

Computer Forensics and Cyber Crime Analysis

Alessandro Milani

October 16, 2024

Contents

I	Legal	3
1	Legal Introduction	4
1.1	GDPR and algorithm bias	4
1.2	Example - Lex Machina	4
1.3	Example - Compas	4
2	Foundations of Digital Forensics	5
2.1	Intro and definitions	5
2.1.1	Digital and Electronic Evidence	5
2.1.2	What is digital Forensics	6
2.1.3	The “Big Five” of Digital Forensics (Council of Europe)	7
2.2	Digital Forensics Procedure	7
2.2.1	Identify the Suspect	7
2.2.2	Detecting and Seizing Digital Evidence	8
2.2.3	Validating Digital Evidence	9
2.2.4	Chain of Custody	9
2.2.5	Analysis of Digital Evidence	9
2.2.6	Presentation in Court	10
2.3	Privacy and Due Process Rights	10
2.3.1	Encryption	10
2.3.2	Case Law on Encryption	11
2.3.3	Mandatory Key Disclosure Laws	11
2.3.4	Remote Forensics	11
2.3.5	Cloud Computing	11
3	Convention on Cybercrime	13
3.1	E-commerce on Dark Web	13
3.1.1	Silk Road	13
3.2	History and objectives of the Convention on Cybercrime (Budapest Convention)	14
3.2.1	Overview of the Convention	14
3.2.2	Aim of the Convention	15
3.2.3	Key points	15
3.3	Harmonization of national laws and international cooperation	16
3.3.1	International Cooperation Provisions	16
3.3.2	Mutual Assistance Provisions	16
3.3.3	24/7 Network for Immediate Assistance	17
3.4	Legal measures against computer-related fraud and forgery	17

II	Tech	18
4	Introduction	19
4.1	Topics	19
4.2	Forensics History	19
4.2.1	Ancient Times	19
4.2.2	Modern Times	19
4.2.3	Digital Field	19
4.3	Computer Forensics Definitions	20
4.4	CF purpose(s)	20
4.4.1	CF Q&A	20
4.4.2	CF Goals	20
5	CF Terminology & revelant concepts	21
5.1	Terms	21
5.1.1	Digital evidence	21
5.1.2	Chain of custody	21
5.1.3	Data acuisition	21
5.1.4	Hashing	22
5.1.5	Write Blocker	22
5.1.6	Forensic image	22
5.2	Scenarios	22
5.3	investigation phases	23
5.3.1	Phases	23
5.3.2	Identification	23
5.3.3	Collection	24
5.3.4	Acquisition	24
5.3.5	Evaluation	25
5.3.6	Presentation	26
6	Non-trusted environment issues	27
6.1	Compromise causes	27
6.1.1	Node infection	27
6.1.2	network injection	27
6.1.3	supply chain attacks	27
6.1.4	Men at work	28
6.2	Advanced persistent threats (APT)	29
6.2.1	APT Attack Process	29
6.2.2	manupulation from the system owner	29
6.2.3	APTxx	30
6.3	Trusted Environment	31
6.3.1	(example of) System Call Interception	31
6.3.2	Examples of Linux system modification	31

Part I

Legal

Chapter 1

Legal Introduction

Before the technology was between us (classic telephone call), but with the advancement of the information technology, the technology is now about us (facial recognition, social media, etc).

Example: The IA act says that it is not possible to utilize IA for real time facial recognition without an "important" reason.

This generation uses technology also for being profiled by an algorithm for various scope and not only for communication.

The "technology was between us" was simpler and the only problem was to check if there is a conversation and intercept it with a good quality. Now, we have the problem of quantity. If we need to analyze data from millions of people we can end up violating fundamental rights and make mistakes (even with OSINT).

Surovsky theory: collective intelligence, the point is that if you have 10k persons say that the restaurant is good and only 10 that say it's a fraud, the collective intelligence says that the majority have right. In Forensics this can not be applied because you need to be 100% secure of what you have. (if in a trial you have only the 1% that a person can't be guilty you need to be in favor of him) [find better term]

In the technology in US, and the advancement of IA is important that the law split what is human and what is not (like being transparent when a content is AI generated and when not)

"Tesla case" when there is an incident it needs to be understood in the percent of error that is from Tesla and the percent from the partners.

1.1 GDPR and algorithm bias

Art.22 of GDPR, says that you always need to have human in the decision process

1.2 Example - Lex Machina

Tool that analyzes all the legal cases from a jurisdiction (like France) and classifies all the cases in different categories. So if you have a case X in Paris with judge Y, you have 60% possibility to win. If the Attorney is Z, the probability is 80%.

1.3 Example - Compas

Algorithm that helps the judge decide if the person can commit other crime or not and so decide if they need to stay in jail or get a reduction in the sentence (sconto della pena)

- A False positive in digital forensics can change people's lives.

Chapter 2

Foundations of Digital Forensics

2.1 Intro and definitions

2.1.1 Digital and Electronic Evidence

By the Scientific Working Group on Digital Evidence (**SWGDE**) a definition of, **digital evidence** is "any information of evidential value whether memorized or sent in a digital format". It's used by the **Council of Europe**

Another definition come from the **Eoghan Casey - 2004** that define a **digital evidence or electronic evidence** as "any probative information stored or transmitted in digital form that a party to a court case may use at trial". It's more related to the juridical part.

A last definition of **Electronic evidence** is information generated, stored or transmitted using electronic devices that may be relied upon in court, defined by the **Council of Europe - 2013**.

