of decision trees:

- Given a dataset with n observations and N inputs
- m = constant chosen on beforehand
- For t = 1,. . . ,T:
  - take a bootstrap sample with n observations
  - build a decision tree whereby for each node of the tree, randomly choose m variables on which to base the splitting decision
  - Split on the best of this subset
  - Fully grow each tree without pruning

Create as much diversity in the classifiers, the higher the diversity, the higher the performances.

Evaluating Random Forests

Random forests can achieve excellent predictive performances.
Their main disadvantage is that they are black-box models because they're based on random decision trees.

4.4 Evaluating a Fraud Detection Model

When evaluating predictive models, two key decisions need to be made:

- **On the dataset split up**
- **On the performance metrics**

4.4.1 Splitting up the dataset

Large dataset

The decision on how to split up the dataset for performance measurements depends on its size.
Split large datasets into:

- (70%) Training dataset
- (30%) Test dataset

There must be **strict separation** between training and test sample.
In the case of decision trees or neural networks separate the training dataset in 40% training, 30% validation.
Stratified split-up ensures that fraudsters/non-fraudsters are equally distributed amongst the various samples.

Small dataset

With small datasets special schemes need to be adopted

- **Cross-validation:** split data in K folds, train models with K-1 folds and test on the remaining one. It will result in K performance estimates that then are averaged.
- **Leave-one-out cross-validation:** every observation is left out in turn and a model is estimated on the remaining K-1 observations, this gives K analytical models in total.

Cross-validation gives multipe models, how should the final model be?

- All models collaborate in an ensamble
- Pick one LOO model at random, they will be quite similar anyway
- Build one final model on all observations but report the performance coming out of the cross-validation

4.4.2 Performance Metrics

- **Classification Accuracy:** percentage of correctly classified observations:

$$\frac{TP + TN}{TP + FP + FN + TN}$$

- **Classification Error:**

$$\frac{FP + FN}{TP + FP + FN + TN}$$

- **Sensitivity, Recall or Hit Rate:** how many of the fraudsters are correctly labeled as fraudster:

$$\frac{TP}{TP + FN}$$

- **Specificity:** how many of the non-fraudsters are correctly labeled as non fraudsters

$$\frac{TN}{FP + TN}$$

- **Precision:** how many of the predicted fraudsters are actually fraudsters

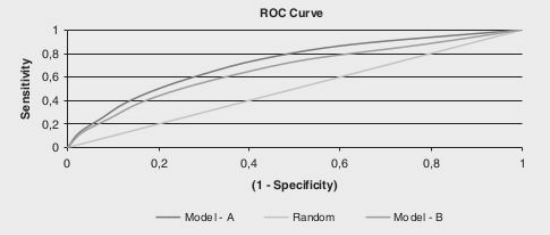
$$\frac{TP}{TP + FP}$$

Usually accuracy is not a good measure in fraud detection domain because frauds are rare, disperse in a huge amount of data.
Frauds represent 0.1 percent of data, if you say that everything is legitimate you will still get an accuracy of 99,9.
The good measures are:

- F1 measure: armonic mean of precision and recall
- Matthew Correlation Coefficient
- Or other graphical measures and precision/recall.

Receiver Operating Characteristic (ROC) Curve

It plots the sensitivity vs 1-specificity: a good model will be near to the extreme



left point of the graph.
A problem arises if curves intersect, **AUC (Area under the ROC curve)** provides a simple figure-of-merit for the performance:
the higher the AUC, the better the performance. It interprets the probability that a randomly chosen fraudster gets a higher score than a randomly chosen nonfraudster.
A good classifier should have a ROC above the diagonal, and AUC bigger than 50%.

4.4.3 Developing predictive models for skewed datasets

Fraud detection datasets often have a very skew target class, frauds are

$$\leq 1\%$$

. This creates problems for the analytical techniques.
It is recommended to increase the number of fraudulent observations or their weight, such that the analytical techniques can pay better attention to them.
Increase the number of frauds by:

- Increasing the time horizon for prediction: instead of six-month, 12-month.
- Sample every fraud twice or more
- Undersample the majority case

Oversampling

Replicate frauds two or more times to make the distribution less skew

Undersampling

Remove nonfrauds two or more times to make the distribution less skew, remove low value transactions or inactive accounts.

Under and Over sampling

They can be combined, they should be performed only on training data and not on test data to give an unbiased view on model performance.

SMOTE: Synthetic Minority Oversampling Technique

Works by creating synthetic observations based on the existing minority observations.
Synthetically generate new samples from the ones available, identify the k nearest neighbors from the sample to replicate and synthetically generate the new one averaging the neighbors.

Cost-Sensitive Learning

Besides interpretability, efficiency is the main problem:
output is going to be investigated. We can estimate the cost associated to the misclassification

Assigns higher misclassification costs to the minority class

		Predicted Class	
		Positive	Negative
Actual class	Positive	C(+,+)	C(–,+)
	Negative	C(+,–)	C(–,–)

Usually C(+, +) = C(–, –) = 0, and C(–, +) > C(+, –).
Minimizing the misclassification cost during classifier learning.

$$\text{Total cost} = C(-,+) \times \text{FN} + C(+,-) \times \text{FP},$$

Part III

Digital Forensics principles

Chapter 5

Introduction to Digital Forensics

5.1 What does forensics mean?

- **Forensics** is the application of scientific analysis to reconstruct evidence.
- **Digital Forensics** is one of the disciplines of forensics, is the application of scientific analysis methods to digital data, computer systems, and network data to reconstruct evidence.

Forensics is strongly related with laws: different jurisdiction means different procedures.

5.2 The Daubert standard (USA)

Generally speaking, we can have two sources of evidence:

- physical evidence
- eyewitness testimony

To provide physical evidence, an expert is allowed to give an opinion in court because of the fact that he is an expert, and the use of scientific methods is the reason why he is listened to.
How do we define what an expert is?

5.2.1 The Daubert Standard

We define as expert witness someone who has witness because of its own experience.
The Daubert standard define an expert witness as follows:

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

- The expert's scientific, technical, or other specialized knowledge will help the trier of fact[1] to understand the evidence or to determine a fact in issue;
- The testimony is based on sufficient facts or data;
- The testimony is the product of reliable principles and methods; and
- The expert has reliably applied the principles and methods to the facts of the case.