For the exam, the first definition is the more important

So, in general, way we can say that a digital/electronic evidence need to be:

- **invisible** to the untrained eye
- Need to be **interpreted** by a specialist
- It may be **altered** or **destroyed** through normal use
- It can be **copied** without limits

Legal Requirements

The main characteristics that a Digital/Electronic Evidence need to have to be accepted in a trial are:

- **Admissability:** it need to be compliant with law and best practices.
What can be seen is not what can be admitted in court (ex. if the italian police enter in a laptop, can only "wiretapping", by enabling mic and camera, and can't use other information like email or files in court)
- **Authenticity:** avoid any digital evidence tampering
- **Reliability and believability:** readily understandable for a judge.
If a judge not understand the evidence can ignore it
- **Proportionality:** respect fundamental rights of parties affected by the measure

Find a digital evidence

A digital evidence can be hidden in different place and a criminal usually use some classical ways (not the cloud because the access is easy from a police force) like hidden folder, usb, external memory etc.. can be hidden everywhere

Categories

Three main types of digital evidence

- **Created by human:** digital data result of an action taken by an human
 - *Human to Human:* Like an email
 - *Human to PC:* like a word document
- **Created independently by the computer:** data that are result of the processing of data by an algorithm and without human intervention
- **Created by both human and the computer:** somethings like a spreadsheet where the data are entered by the human, and the result is worked out by the computer

Julie Amero Case

This case is not important for the exam

Julie Amero is a supply teacher at Kelly School in Norwich, Connecticut who was found guilty of showing pornography to children under the age of 16 for some popup that appear during a lesson

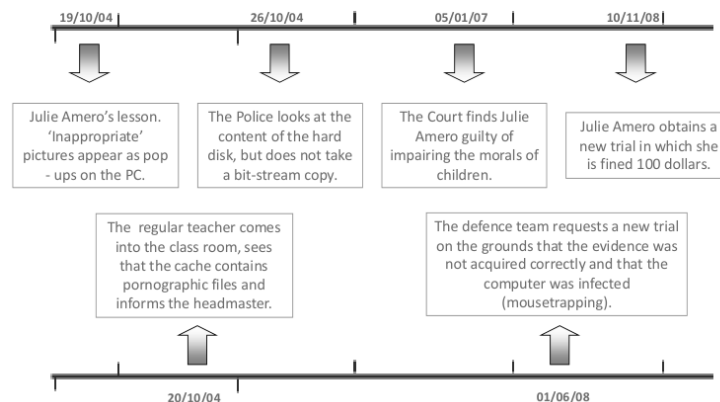


Figure 2.1: Timeline of the case

2.1.2 What is digital Forensics

Digital Forensics is get hold of evidence without modifying the IT system in which that evidence is found, ensure that the evidence acquired in another medium is identical to the original and analyse data without modifying it.

2.1.3 The “Big Five” of Digital Forensics (Council of Europe)

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

2.2 Digital Forensics Procedure

Six phase of digital forensics procedure:

2.2.1 Identify the Suspect

There are 3 main phase for identify the suspect:

- **Osint and Socmint:** Very usefull for collect information regarding criminal (even mafia ones), from social media, and other public sources.
- **Data Retension Directive in EU:** The investigator uses the Court System to compel the ISP to reveal a physical location that corresponds likely to the source of Network (IP Address)
- **Multiple User ID or multiple Ips over time, open Wi-Fi, Proxy, Botnet:** Under a warrant (depending from the Jurisdiction) the location is searched and any computer or other device is seized

Data Retension

With the Directive 2006/24/EC, the EU member states are required to store data for a period of **6 to 24 months** (but can change from state to state). The data stored are gerally call detail records (CDR) of telephony and internet traffic and location data (IPDRs).

So evert single country and ISP have different data retention policy, and this can be a problem for the investigator, but from a privacy point of view a short time or null data retention is better.

Transparency report: Every year the ISP need to publish a transparency report where they show the number of request of data retention and the number of request that they have accepted.

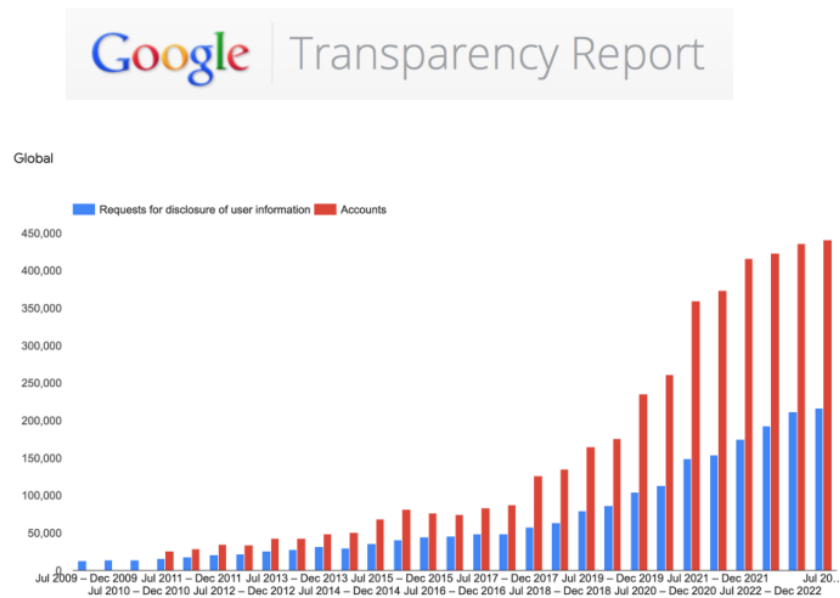


Figure 2.2: Timeline of Google transparency report

Identify the suspect

it's right use face recognition to identify a suspect?

For the moment if the situation is critical is possible utilize a live facial recognition.

AI act also regulate the use of facial recognition.

2.2.2 Detecting and Seizing Digital Evidence

The seize of digital evidences has to respect two fundamental rules: **bit-Stream Copy** and **Hash Function**. (definitions are already known)

Where and how is the digital/electronic evidence hosted?

The digital evidence can be in the suspected PC or in a third party server.

In the first case, we need to manage the encryption of the data, and it's possible get a Key Mandatory Law.

In the case of evidence in a server, is needed a collaboration from the ISP/Telco/Bank, and so need there is Jurisdiction problem.

The role of third parties during digital investigation

A third party can give a lot of useful information.

For example and **ISP** could reveal from which place the email was sent, the **Mail Account Provider** could reveal from which places the email account was accessed and a **Credit Card Company** could reveal where the goods bought with a cloned credit card were delivered

2.2.3 Validating Digital Evidence

There is some tool that help to validate online digital evidence.

These kind of tool are usefull during OSINT because they permit to collect data (like a story, a reel), that are not sure remain online, in a proper way for a court.

2.2.4 Chain of Custody

Digital storage media last less than analogue media and devices to read such media last even less. For example a LaserDisc last only 15 years, where there are books form thr 1086 (domesday book). So there at the moment, for trial, there are a lot of hard drive kepted in proprer way to avoid the data loss. It's a real mess.

2.2.5 Analysis of Digital Evidence

Start after the sieze of suspect's device, and need to be performed besides a precise chain of custody.

To perform the analysis are usually used some automatic tools, but in the recent time are used also some AI tools but only for post analysis and not for prediction policies (limitation imposed by the AI act) for example AI can not be used for kidnapping cases because the crime is in progress and not "finished".

- **Text searches:** aimed at scanning files, directories and even entire file systems for specific text terms (generative AI can be used for summarizing and analyzing documents, but it's not very precise, plus alucination)
- **Image searches:** aimed at identifying image files in various formats, and at generating still frames of digitally stored video footage. Mainly analysis of child pornography that can be lead also to false positive (like a video of a mum and child in a bath)
- **Data recovery:** aimed at recovering all files stored on mass memory units, including deleted or damaged data. Destroy data can also be a crime (even if there are some backups), based on the intention of the crime (like delete file to hide evidence)
- **Data discovery:** targeted at accessing hidden, encrypted or otherwise protected data
- **Data carving:** focused on reconstructing damaged files by retrieving portions of their content
- **Metadata recovery and identification:** this digital forensic tool is particularly useful for retracing the timeline of web accesses and file changes

Some other problem withthe use of **AI for prediction** of crime are: the bias of the AI and possible consequences for privacy and **social control** by not very democratic government.

Two Italian issues

Repeatable or Unrepeatable forensics analysis: **Repetable** when you can do a bit-stream copy of the data and also give one copy to the defence to do the same analysis, and more in general i can repeat analysis again and again (when i want).

In a **non-repetable** analysis, we need do live forensics activity and ofter occure with mobile phone, where there isn't the possibility to do a bit-stream copy of the data. In the live forensics i also need the presence of the attorney or the defender when i do the analysis to make it admissable in court.

Open Source or Closed Source: Open source can be more transparent

2.2.6 Presentation in Court

The presentation of digital evidence findings is a **crucial stage** for prosecutors, judges, and lawyers (the evidence need to be presented in a way that the judge can understand it otherwise he can ignore it). The outcome of the trial relies not only on the results of the investigation but also on the **clarity and comprehensibility** of the report provided.

Operational Recommendations:

- **Presence of an index:** The report should include a clear index for easy navigation through the document.
- **Glossary and Reference Notes:** If technical terms are used, a glossary and reference notes should be provided to ensure that all parties understand the terminology (judge and lawyers are not IT experts).
- **Timeline Table and Flow Charts:** A timeline table or flow charts should be included to visually represent the sequence of events and digital activities.
- **Presentation Slides with Photos:** Visual aids, such as presentation slides with photos, help in simplifying complex technical details.
- **Video Recording:** Where applicable, video recordings of the operations carried out during the investigation can provide further clarity.

Presentation in Court of the Digital Evidence Findings: Murtha Case

2.3 Privacy and Due Process Rights

2.3.1 Encryption

Encryption can be used to hide the fact that encrypted messages are exchanged and used by criminals can lead to difficulties collecting the necessary evidence

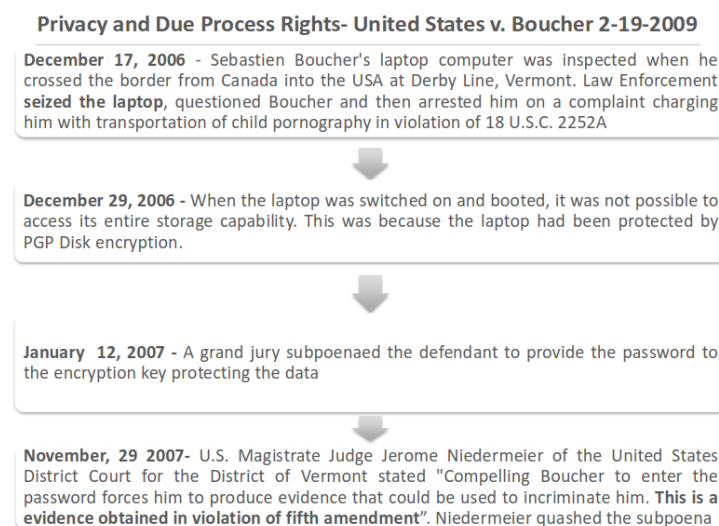


Figure 2.3: Case correlated with the use of encryption

2.3.2 Case Law on Encryption

Another the previous case, some states are starting to create law about “Mandatory Key Disclosure” that force the suspect to give the key/password to decrypt the data. (some are Australia, Belgium France etc. . .)

2.3.3 Mandatory Key Disclosure Laws

These cases of legislative instrument doesn't work for two main reasons:

- **technical reason:** An expert could always find a way to hide a file
- **Possible violation of European Convention on Human Rights:** Article 6 Everyone charged with a criminal offence shall be presumed innocent until proved guilty according to law

2.3.4 Remote Forensics

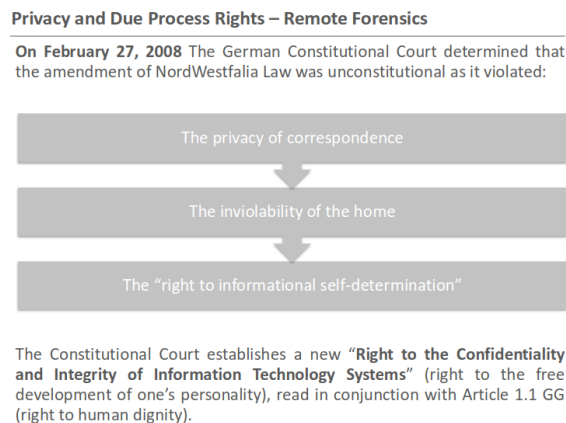


Figure 2.4: Case correlated to remote forensics

2.3.5 Cloud Computing

Cloud computing services face two key legal challenges: **Jurisdiction** and **Privacy**.