Reliable principles and methods means that they are scientifically found. You need an expert to use scientific method to establish those facts:

5.2.2 What is scientific? (Italy)

- Galileo: scientific means repeatable, you actually make an experiment to demonstrate that something can happen
- Popper: scientific means falsifiable, if you're able to create, at least in your mind, an example which proves the opposite of a statement, then the statement is scientific.
  - Example: *Stefano is a kind person* is not a scientific statement, because it is not falsifiable.

Some of the questions are not falsifiable, hence not scientific. An example is a criminal which keeps images of kids on his own computer. Did he do it willingly or not? This is not falsifiable.
What is possible to do is to look at the folder to see if it was opened multiple times, look at the file to see if it was in the browser's cache or in a specific folder on the criminal's computer.

5.2.3 Daubert Test for scientific

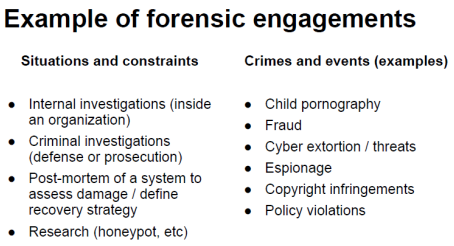
Factors to consider (USA) to establish if something is scientific or not:

- Whether the theory or technique employed by the expert is generally accepted in the scientific community
- Whether it has been subjected to peer review and publication
- Whether it can be and has been tested
- Whether the known or potential rate of error is acceptable; and
- Whether the research was conducted independent of the particular litigation or dependent on an intention to provide the proposed testimony.

[1] the judge

5.3 Example of forensic engagements

We do forensics for different situations in different contexts



Depending on which part of the lawsuit you're working for: *("prosecutor", "judge", "lawyer")*, the things that you can do are different.
The procedures has always to be contextualized in what are the purposes and which constraints the purpose has.
Crimes:

- The first investigated crime is the one that has to do with children. The reason why is that it is one of the easiest to prosecute.
- Fraud is the second one. Prosecuting fraud is difficult because there is a large amount of them which are perpetrated by people who live in countries where the laws kind of permit it. It is still denounced because people can get their money back.
- Cyber extortion is the crime which consists into stealing personal data to get money from them. Extortion happens with family, work related, revenge cases.

There exist also a lot of non-cyber crimes which involve digital components:

- Search for traces in digital devices in murder cases
- Tracking or geo-localization of mobile devices

Digital components can be fundamental.

5.4 Phases of an investigation (Pollitt)

- Source acquisition: how we preserve digital evidence
- Evidence identification: how we analyze digital evidence
- Evaluation: how we take evidence and pack with the specific case
- Presentation: how we put togheter all of this in a court

Part IV

Acquisition, analysis, evaluation and presentation of evidence

Chapter 6

Acquisition

Forensics was born in the USA, it borders with law so it was developed with US laws in mind.
The legal system of USA and the EUs ones are extremely different, and so is the court approach:

- In the U.S. most of the cases are tried by juries of peers, the judge works to make the court work. In fact, the judge is the one to decide if evidece is admissible or not. If not admissible, that could not be talked about from jury and lawyers or taken into account. Admissibility of scientific evidece is completely based on the concept of **chain of custody**: *tracking where the evidence was taken, who was in custody, where was it stored, who analyzed it, what was done to it.* If that is broken, evidence becomes inadmissible.
- In Italy is the judge the one who takes the decision. The jury exists only in Corte d'Assise and it is made by people extracted from a certain set. Is the judge to evaluate the admissibility of the evidence, it is inadmissible only if it was taken in violation of laws.

By the way, the Counsil of Europe has an international law agreement called **Convention of Budapest on cybercrime**, it was written in 1999, and was incorporated in the Italian law in 2008. Also international standards from ISO/IEC can be applied.

6.1 Brittleness of digital evidence

Digital evidence is **brittle**[1], it is because if it is modified there is no way to tell. In other words, it is not **tamper evident**[2]. This means that there is no way to say if the chain of custody was violated.
Digital evidence can also be fake created:

[1] fragile
[2] non mostra segni di manomissione anche se manomessa

- By changing for example the clock of a computer, it is possible to create a fake file which was modified in a different time, and there is no way to figure out if that happened.

We need an entire process of acquisition to create ways to make digital evidence as far as possible tamper evident, there is a need to ensure:

- Legal compliance: evidence must comply with the laws.
- Ethical behavior from all parties: even the police can act unhetical
- Detection of errors in good faith
- Detection of natural decay

6.2 The usage of hashes in digital forensics

Hash function are used to record the state of an object at a given step of acquisiton. It is constantly checked to ensure authenticity and non-tampered state of that at any further phase of acquisition. They are used to *"freeze"* the crime scene.

- We don't know what happened before that first step
- We can only tell if something has been tampered, but not what has been tampered. We also cannot restore data if tampered.
- Hashes are used to prove that something has been modified, but cannot tell what.

Since evidence is going to be stored on a certain media, hashes must be put in another place *(e.g. writing them on a physical register is the most frequent, or use digital signatures).*

6.3 Hardware and software for acquisition

- Hardware:
  - Removable HD enclosures or connectors with different plugs
  - Write blocker
  - External disks
  - USB, firewire, SATA, e-SATA controllers if possible
- Linux: extensive native file system support, ease of accessing drives and partitions without mounting them

6.4 Bitstream images

What we want to acquire, if possible, is a **bitstream image**, a bit-by-bit clone of the original evidence media. The main reason is because if we only copy the allocated content we potentially lose information stored in the unallocated part of the disk. This is called a **forensic clone** or clone copy or image of the disk.

6.5 Basic procedure of acquisition

- Disconnect the media from the original system, if possible
- Connect the source media to the analysis station, if possible with a write blocker
- Compute the hash of the source:
  - Linux: *#dd if=/dev/sda conv=noerror,sync — sha256sum*
  - conv=noerror means *"keep going even if the system generated some errors"*
- Copy the source:
  - Linux:
  - *#dd if=/dev/sda of=/tmp/acquisition.img conv=noerror,sync*
- Compute the hashes of the source and the clone
  - Linux: *#dd if=/dev/sda conv=noerror,sync — sha256sum*
  - *#sha256sum /tmp/acquisition.img*
- Compare the three hashes

If the hashes are different, it means that the copy didn't happened correctly or that you tampered with the drive, or that some damaged block read broke it, or that some damaged block reads everytime something different ...
It could be also good to compute MD5 and SHA-1 hashes, for redundancy and backward compatibility.