Jurisdiction The “**loss of location**” of digital evidence in the cloud introduces significant jurisdictional issues. In a cloud environment, the question arises: are the documents governed by the laws of the state in which they are physically located, the location of the company possessing them, or the laws of the state where the individual resides?

Over the last few years, various legal frameworks and approaches have been proposed to address this complex issue, but it remains an area of ongoing debate.

Privacy Cloud computing introduces several privacy concerns, including:

- **Lack of Control:** Cloud clients may no longer maintain exclusive control over their data, limiting their ability to implement necessary technical and organizational measures to comply with **Data Protection Laws**.
- **Absence of Transparency:** Cloud providers may not provide sufficient information regarding how data is processed, leading to significant risks in terms of compliance with data protection regulations.

Lack of control over the data



Jurisdiction

In addressing the “loss of location” issue within the realm of cloud computing, we have four possible legal principles that can be applied:

- **Territorial Principle:** The court in the jurisdiction where the data is physically located has authority. Jurisdiction is determined based on the geographical location of the data.
- **Nationality Principle:** The nationality of the perpetrator is used to establish criminal jurisdiction. The legal framework of the country of the individual committing the crime applies.
- **Flag Principle:** This principle applies to crimes committed on ships, aircraft, and spacecraft. They are subject to the jurisdiction of the country whose flag the vehicle flies under.
- **Power of Disposal Approach:** This approach focuses on who has control over the data. A regulation based on this would enable law enforcement to access a suspect’s data in the cloud, regardless of its physical location, by considering the individual or entity with power over the data.

Chapter 3

Convention on Cybercrime

3.1 E-commerce on Dark Web

The Dark Web provides a platform for buyers and sellers to engage in e-commerce transactions, often involving illicit goods and services. This anonymous marketplace operates with the same principles as traditional e-commerce, but with heightened security and privacy measures to conceal identities.

Vendors on the Dark Web offer a wide range of products, from drugs and weapons to stolen data and hacking services. Buyers can browse listings, read reviews, and complete purchases using cryptocurrencies, all while maintaining a high degree of anonymity.

3.1.1 Silk Road

Silk Road, often referred to as the "eBay of drugs," was an online marketplace that facilitated the sale of a wide range of illegal substances, including narcotics and controlled substances. At its peak in 2013, Silk Road had a reported annual revenue of \$89.7 million

- **Combining Tor, PGP, and Bitcoin:** Ross Ulbricht leveraged the anonymity of Tor, the encryption of PGP, and the decentralized nature of Bitcoin to create the Silk Road marketplace.
- **Bitcoin-only Payments:** Silk Road required all transactions to be conducted using Bitcoin, providing an added layer of anonymity and making it harder to trace purchases.
- **User-friendly Interface:** Silk Road featured a well-designed interface that allowed users to easily navigate the site and leave feedback on their transactions.
- **Intermediary Role:** Silk Road acted as an intermediary, handling the payment processing and logistics of shipping items purchased on the marketplace.

Silk Road Investigations

- **Operation "Marco Polo":** Undercover agents from DEA and Secret Service involved in extorting money from Ulbricht and attempting to threaten him.
- **Silk Road's Scope:** At its peak:
 - 950,000 registered users
 - 1.2 million transactions
 - \$79 million in commissions
- **Incriminating Errors:** Ulbricht made several mistakes that led to his arrest, including using his real email address and have counterfeit documents delivered to his home.

Charges Against Ross Ulbricht

- **Summary of Charges:** Ulbricht faced 7 key charges, including drug trafficking, money laundering, and computer hacking, which were consolidated into 3 main counts against him
- **Legal Process:** The trial lasted just 13 days and resulted in Ulbricht's conviction and life sentence, plus \$180 million in damages

The investigation was possible because involve US citizen, US platform, and US law enforcement. In another country, the same investigation would have been more difficult.

3.2 History and objectives of the Convention on Cybercrime (Budapest Convention)

The convention involve 65+ states, and its main objectives are:

- Harmonizing national laws on cybercrime
- Improving investigative techniques
- Increasing international cooperation

3.2.1 Overview of the Convention

Timeline and Ratification

The Council of Europe \neq Europe Union, and it's composed by state that are part of the European continent, and its conventions are open to non-European countries.

- Full Adoption: Committee of Ministers of the Council of Europe, November 8, 2001
- Signature: Budapest, November 23, 2001
- Entry into Force: July 1, 2004
- Participating States (as of April 2023):
 - 68 States have ratified
 - 2 States signed but not ratified (Ireland, South Africa), so for them is not binding

Criticism and Opposition

- **India:** Initially refused to adopt due to non-participation in drafting
- **Reconsideration** (since 2018): Surge in cybercrime, but concerns about data sharing with foreign agencies remain
- **Russia:** Rejected due to concerns about sovereignty, limited cooperation in international investigations, even after some articles were revised to address these concerns

New Global Cybercrime Treaty (UN, August 8, 2024)

Content:

- Criminalization of unauthorized access to information systems
- Crimes related to online child exploitation and non- consensual explicit content distribution

Criticism

- Concerns over human rights and press freedom
- Issues with data privacy and overly broad definitions of cybercrime

3.2.2 Aim of the Convention

The convention aims to assist in combating crimes that are inherently linked to the use of technology, where devices serve as both the tool for committing the crime and the target of the crime

Note: there are differences when a crime is committed through a technology tool, and when the technology is the target of the crime.

as well as crimes where technology is used to enhance other criminal activities, such as fraud. It provides guidelines for countries to develop domestic cybercrime laws and acts as a foundation for international cooperation between parties.

The **first additional protocol** focuses on criminalizing the dissemination of racist and xenophobic material through computer systems, along with threats and insults motivated by racism and xenophobia.

The **second additional protocol** establishes common international rules to enhance cooperation on cyber-crime, particularly in the collection of electronic evidence for criminal investigations and proceedings.

3.2.3 Key points

Original Convention (1 July 2004)

- The convention covers:
 - The criminalisation of conduct: ranging from illegal access, data and systems interference to computer-related fraud and dissemination of child abuse material;
 - Procedural powers to investigate cybercrime and secure electronic evidence in relation to any crime;
 - Efficient international cooperation between parties.
- Parties are members of the Cybercrime Convention Committee and share information and experience, assess implementation of the convention or interpret the convention through guidance notes.
- Of the 27 Member States, 26 have ratified the convention - Ireland has signed but not yet ratified it.

Additional Protocol 1 (1 March 2006)

This protocol extends the scope of the convention to cover xenophobic and racist propaganda disseminated through computer systems, providing more protection for victims. It furthermore:

- Reinforces the legal framework through a set of guidelines for criminalising xenophobia and racist propaganda in cyberspace;
- Enhances the ways and means for international cooperation in the investigation and prosecution of racist and xenophobic crimes online.

Additional Protocol 2 (8 August 2024)

This protocol aims to further enhance international cooperation.

It addresses the particular challenge of electronic evidence relating to cybercrime and other offences being held by service providers in foreign jurisdictions, but with law enforcement powers limited to national boundaries. Its main features are:

- A new legal basis permitting a direct request to registrars in other jurisdictions to obtain domain name registration information
- A new legal base permitting direct orders to service providers in other jurisdictions to obtain subscriber information
- Enhanced means for obtaining subscriber information and traffic data through "government to government" cooperation
- Expedited cooperation in emergency situations including the use of joint investigation teams and joint investigations. Key points of the Additional Protocol 2 (8 August 2024)

3.3 Harmonization of national laws and international cooperation

3.3.1 International Cooperation Provisions

- **Cooperation Principle:** Parties are to cooperate "to the widest extent possible" in investigating electronic evidence.
- **Expedited Mutual Assistance:** Issue with Current Mechanisms: Mutual assistance requests are often slow and take months.
- **Convention solution:** Allows for expedited requests using "expedited means of communication" and Expedited means must provide adequate levels of security and authentication. (So can be used encrypted physical device for data transfer)
- **Voluntary Information Sharing:** Parties may share information without a formal request if it would assist in investigations or help the receiving party with any related offences.

3.3.2 Mutual Assistance Provisions

- **Procedural Powers for Assistance:**
 - Expedited Preservation of stored computer data.
 - Expedited Disclosure of traffic data.
 - Real-time Collection of traffic data and interception of content data: parties provide assistance according to domestic laws and applicable treaties, subject to any reservations.
- **Art 23 - General Cooperation Principle:**
 - Mutual assistance "to the widest extent possible" for:
 - Cyber-related offences.
 - Collection of electronic evidence for any criminal offence.
- **Restrictions:**
 - Cooperation may be restricted in cases of:

- Extradition.
- Mutual assistance regarding real-time collection of traffic data.
- Interception of content data

3.3.3 24/7 Network for Immediate Assistance

Provision for Constant Availability:

- Each party must designate a contact point available 24/7
- Purpose: Provide immediate assistance for cybercrime investigations, proceedings, or the collection of electronic evidence
- Based on the G8 network of contact points model

Significance: aims to expedite the processing of urgent mutual assistance requests, overcoming current delays in traditional bureaucratic channels

3.4 Legal measures against computer-related fraud and forgery

Part II

Tech

Chapter 4

Introduction

4.1 Topics

- **Forensics Analysis**
use of logic and meaningful knowledge and methodological approach to legal problems and criminal investigation.
- **Computer Forensics**
Collection, preservation and analysis of digital evidence (inside file system, email, cloud account etc...) to support investigation and legal proceedings

4.2 Forensics History

4.2.1 Ancient Times

Forensic science dates back to **Babylon (1900 BC)** where fingerprints were used for identification, and **China (1248 AC)** with forensic pathology. In the **UK (1835)**, bullet comparison solved a case, and by **1892**, the first murder was solved using fingerprints.

4.2.2 Modern Times

Forensic standards grew in police departments, with the first crime lab in **1923**. DNA fingerprinting began in the **1950s**, and DNA profiling was developed by **1985**. **AFIS** systems emerged in the late **1980s**. Today, AI, toxicology, and digital forensics are key areas of innovation.

4.2.3 Digital Field

Early Times

In **1989**, Robert Morris was convicted under the Computer Fraud and Abuse Act, marking the first use of computer logs in forensics. That year, **IACIS** was founded, followed by **IOCE** in **1995** to share digital forensic practices.

Recent Times

- **1990**: Forensic tools like EnCase emerged
- **2000**: Digital forensics became widespread in law enforcement

- **2010:** Growth of cloud and mobile forensics, automation, and machine learning
- **2020:** Advances in crypto, blockchain, and AI improve digital forensics

4.3 Computer Forensics Definitions

US_CERT: The discipline that combines elements of law and computer science to collect and analyze data from computer systems, networks, wireless communications, and storage devices in a way that is admissible as evidence in a court of law.

A. Ghirardini -Computer Forensics: The discipline whose goal is preservation, identification, analysis of information system to the aim of identification of evidences during investigation activities.

NIST glossary: The application of computer science and investigative procedures involving the examination of digital evidence - following proper search authority, chain of custody, validation with mathematics, use of validated tools, repeatability, reporting, and possibly expert testimony.

4.4 CF purpose(s)

4.4.1 CF Q&A

During an investigation, digital forensics need to analysis data to answer some key questions:

- | | |
|----------------------------------|-------------------------------------|
| • What happened? | • Where did it take place? |
| • Who was involved? | • Why did it take place? |
| • When did it take place? | • How did an incident occur? |

The answers to these questions are essential for support legal proceedings and mitigate possibility of future incidents with a preventive approach.

4.4.2 CF Goals

The goals of computer forensics (CF) are multifaceted and aim to provide a comprehensive understanding of digital incidents.