6.5.1 Challenges: time

A typical hard drive capacity today is 1TB. Transfer speeds are in the order of 600MB/s *(SATA 3)*, but mechanical drives reach an average of 80MB/s *(SSDs are actually fast as SATA3)*, USB connectors are up to 100MB/s.
This means that for a 1TB drive you can expect to wait several hours to complete a copy or to run a hash. Tools like **embedded duplicators** or software like dcfldd can make this run in parallel to save time.

6.5.2 Challenges: size

File systems limit to 4TB mostly for the size of a file. If a drive is larger, it has to be spliced in more files.
Space is also needed to keep data for all the devices used in an investigation, which can be in the order of the hundreds of terabytes, that's why NAS or SAN systems are common in forensic shops.
Sometimes it can be also useful to move images across a network, for exaple by using netcat to listen for a stream generated by the computer which is taking the acquisition.

6.5.3 Challenges: encryption

Many machines use encryption. In some cases the key for decription is stored on the motherboard of the computer, and so the image became useless without the computer.

6.6 Alternative procedures

6.6.1 Alternative 1: booting from live distribution

Sometimes it is mandatory to work directly on the actual machine, because of systems with weird connectors, RAID devices or because of specific investigation constraints.
In these cases it is possible to live-boot the system using a Linux distribution targeted for forensic analysis, like *Tsurugi* or *BackBox*[3].
This process has to be performed with great care, different systems have different ways to access the boot menu, and if by chance the original operating system is booted, it may be cause a loss of useful data.

6.6.2 Alternative 2: Target powered on

For many reasons can happen that the target machine is turned on:

- Maybe it is running a critical service for the company we're working for
- Maybe it was seized while turned on

Working with a powered on machine is called **live forensics**. Can we turn it off? Probably not if the machine is a critical one. Should we turn it off? Maybe not if we're trying to do live analyisis of an intruder. We want to exclude an intruder? We can disconnect the network.
Since the machine is turned on, our actions are going to modify its state, our commands, even turning the machine off will change it, because of the operating system's operations.

[3]The reason why we use specific distribution, is because we do not want to overwrite linux swap partitions on the media

Now the prospective changed, before we thought of evidence as immutable as possible, now we don't.
Since our actions are going to tamper with the state of the system, we need to **document what we did** *(to preserve chain of custody (USA))*, here we don't have the safeguard of the hash. Work in volatility order:

- Dumping the memory may be useful, for example most of the break-in attacks are file-less.
- So is saving runtime information: network, process information, etc...
- Consider encryption before turning the machine off *(maybe the device can become unusable after)*
- Finally, disk acquisition

If possible, perform the acquisition without a shutdown, if not, if the machine has to be reboot after, shut it down properly. If it is not crytical it is possible to pull the plug to not tamper the disk.

Useful commands

- Network data: *"ifconfig -a ; netstat -anp ; route -n ; arp"*
- Process data (to store process information): *"ps aux ; lsof file"*
- Users data: *"who; last; lastlog"*
- Memory acquisition: *"Mantech mdd, win32dd, Mandiant Memoryze"*

6.6.3 Alternate 3: live network analysis

In enterprise environments if we want to observe an attacker *live* whithout him being scared of our actions, we can use two observation points that are outside of the machine: logs and network traffic *(we'll have a separate class)*

6.6.4 New Challenges (separate classes)

- Forensics in cloud environments
- Mobile forensics
- SSD drives work different than magnetic drives

Chapter 7

Analysis or Identification

The purpose of Identification is to perform actions that can fall anywhere in computer science.
We'll look at either the methodologies and specific forensics cases.
Tipically we use Linux:

- extensive native file system support
- we can just dump a disk image as a file with linux, and we can mount the copy as if it were a drive, also read only to prevent writing and keep the copies not touched.

Why not Windows? Nobody uses windows by itself. It must be confined, it stinks.

- It tampers with drives by automatically mounting them and writing on them.
- No image handling or hotswapping of drives.
- No support for non Windows filesystems.

The best thing to do is to use linux as host with windows as guest on virtual machines:

- Work the images with Linux, mount them read-only and exporting them via Samba or vmware *"not persistent"* to Windows.
- Use specific windows tools

Keep in mind that sharing something like this, is doing it file level: any utility that analize drives cannot work in this way.

7.0.1 What does scientific mean?

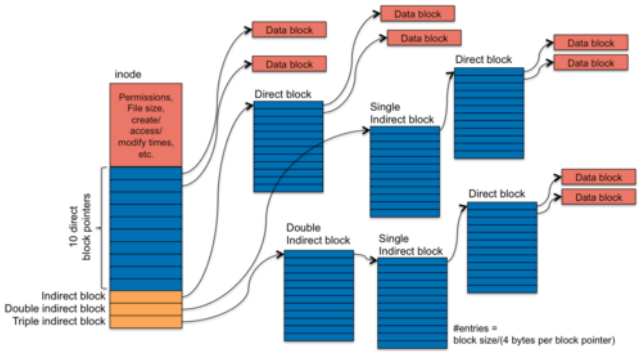
We defined scientific as repeatable. Any other expert will be able to perform the same experiment, on a clone of the image, obtaining the same results I obtained. The experiment is not just a black box tool with an input and an output, the expert must be able to perform the same analysis by hand (at least in theory). This means that analysis software needs to be open sourced, and possibly free. Proprietary or "law enforcement only" tools can be used but the analysis must be performed again with something repeatable. *(Judges, Lawyers, Marescialli, they can access the source for "law enforcement only" tools).*

7.0.2 What does analyisis mean?

During the analysis phase, we may need to apply many techniques from computer science, that's the reason because people goes to a forensic expert. Forensic experts are able to do analysis because they are competent in computer science in general. They can also be ethically prone to call other experts in fields in which they're not so expert.

7.1 Recovery of deleted data

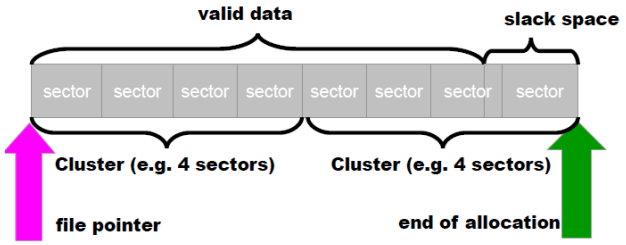
Data/evidence may have been voluntarily or unvoluntarily deleted because time passed, because of the OS, because the drive was formatted, because the drive was faulty. In many of these cases, we may be able to reconstruct all or part of the information, which is one of the most typical tasks computer forensics experts perform.
Let's look at the UNIX file system again for the millionth time to recall basic elements on data storage by OSs:
A file system is a way to organize the space on a drive to store information.