Firstly, CF seeks to retrieve what has been the input, such as what has been typed. It also aims to determine the actions performed, for example, what programs have been run and what peripherals have been connected. Additionally, CF involves analyzing used files to understand what modifications have been done and when these modifications occurred (and the information from an OS are not enough, because are an abstraction managed by the file system → needed bit analysis (like for ereased data)).

Another critical goal is to identify the damage done, such as what data have been erased. In essence, the overarching goal of CF is to **gain a technical comprehension of what happened during the incident** (from a technical point of view).

Chapter 5

CF Terminology & relevant concepts

5.1 Terms

5.1.1 Digital evidence

Technical definition of Digital evidence is very similar to the legal ones:

Data stored or transmitted in digital form that can be used in court.

The cornerstones of digital forensics are the different levels of **abstraction**, requires **interpretation**, are **fragile**, may be **voluminous** and the difficulty to discover **connection** between data and reality (the connection need to be done before entering the court).

Digital evidence also requires a deep technical understanding of the possible types of data (files, emails, logs, metadata) and the legal requirements for each of one to collect and preserve it. (To make all of this effective, knowledge of file systems, network protocols and encryption are essential)

5.1.2 Chain of custody

Documented and **unbroken** process of handling evidence from the time it is collected until it is presented in court

This procedure is essential to ensure the integrity of the evidence and to avoid that them to be tampered or accessed by unauthorized people.

Keep the chain of custody requires knowledge about how to document evidence collection, storage, and access (logging procedures, secure storage, legal protocols etc..)

If the chain of custody is **broken**, the evidence may be considered **inadmissible in court** (so is needed know the regulation of the state to decide how to manage it).

5.1.3 Data acquisition

The process of collecting digital evidence from devices without altering or damaging the original data.

One of the biggest problem for the management of digital evidence, because it's needed to be performed on hostile systems (that can be infected, compromised, have a malware or system to avoid copy like edited system call for make other program to fail). in a not controlled environment like a crime scene. So are needed knowledge of disk imaging and live data capture in order to not alter what's going on on the suspect system. Are also required expertise in forensics acquisition, analysis tools (like FTK Imager, EnCase) and knowledge of file systems, write-blockers, and hashing (crucial for ensuring integrity).

5.1.4 Hashing

The process of converting data into a fixed-length string of bits, which represents the data uniquely

It's used in the chain of custody for ensure the integrity of the digital evidence and so verify that a file has not been altered.

Require understanding of hashing algorithms (strengths and weaknesses, e.g. MD5 collision), formats (hex, base64 etc...) (if wrong formats are used, the chain of custody is broken and each information gathered from that point is considered not valid) and expertise in hashing tools (sha256sum, hashdeep, FTK imager, Autopsy).

Have to be used any time an evidence is "managed" (copied, moved)

5.1.5 Write Blocker

Hardware or software tool used to prevent any data from being written to a storage device during analysis, preserving the original data content

To be operated, require understanding of how write-blocking devices work and how they can be implemented in forensic procedures.

It's essential for the legally defensible acquisition.

5.1.6 Forensic image

A bit-by-bit copy of digital media, including deleted files and data in slack space, which is an exact replica of the original device

The goal of a forensic image is to preserve the original evidence and avoid the modification of the original data.

To be performed in a correct way, requires understanding of mechanisms to copy information in digital devices (file system knowledge and behavior) and familiarity with bit-by-bit copy tools (DD, FTK Imager, EnCase, Guymager).

As the hashing, it's need to be used any time an evidence is "managed" (copied, moved)

5.2 Scenarios

There are some possible scenario that a computer forensic investigator can face:

- Internet abuse from employee
- computer-aided frauds
- Data unauthorized manipulation (theft or destruction)
- Computer/network manage assessment
- ...any other case that include digital evidence

5.3 investigation phases

A Computer forensics investigator usually follow standard phases that guide him. There are different standards like: NIST family, ACPO guidelines (UK), ISO/IEC 27042, SWDGE.

5.3.1 Phases

- **identification:** When the investigator come for the first time to the crime scene and need to identify potential source of relevant digital evidences.
- **collection:** The letteral pick up of the evidence (like a computer or a smartphone) or a remote taking possession of the evidence (like for a remote server) and its connection (e.g. network or physical, like USB disk).
It's splitted from acquisition because it's a critical phase where the evidence can be altered and lost utility for the investigation (es. data corruption, lost of metadata etc...)
- **acquisition:** Electronically retrieving data by running various CF tools and software suites
- **evaluation:** Evaluating the data recovered to determine if and how it could be used against the suspect (e.g. for prosecution in court)
- **presentation:** Presenting the evidence discovered in a manner which is suitable for lawyers, non-technical staff/management and the law (and internal rules)

5.3.2 Identification

During the identification phase is important **recognize** all the **relevant data sources** before any acquisition, even if no physical present, like data in the cloud

A imple **list of example** are: hard drives (HDD/SSD), memory (RAM), mobile devices (smartphones, tablets), cloud storage, network traffic, removable media (USB drives, DVDs), IoT devices and embedded systems (like smart washing machines)

For identify these sourcer, the investigator can perform some actions, like:

- Perform an initial survey of the scene (physical or network environment)
- Identify key devices and data locations (local storage, remote servers, cloud services)
- Check for connected devices, including peripherals like printers, removable media, or network-attached devices
- Map all potential data sources using network topology diagrams or asset inventories

A particular aspect that need to be considered is the possible present of "ephemeral" storage or data, like cloud syncing, hidden sector, tmp, dat in ram etc...

5.3.3 Collection

During the collection, the focus is on gathering evidence from identified data sources while ensuring the preservation of its integrity. An important key point is the implementation of methods that **minimize the risk of evidence tampering or data loss**.

For enforce this key point, it's important to **isolate devices** to prevent them from being tampered with remotely (e.g., disconnect them from the network), use devices to **block external communication** for mobile or wireless devices and use network isolation tools for virtual and cloud environments to prevent remote access (like use a virtual private cloud).