It's an archive-like thing. The basic idea is to have units of information *(files)*, to put on the storage drive, and then to have an index of what is stored and where. Following the index I can find the place where things are stored.
Whithout the index I may not reach them *(if they were deleted)* but they can

still be stored:
when we delete a file the operating system just marks the file as deleted and just doesn't shows it no more. The actual data will go away when overwritten by other files.
Indexes and data will go away randomly and independently from each other, statistically on a large hard drive there is a good possibility that metadata goes away before data blocks.

7.1.1 Disk Geometry



When reading and writing, the minimum portion of a track is called sector, each drive has a different sector size, and the OS can't manage it for each of the existent models.
OS's filesystems are made to work on clusters of sectors *(NTFS uses 4kb clusters)*, independently of the real dimension on the drive.
So, storing for example a file of size 5kb needs 1 cluster of 4kb + a fourth of a cluster, and leaves the remaining part as unused slack space.
In the slack space there can possibly be remainances of old files, obviously they can be small portions, so only small files can be retrieved *(e-mails, text files . . .)*, while other filetypes can't because they need to be full to be read *(images, encrypted files . . .)*
Sectors are one after the other, if we place a file on a drive there is a good possibility that all the sectors of it are one after the other, so by scanning *(cramming)* the drive can be possible to find them and restore them too.
The issues are with fragmented files, it was a big problem on small drives, now in most cases drives are so large to not fragment and most of OS's try to not break files for performance reasons. There still are techniques to find fragments of files and restore them.
Another issue is with encrypted, compressed files. In those cases we really need the file to be complete to be able to read them.
Headerless files because they were taken away header or footer cannot be interpreted. Headerless compressed/encrypted files are unpracticable.

7.1.2 Free software tools for data recovery

- TSK and Autopsy can perform data recovery under linux, they support NTFS, FAT, FFS, EXT2, EXT3..
- Foremost can perform file recovery trough carving
- gpart, testdisk can perform partition recovery
- photorec can seek specifically for images or videos

7.2 Antiforensic techniques

They are techniques designed to create confusion in the analyst and/or defeat tools and techniques used by analysts.

- Transient techniques: aim to confuse or mislead the analyst, can be defeated if detected
- Definitive techniques: make impossible to recover data

7.2.1 Critical failure points

Acquisition *(usage of tools for repeatable cloning and custody)* and identification *(usage of tools for analysis of file systems, data reconstruction and carving)* Interfering, it is possible to compromise the process. We talk about transient anti-forensics if we interfere with identification, definitive anti-forensics if we interfere with acquisition.

7.2.2 Timeline tampering (definitive)

Files' metadata store also information about modified, accessed, created dates *(MAC)* + Entry Changed (E) value on NTFS.
A typical way to use these data is to take all of the MAC entries of all files in the drive and put them in reverse order starting from the most recent, to display a so-called timeline.
As we go back, since every action overwrites the previous value we just will see the last one, as far we go, old data are shown because they maybe were stored and never touched.
MAC information can be modified to make them appear separated, close, randomized or moving them completely out of scope.
MAC though are not inforced by the OS, they're only declarations. There exist tools just to change these values, those values are not there for security reasons. In Unix you can use the touch command to change the thing for example, and there is no track of this, we can have a drive with manipulated access times.
So, as forensic experts we cannot guarantee that these values are correct, and if we use them to support or not support theories we must keep in mind.

7.2.3 Countering file recovery (definitive)

File recovery uses data remnants, but some tools can perform *secure deletion*:

- they can overwrite blocks, fill not allocated parts and/or slack spaces with zeros
- they can go on specific blocks and overwrite them

Encrypted drives can create problems with file recovery, also because they're managed in a way which is different from the usual filesystem's usage of the blocks.
Virtual machines, when dynamically allocating drives *(which are actually files)* have they to shrink if data are deleted and to enlarge when new data is written, often implicitly countering file recovery.

The talk

Gutmann in the 80's said that probably it was possible to understand if a zero was written in a sector or a one was, because of the magnetism. In reality it was difficult to do that (impossible) even in the 80s, with nowadays dense disks it is certainly not possible.
But we still have the Gutmann patterns to *"erase"* disk contents.

7.2.4 Fileless attacks (definitive)

Many modern attacks tend to be fileless, with no traces on the disk at all.
For instance, Metasploit's meterpreter is injected in a process' memory space, and gives attacker control without writing anything to disk. So, if the machine is off, the evidence is lost, the only thing that can save us is to dump the memory of the machine, if possible, in our acquisition phase.

7.2.5 Filesystem insertion (transient)

Data placed where there's no reason to look for them, in particular inside filesystem metadata. Inside a partition table there is space for $\sim$ 32KB of data.
In ext2/3:

- RuneFS can write in bad block inodes
- WaffenFS adds a fake ext3 journal in an ext2 partition
- KY FS uses directory inodes
- Data Mule FS puts data in padding and metadata structures of filesystem ignored by forensic tools

7.2.6 Log analysis (∼ transient)

On most machines one of the big sources of info for forensic analysis are logs.
Logs tend to be long text files with structured text inside, so they can be manually scrolled or checked by use of regular expressions or scripts.
If attackers can inject stuff in the logs (very likely), they can try to make your scripts fail, or even to exploit them.
Maybe they use an username that contains a linebreak character, the line would be skipped (theoretically example).
Techniques which are transient in the sense that they're trying to confuse the analysis tools. In reality however some of them could be modified in the way in which the log becomes definitive.

7.2.7 Partition table tricsk (transient)

"sci-fi"
Partitions can be uncorrectly aligned, this may be enough to make a forensic tool to fail.
Adding a number of extended partitions which can be managed by the Operating Systems but not by forensic tools, or a sufficiently large number of logical partitions in an extended one to make the tools fail.
Transient techniques like this work well when for example the drives are a lot and the analyst has no clue of where evidence can be. But, with a small number of drives, if an expert knows what to search for and where to, most of the times he can find the evidence.

Chapter 8

SSD Forensics

8.1 SSD technology

Today, most of the drives are SSDs. SSDs are storage drives made of NAND flash memory chips, which are faster, and were cheaper than an Hard Drive.
By the point of view of the Operating System, there is no difference between *"talking with"* an SSD or an HDD, but the two technologies are slightly different.
NAND flash memory though, has a limited lifetime, and there is the need to manage how to make them to last longer, another difference is that those kind of memory's blocks are only **fully writable/erasable**, so we don't see slack space in this kind of drives.
When a block is re-written, it has to be *blanked* before, this is a big disadvantage for forensic experts.
What manages SSD behaviour is the FTL controller, which fakes for example to the Operating System the SSD as working as an HDD. It manages a lot of things:

- **Write caching**: it keeps data in a cache and writes them on the SSD only when needed, to preserve blocks life.
- **Trimming**: it blanks no more useful blocks any time it is idle. (even with write locker connected)
- **Garbage Collection**: they were able to figure out what to delete when Trimming wasn't supported by Operating Systems.
- **Data Compression**: they can store more data in less space, to preserve blocks life.
- **Data Encryption/Obfuscation**
- **Bad Block Handling**
- **Wear Leveling**: they manage blocks in a way to make them be in the same deterioration state.

In HDDs when the OS addresses a sector, it gets the exact same sector from the drive, while in SSDs, the FTL manages the way in which a logical address requested by the OS is mapped into the respective physical block address on memory chips.
The underlying mapping is completely transparent and can be modified by the FTL at any time for any reason. The FTL may move data around or blank data even if the OS is not running.

8.2 Can we bypass the FTL?

We don't know hot the FTL works, it is proprietary information of the company which produces it. We need to bypass it.
It is not possible by software, it is in theory possible by directly reading the memory chips, using a custom setup built to interact with flash memory chips using an FPGA. This is in any case:

- Extremely time and money consuming: there is the need to buy and manage custom hardware, reverse-engineering of the FTL data management.
- Non repeatable: the operation of physically extracting the memory chips may be destructive or alterating.

Mobile devices work in a similar way, so they have the same problems.

8.3 Challenges in black box analysis and goals

8.3.1 An unclear picture

Most of our recovery of deleted files works because of leftovers on the drive. If the chip trims the drive, the leftovers are no more there: we need to know how fast trimming happens. Since this is performed by the FTL chip, it is performed when the drive is turned on, even if the drive is not connected to any computer. Write blockers don't work too.
USB sticks has the same fastidious characteristic. Notice that this causes the hashes to be different everytime.
Carving don't woek at all on SSDs.
What the professors did was to test different SSDs to determine what was possible to do, and which was the impact of FTL on the use of black-box tools.

8.3.2 Testing methodology

They tested for:

- Trimming: negative impact on forensics as data persistence is reduced, it happens even with a write blocker. They wanted to determine the percentage of erased blocks and how fast.

- Garbage collection: they wanted to determine whether it is employed by the SSD under examination.
- Erasing Patterns: they found out that particular behaviors were shown by some SSDs when using trimming.
- Compression: to verify if it was active
- Wear leveling
- Files recoverability

8.3.3 Test drives

They used three different test drives, equipped with a small amount of DRAM-based cache memory to reduce physical writes, which was disabled.

8.3.4 Trimming

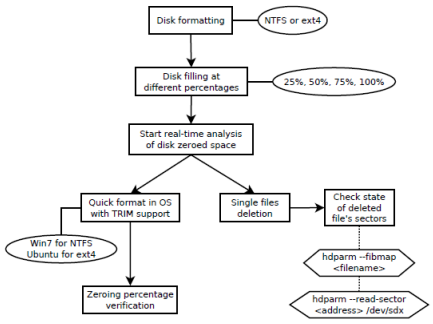


Figure 8.1: trimming test flow

Their results:

- If trim is present and active, it activates in 1-10 seconds
- NTFS wiped the disk in less than 10 seconds
- ext4 erased the disk in about 15 seconds when formatting, some drives did not erase on delete, some only when unmounting.

Consider that nowadays most of the filesystems support SSDs, so they will work correctly. But it may be useful to check for any remainance.

8.3.5 Garbage Collection

In their tests, none of the SSDs performed garbage collection, even when trying to repeat a documented experiment with the help of the author of that experiment.

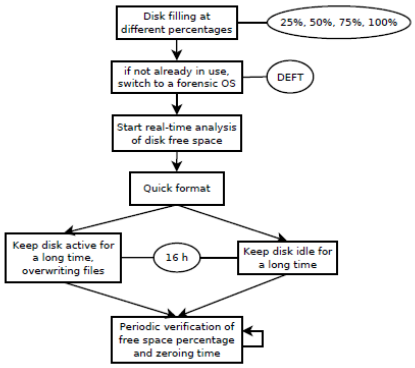


Figure 8.2: garbage collection test flow

8.3.6 Erasing Patterns

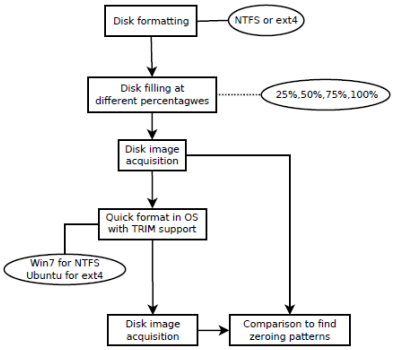


Figure 8.3: erasing patterns test flow

They found out that certain SSD controllers may exhibit unexpected trimming patterns, leaving some parts *more recoverable* than others. What the professor thinks is that the drive could be faulty, but this is important to check then, faulty drives may still keep data.

8.3.7 Compression

They found out that some drives had less data transferred when transfering the same file, thus compression.