A particular note is for the management of live systems where is needed ensure evidence integrity while maintaining system uptime (so not shutting down the system for avoid the loss of volatile data).

Create a **detailed record** of the condition and state of the evidence

- take photographs of the devices in situ, including connected peripherals and the physical state
- record serial numbers, device models, and any other identifiable information
- document the scene, noting which devices were running, whether screens were active or locked, and any other visible indicators

hint: complete documentation is crucial to prevent legal challenges regarding the integrity of the evidence. Before proceeding to the acquisition, is needed to ensuring no alteration will take place, so do some things like enable write blockers for physical storage devices, disable connection and syncing. particularly complex is maintain integrity on live systems (e.g., using remote collection methods that minimize data alteration risks)

5.3.4 Acquisition

The act of performing a forensic copy (so a bit-by-bit copy) of the original data with the goal of ensure that the acquired data is a faithful replica of the source so to maintaining data integrity.

There are two main acquisition methods:

- **Static:** When the system is powered down, it's the most common method for acquiring data from hard drives and external memory
- **Live:** The system is running and it's needed to deal with volatile data like RAM, network connections, or running processes.

Static acquisition

1. shut them down carefully to avoid losing data
 - e.g. for encrypted devices, consider methods for capturing data without triggering loss of access (e.g., before the decryption key is wiped from RAM)
2. attach the device to a forensic workstation using a write blocker
3. use forensic imaging tools to create a complete image of the storage device
4. generate a hash value (e.g., SHA-256) of the original media before and after acquisition to verify integrity
5. store the image on a secure forensic storage device

hint: pay attention that data is properly hashed and verified post-acquisition, not perform steps like an automata.

Live acquisition

1. choose a method that minimizes system interference while capturing volatile data
2. dump RAM (memory acquisition) and capture data from running processes or network connections.
3. perform network traffic capture
4. document all acquisition actions and steps to ensure chain of custody and admissibility
5. hash the volatile data wherever possible to maintain data integrity

Integrity

In this phase is needed to ensure that the acquired data is an exact replica of the original and has not been altered.

Performed mainly by the use of hashing algorithms.

Some general steps are:

- choose a method that minimizes system interference while generating a hash (MD5, SHA-256) of the acquired image or data dump
- compare the hash value to the original data hash (for static data) to verify its integrity
- document the hashing process, including the algorithms used and the results, in the chain of custody documentation

hint: be careful! any discrepancies in hash values would require re-acquisition and could damage the credibility of the evidence

Chain of custody

This section need to be performed in paralld with all the other phases to ensure a complete, documented chain of custody for the evidence throughout the acquisition process. (Record every step in the acquisition process, including personnel involved, tools used, date, and time of acquisition.

Store the data and evidence securely to avoid unauthorized access or tampering)

5.3.5 Evaluation

analyzing, verifying, and validating the evidence to ensure it remains unaltered and trustworthy for legal proceedings or further analysis. It's possible alter the evidence only if the evaluation is performed on a copy of the original data.

More in practice, the main actions that are performed are:

- timestamp and metadata analysis
 - verify file creation, access, and modification dates of data to ensure they match the timeline of the incident
- timeline reconstruction
- cross-reference analysis/consistency verification
 - correlation of digital evidences with external logs or other data to countercheck it is related to the suspected system or device

- comparison of data from different sources (e.g. logs, email)
- compliancy with current legal/internal standards
 - collection, preservation, evaluation must be coherent to applicable legal procedures...and the documentation must keep track of that
- review of possible anti-forensics techniques

5.3.6 Presentation

Preparing and presenting the findings of the investigation in a clear, accurate, and legally admissible manner is essential. The goal is to **translate** the technical details of the forensic analysis into a format that can be understood by **non-technical stakeholders**, such as lawyers, judges, or company executives.

Hint: The quality and clarity of the presentation can significantly influence the outcome of legal proceedings or internal investigations.

Actions:

- Review all the data collected, analyzed, and interpreted during the investigation.
- Identify the key pieces of evidence.
- Verify that all conclusions are directly correlated to verifiable evidence.
- Document the entire forensic process in a formal report, free from technical jargon, so that it can be submitted as legal evidence.
- Securely manage the report to ensure its integrity.

Hints:

- Avoid "personal interpretation" unless explicitly asked to provide expert opinion.
- Include appendices with timestamps, metadata, hash values, and other forms of technical evidence as "reinforcement."

Chapter 6

Non-trusted environment issues

We need to not trust an environment by default, because it can be compromised, and there are many ways to do so.

6.1 Compromise causes

6.1.1 Node infection

Nowadays, a node infection is obtained through a social engineering attack, that lead to the download of a compromise file/software.

- Legitimate software containing malicious code (trojan horses) (a free version of paid software is always a good bait), social engineering, physical access, bug or configuration error exploitation (OS syscall, device driver, application, firmware and BIOS, browser ...)
- Backdoors creation, data stealing, hidden (or not so much) processes disruption, ...
- Persistent unauthorized access to a system (as root - i.e. rootkits)
- Spyware (sensitive information collection)
- Ransomware (encryption of sensitive data)

6.1.2 network injection

- nodes capable to read and write data while in transit, actors capable to "poison" routing mechanisms
- access and modification of network data flow, redirection versus illegitimate destination
- Sniffers and (growing) family of Man-in the-X attacks_{6.1.4}

6.1.3 supply chain attacks

- compromise of service, hardware, software of a third-party vendor or partner used (and trusted) by the target organization
- gain access to the target organization, inject unauthorized behavior
- infrastructure for update management
- – e.g. SolarWind Orion Attack
 - malicious code into software updates of Orion network monitoring platform.
 - distributed to over 18,000 customers, including government agencies and large corporations.

- libraries and dependencies
- hardware during manufacturing
- IT infrastructure management service
- ...

6.1.4 Men at work

man-in-the-middle

An attacker secretly intercepts or alters communication between two unaware parties.

Examples include **HTTP session hijacking**, where the attacker intercepts session cookies to impersonate a user, and **ARP table poisoning**, where ARP tables are altered for traffic redirection.

man-in-the-browser

Infection occurs in the browser to alter web pages or transactions.