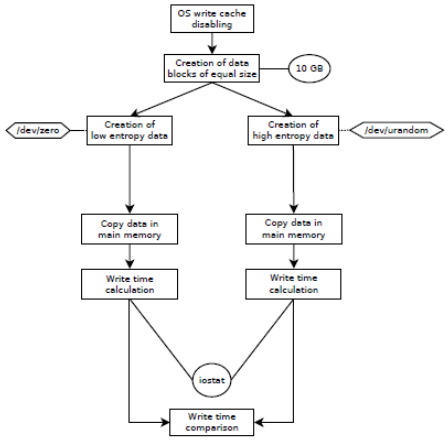


Figure 8.4: compression test flow

8.3.8 Wear Leveling

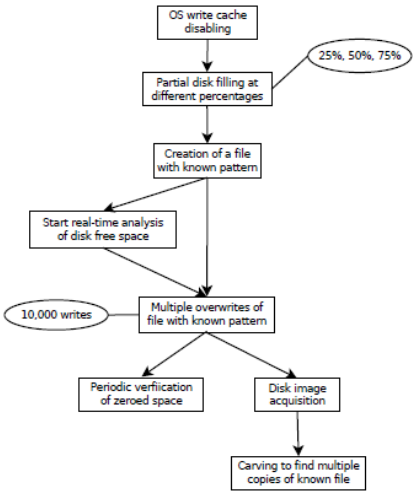


Figure 8.5: wear leveling test flow

They wanted to find if more copies of the same data were created while doing wear leveling. They weren't.

8.3.9 Files Recoverability

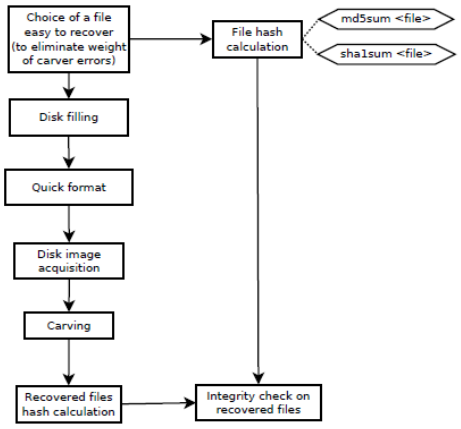


Figure 8.6: files recoverability test flow

Only the drive which presented trimming had recoverable files.

Chapter 9

Cloud Forensics

Cloud computing is a computing-as-a-service paradigm, which has different declinations: IaaS *(Infrastructure as a service)*, PaaS *(Platform as a service)*, SaaS *(Software as a service)*.
We also distinguish between public cloud and private cloud *(company owned)*, but we will consider mostly public ones.
Different issues are present:

- Issues with acquisition and access to evidence
- Analysis issues
- Issues with attribution
- Issues of legal status

9.1 Acquisition issues

In general, no control is given to the user on hardware and storage space, so investigators cannot really access the metal. This makes traditional acquisition procedures unfeasible for the host, but they are feasible for guests.
It is possible to acquire virtual machines for example. Levels of access vary:

- SaaS: the service provider is the only one to have logs/data
- PaaS: the customer may have application log, network log, database log or operating system depending on the content security policy
- IaaS: logs until OS level are accessible to customers

The real problem is that data can not exist sometimes. Instagram pages are not stored, but composed at the moment they're seen. This is hard to understand for court people, and also brings up situations like editorial responsibility.

9.1.1 Simple case: web pages

What can be possibly go wrong?

- There's dynamic content on the page: how do we capture it, how do we reproduce it in courts? If it's coming from external sites, what is its legal status? If it's an important part, we may be in need to acquire multiple copies to show how they change.
- Visualization is different from data
- Attribution: where is physically located the server that deploys the web page? It's not easy to determine where the datacenter is physically located.

How do we find the exact location of a web infrastructure with a certain IP address? How do we prove where it is?
It's not an easy question. Expecially if we want to prove where it was in the past. We may be able to prove what the IP address was, but it's very hard to prove where physically the hardware was located. The infrastructures are internal, even for who is running the location. Tracking down the specific owner of a machine requires the collaboration of different entities.
Whois allows us to track who owns some parts of the ip address space, who owns a certain domain address. In some cases we need to demonstrate that a certain domain resolves to a certain ip address, this must be proven from multiple points because the DNS resolver may be lie.
Geolocation of hoster: there is no real geolocation of IP addresses. There are good guesses.

9.2 Analysis issues

Virtual machines *(which are the lowest level acquirable things)*, when acquired, have dynamically allocated storage, so their storage will not have remnants of data *(with few exceptions)*. Metadata of VM disappear easily, for instance everytime a snapshot and a restore is done.
Are we going to try? Yes. Is it surely working? No.
If the case is one of hypervisor-level compromises which allowed an attacker to bypass the control of the service provider and jump from one tenant to another, there is no way to investigate this. Usually because they are custom hypervisors with no tools and research. Fortunately it is a very corner case.

9.3 Attribution issues

Cloud infrastructures create an additional layer of uncertainty. An IP address with a timestamp will give us a machine, we need to get from there to the specific person, sometimes it is easy *(someone who lives alone)*, sometimes it is hard *(example: families)*.

If you reach the router of a family, and no family member confesses the crime, how do you find the one who committed the crime? It is hard to find specific attribution.
With cloud providers, we have another difficulty: their cooperation. It is not so easy to get that, because most of the times they are not in the same jurisdiction as we are, and we need a *"rogatoria"* to access data from another jurisdiction, it takes time and money.
Since in cloud service providers, there is the need of them to cooperate, this makes investigation of transnational crimes with them involved even harder than normal.

9.4 Legal issues

- Many types of legal procedures depend on physical geographic location
- Electronic data may span across different locations, or created during a transaction and not existing in any location. We may need to explain in detail where and how data is stored, generated.
- Under Budapest convention there is this: ordinarily search and seizure happens physically. A prosecutor issues a warrant for a search *(Mandato di persecuzione)*, and this is done by the Carambineris physically. They introduced the concept of electronic search and seizure: if the thing i need to search is on dropbox, email, .. I am searching the digital possessions, while on a surface level this sound as reasonable, it brings a lot of considerations: in some cases the search and seizure is not really a search, it involves something generated by a transaction, this is documenting. There are specific laws for different procedures, we need to be able to describe the differences in a way which is understandable by lawyers and judges.
- Removal of obstacles: when a search warrant is issued in Italy, it usually authorize the police to remove obstacles *(sfondare la porta)*, and this has the effect to make it not a crime because it has been authorized. Giving execution to an order is not a crime in Italy. If the main server to which the access control system needs to be removed is in Norway, it is highly debatable to determine if it is a crime or not. Technical experts must be aware of it. You need to request to Norway's authorities. In digital world it may be not realizable, *(Realizzabile nel senso realizzare mentalmente)*, so you need to be aware.
- In our agreements with CSP we need to include the forensic assistance that may be used *(The user ask forensic data to be given to the court)*, because otherwise Microsoft can say "No dati non cielo".
- More complex because CSP use each other: at the beginning Dropbox used AWS as a storage provider. Interesting issues.

9.5 Forensically-enabled clouds

When we create a contract with a CSP, the fact that it is able to empass with forensic information, may be good for our contract because law may require that my business needs to have a certain kind of accord.
There are also requirements to keep data in a certain location. This may impact optimization. But to be able to offer the service in that location, they do it.
In order to be auditable we may ask to provide snapshots, proofs of past data possessions.
Encryption and key management is important, if certain kinds of keys must be kept by certain organizations, and they store them (?)

9.6 Dual considerations: cloud-enabled forensics

A typical laptop today has a 1 or 2TB hard drives. With 100s laptops seized, you need to search in 100sTB of data. Doing this through cloud enabled services it may be useful, but in some cases we're legally obliged to mantain control of the data.
Transfering data on a cloud provider can trigger challenges related to personal data, not all seized people are indagati, we have their data and a duty of care wrt them *(and convicted people)*.
If those data get stolen, indexed because of our error, it may be lead to problems.
As forensic experts, this is a significant challenge.
Also transnational issues because we're performing forensic things in another state.

Chapter 10

Evaluation and Presentation

10.1 Evaluation Phase

Evaluation is the phase in which we're trying to match evidence elements with the required elements that are part of a legal theory.
What are the elements of the legal theory for kid's bad things are:

- User has pictures, found trough carving or search. We need to demonstrate that they are what we are saying them are.
- Willingly detaining? We can propose facts to support or not this legal theory: were the images in a specific folder, sorted, access dates, in the cache of the browser?
- If the images were deleted, we have not the metadata to understand if they were willingly detained.

10.2 Items to evaluate

The elements to support the indictment. Alternative explanations to the elements: does the elements univocally support the theory or there is another theory? Defense experts job. Prosecutor expert witness, may or may not be obliged to search for them. We can also analyze something which comes from another person's investigation, so when analizing evidence or sources of evidence, we know that (?) Analyzing what can be said, what cannot, and what further experiments would be needed to say more: sometimes defense can under certain circumstances ask the prosecutor to do something, the data is not there at the moment, how likely is this data to go in my direction or in the other?
If you're working as Prosecutor's expert witness, the answer is to perform the

experiment to retrieve what's missing. For the expert witness of the defense, this is an important consideration.

10.3 Relationship with lawyers

Lawyers own the choice of defense strategy, they decide what is the legal theory to present to the court. We can give elements but the choice of the theory is left to the lawyer because it has implications that we as engineers are not trained to handle. Also, legal professions have a specific trust relation with the client, they own the relationship with the client, experts should never overshadow this relationship. If you think that the lawyer is hurting the client, you at most can resign.
On the other hand, we don't know what the lawyer knows and vice-versa, we're collaborating. Our goal is not to damage our trust relationship.
The lawyer *(the customer)* is paying us, but, they don't dictate you what to write or to say. The most they can do is to omit something you found if it's not good for the defence strategy. Lawyers may hide things or lie to the court. Witnesses may end up in jail if they lie. Also, in the Italian jurisdiction, some professionals have a right to professional secret *(doctors, lawyers, priests)*, but we as engineers cannot lie if we know something, we end up in jail if asked and we lie or say to don't know.
There may be things that the customer knows that is better that the expert witness don't know, because if asked he might answer.

10.4 Relationship with the customer

Any expert witness assesses one of the parts of the judgment, which in civil law are equals, but in criminal law they are the supposed criminal and the law, in Italy they're not equals. By working with the defense you will be considered less credible than the prosecutor witness, especially if a carabiniere.
The mandate of an expert is to find what helps the customer, which is not the same as *helping someone escape law*: by helping any of the parts in a criminal procedure to put forward the best possible evidence in their favor, we're helping the process of determine whether or not the punishment would be applied.
"Process Truth", is different from historical truth. If something is not in process truth, against our customer, we must do anything in our power to not let it end up in process truth.

10.5 Relationship with prosecutors/police

Working with the prosecutor doesn't necessarily entail moral superiority. It's still very important to stick to science and facts. It's important to not get your words or thoughts shaped by *justice*.
It's easy to fall into this trap. It's hard to stick to the facts. We have ethics, we

want the right thing to happen, but criminal justice is not founded on ethics. So, we cannot act based on our sense of ethics or justice.

10.6 Evaluation: analyzing the documents

How does someone read a written report of someone else? In particular, written reports of other expert witnesses and investigators.
Judges read written things, which is the real way in which they work, they'll read our reports, the documents, everything that is written. So what do we look for in others' reports:

- Technical or factual errors (or omissions).
- Unclear reasoning, methodologies or descriptions: because if it is unclear enough, it may give us the possibility to say that certain analysis is not credible, and because in court whoever is clearer is more convincing, so if their writing is unclear it is a great opportunity to be clear and explain things in our way.
- Suggestive writing: writing which insinuate or suggest something without proof.
- Opinions and hypotheses not clearly distinct from facts and not substantiated: particularly true in criminal law settings.

10.7 Typical technical and factual errors

- Acquisition:
  - Search and seizure: process, chain of custody, seals ...
  - Description of seized/analyzed materials: serial numbers ...
  - Hashing/Cloning procedure *(e.g. write blockers)*.
  - **Unless this is really important, judge will ignore it**
- Analysis
  - Steps where hashing wasn't verified
  - Use of proprietary programs or with known bugs and/or vulnerabilities
  - Description of the process
  - Technical mistakes

10.8 Typical presentation errors

- The first and most typical that we'll use agains others is the lack of exploration of alternative hypotheses. *(this element could have happened because of this but this do not explain this other thing ...)*. If there is an alternate hypothesis that explains piecese of evidence taken alone but they do not work in the context usually help the thesis because you already explained and discarded those hypotheses. However, if the other explanation is not credible, I'm not going to help the customer well.
  "Trojan Defense clichè:" there are trojans and botnets that maybe put images in a computer to make it look like it's owner committed a crime. But things like *"a trojan download images and then perfectly deleted itself"* is not falsifiable. It's a very typical potential defense, so typical that becomes a clichè if not coupled with scientific proof.

- Is the presentation neutral or biased?: each expert witness is working for a part, but if while writing a report we put a lot of bias in the report, it end up being less credible to the judge. The less bias we put, the better the job will be. Counter-examples are a powerful tool to deny the credibility of a report. Since the process is a debate, rethoric and counter examples can be a way to diminish the other's arguments. If you have 5 arguments and 3 of them are very strong, bring 3.

- Can we find counter-examples for some of the assumptions? Counter-examples are a powerful tool to deny the credibility of a report. Since the process is a debate, rethoric and counter examples can be a way to diminish the other's arguments. If you have 5 arguments and 3 of them are very strong, bring 3.

- Are there missing explanations that we can provide in order to shift the understanding of the judge? Whoever is clearer is more credible. Did the counterpart not explained something to the judge? Give them an explanation, establish credibility. The trier of facts do not understand the subject, when you're writing your text think about it.

10.9 Presentation: writing your report

Extensive: they need to cover everything. Sometimes your report is the only report. If there is a collective of experts, in those cases you'll have more opportunities to comment on things. When you are one, you need to write everything and also to be concise: the ones reading have no time and patience to read something very long. In fact, in most cases, if you write a technical report for a judge, he'll look only at the conclusions. So, you're writing it trying to be concise and do cover everything, and to also be clear to the judge and the lawyer who don't know about technology, or they would have not called. Probably technology is not even in their interests. In Italy you probably want to avoid

english terms, and if they're not avoidable you need to provide an explanation for them. Writing things oversimplyfying them, can make the poor judge to feel treated as a stupid. And since we are technologists this is the part of our profession we don't care about. But in a tribunal the only thing that matters is what you write and how you speak about it. Notice that the report is also going to be read by the other experts, which will apply the same checklist above to your report. We need to make what we write to matter, we need to map the technical things to the legal matter. If something is not relevant, it doesn't map directly to something, it doesn't matter.

10.10 How not to write a report

Don't write things like *"it doesn't work"*. Why? Clear the meaning of what you're saying. It may be obvious to you but not to the reader. Don't suggest, don't use innuendo. Explicitly say *"this is important because it may lead to that", "it is compatible/incompatible with", "it does/doesn't support it"*. They're all very clear in how much this proves things. We need to prove with facts what we think.
Since we know that most technical things require very in-depth knowledge, we tend to obscure it as demonstration of expertise, and it works with other experts. But if you are too technical in your report, it just will be ignored. The clearest explanation wins when talking with a non-expert. When talking to an expert, the detalied explanation wins. You are never going to convince the other expert because he's payed to not to be. If you're in the setting of evaluating something with another expert, this is different. But, if the report is for the judge, you need to be clear.
Another typical mistake is to show biases, you do not need to show solidarity with your client, this is expected. It's not like working for the accused person because you don't care about the victim of the crime. You don't need to show in reports the solidarity with the victim. Everyone in the court is equally solidal with the victim. No solidarity with the client, he is paying you. You need to analyse everyting in a very cold and factual way.
Don't show excessive deference to the judge.
Don't use sarcasm, however iff you're very sure of yourself, and the other part made an egregious mistake, then a *little* irony. Don't use a weak argument if you have a strong one.

10.11 Structure of a report

Model it on a scientific paper or report:

- Introduction: what I'm going to say and why it is relevant
- Facts

- Discussion and analysis: each block with a small introduction saying what you want to explain and a conclusion summarizing what you just describe
- Final conclusion: important things you want to convince the judge of. Leave out any doubtful statement.

Structure it like an obstacle course, each obstacle a little taller than the previous one: make the judge sweat to ignore all of your obstacles.

10.11.1 Example

- Small introduction of yourself as expert witness
- Foreword: I have examined these documents and these evidence sources, I was asked to report on X
- Introduction: Say what we're going to show and how are we going to show that: *"I will show A,B,C,D"*. Since we do present specific parts of evidence, we need to explain to the judge why we're writing about this.
- Acquisition issues: if we're working for the defence, we show which are the issues in the acquisition part (hash not computed, ...). *"this is the lowest obstacle in the obstacle course"*.
- On the technical analysis: start with factual errors.
  - Even ignoring all of our fundamental issues with the evidence integrity...
  - The connection wasn't from A to B, but rather to C. This changes the reconstruction like this, and also makes these paragraphs of the adversarial report wrong
  - Even ignoring these factual errors, the evidence is not best explained as the adversary did, but rather is much more compatible with these other explanations
  - The following experiment has not been performed, or the following evidence is missing, and it could...
- Conclusions: *the only thing most of the judges will consider.*
  - In this report we have shown:
    * Evidence was not properly acquired and thus these manipulations may have contamined it
    * Item I did not happen as described, making conclusion C completely wrong (don't fix it for them: if you find a mistake, if it is relevant (changes something), we know that this is a weak point, correct it in our minds but then the fix is up to the other side, if they have time, knowledge, understand the problem)

    - Theory T fits the facts better
    - Nobody checked if F was true or false, which could have led to G
  - We must therefore conclude that **(one strong, inevitable phrase)**

10.12 Testimony as a witness

In most jurisdictions, the expert witness work do not end with the report. In some of them you may be called as a witness in a tribunal. In Italy, in criminal courts, you'll be called, your report is included in the trial only if you testify. In civil courts there is no need to testimony.
You'll be called to do a sworn testimony. Where you swear that you tell the truth. In Italy no swear. If in specific circumstances you say something that is false, you commit the crime of perjury.
This means that expert witness cannot lie or claim confidentiality or professional secrecy.

10.12.1 Direct examination

What is going to happen is that you're being called as a witness by your side. You start with the "friendly" part of examination, which you may have prepared with the lawyer *(list of questions that they're going to ask)*. In direct examination you're going to be the most clear as possible. Make sure you explain everything to the judge, take your time (not too much). However in Italy, the judge can ask questions at any point in time, so prepare, check previous records of the judge, obviously you can't prepare as for the direct examination.
After direct examination, you stand cross-examination.

10.12.2 Cross examination

Examination by the counterpart is less friendly, sometimes downright hostile.

- Prepare: check previous records of the lawyer, prosecutor or judge
- Use your report as a shield and as a way to get time and think about the answer
- Be curt if you can "yes" or "no", if you cannot, be very complex and difficult to understand
- If unexpectedly a question is positive, immediately go back to being extremely clear and helpful
- Don't get angry *("Quindi per non essere io quello che finisce in carcere, mi conviene escludere la possibilità di fare questo lavoro")*
- Don't be surprised if your competency is called into question