An example is banking trojans like **ZEUS**, which modify online transactions.

man-in-the-cloud

This involves stealing credentials or tokens to access a user's cloud environment.

For example, the interception of a Google Drive OAuth token can allow access to the victim's files.

man-in-the-mobile (MitMo)

Mobile infection is used to intercept communication or two-factor authentication (2FA).

An example is ZitMo, which intercepts SMS and forwards them to a command and control (C&C) server.

man-in-the-disk

This exploits vulnerabilities in handling external storage.

For instance, an attacker can modify temporary files stored on an external device.

man-in-the-memory (MitMem - guest star)

In this case, an attacker intercepts or modifies data while it is in RAM.

A notable example is fileless (stealth) malware.

man-on-the-side

An attacker observes and injects communication without modifying it.

An example is China's Great Cannon.

man-at-the-end

This type of attack compromises end-point communication.

For example, a keylogger infection can capture sensitive information.

6.2 Advanced persistent threats (APT)

- **advanced**
 - use of sophisticated techniques
 - * customised malware, zero day vulnerabilities, evasion strategies
 - targeted to specific victim
 - * high budget and expertise, careful preparation
- **persistent**
 - Item compromise maintained for extended period
 - * possible escalation and infection diffusion
 - low-profile operation (during infection)
 - * stealth techniques, limited bandwidth usage, mimicking legitimate traffic
- **threat**
 - highly skilled individual aiming strategic goals (espionage, foreign country intelligence, ...)

6.2.1 APT Attack Process

The Advanced Persistent Threat (APT) attack process consists of several key stages:

Initial Intrusion

The attacker gains access through a weak entry point, such as exploiting zero-day vulnerabilities or using spear phishing techniques to infiltrate the target system.

Foothold Establishment

Once access is gained, the attacker sets up persistent access by installing backdoors or infecting the system with (stealth) malware to maintain control over the compromised environment.

Privilege Escalation

The attacker escalates privileges to gain further control over the target system. This involves techniques like credential stealing or vulnerability exploitation.

Lateral Movement

The infection spreads across the target organization as the attacker moves laterally (like over different device/account with same/similar level of privilege), using stolen credentials (social eng.) or exploiting vulnerabilities to compromise additional systems.

Goal Achievement

The attacker eventually reaches their goal, which often involves data exfiltration or sabotaging critical systems.

6.2.2 manipulation from the system owner

If the system owner is technical-savvy, he can manipulate the system to hide the compromise, or to make it more difficult to detect, by installing modified application, compromised drivers or edit system calls.

6.2.3 APTxx

APTxx refers to organized hacker groups involved in advanced persistent threat (APT) activities. An example of such a group is **APT28**, also known as Fancy Bear.

APT28 (Fancy Bear)

APT28 is a **Russian state-sponsored group** that operates during Russian business hours and closely aligns with Russian government strategic interests, particularly in regions like the Caucasus.

The group has been active since the mid-2000s, with documented operations dating back to at least 2008. APT28 targets a wide range of sectors, including aerospace, defense, energy, government, media, and dissidents, engaging in activities such as espionage, political influence, and cyberwarfare.

Notable Attacks: In 2016, APT28 was responsible for the breach of the **Democratic National Committee** (DNC) during the U.S. presidential election. This attack led to the leakage of sensitive information with the intent of influencing the election outcome.

Another major attack occurred in 2017 with the NotPetya ransomware, which was initially designed to target **Ukrainian institutions**. However, the malware spread globally, causing billions of dollars in damages.

APT28 Typical Behavior

APT28 targets a wide range of devices, including desktops, laptops, and mobile phones. It often employs (*spear-*)*phishing* messages to direct victims to realistic websites for credential harvesting.

- APT28 registers domains that closely resemble legitimate organizations (e.g., `gov.hu.com` for the Hungarian government `gov.hu`).
- It uses URL-shortening services to obscure the true destination of malicious links.

In addition, APT28 delivers highly-realistic and targeted emails, often containing "weaponized" attachments such as `.docx` or `.pdf` files.

They are also used to implant custom malware, such as **X-Agent**, a multi-functional malware implant used for:

- Data exfiltration,
- Keystroke logging
- Multiplatform operations (Windows, Linux, Android, and iOS).

After gaining initial access, APT28 actively seeks to harvest credentials through techniques like keylogging and central memory dumping. To evade detection, APT28 adopts various **evasion techniques**, including:

- Malware code obfuscation,
- Use of compromised certificate signatures,
- *Timestomping* (modifying timestamps), and
- Encrypted communication channels.

APT28 also engages in **lateral movement** within the compromised organization by exploiting harvested credentials. This lateral movement involves:

- Remote Desktop Protocols (RDP),

- Windows Management Instrumentation Command-line (WMIC) and **PsExec** to execute commands on remote Windows machines, and
- **SSH** to connect to remote Linux systems.

At this point, APT28 escalates privileges by exploiting harvested credentials or vulnerabilities in the system.

Finally, engages in **data exfiltration** using custom **Command-and-Control (C2)** communication frameworks, such as **Zebra C2**. The exfiltrated data may be optionally compressed, especially if large, and is transmitted via encrypted channels like **HTTPS**, **FTPS**, or even custom protocols.

Although primarily known for espionage, APT28 has also been involved in **destructive attacks**. These include the use of **wiper actions**, such as:

- **KillDisk**, designed to destroy the master boot record, and
- Disk wiping tools, particularly in the energy sector.

6.3 Trusted Environment

The analysis must be performed in a **trusted environment**, as rootkits can **alter the normal behavior** of the operating system, making traditional tools unreliable.

Rootkits are capable of **modifying file system utilities**, such as: **ls**, **cp**, **mv**, and other basic commands.

Additionally, rootkits can intercept and **modify file system calls**. For example, they may intercept system calls like **open()**, **chdir()**, or **unlink()** to avoid displaying or acting on specific files, making it difficult to detect their presence.

6.3.1 (example of) System Call Interception

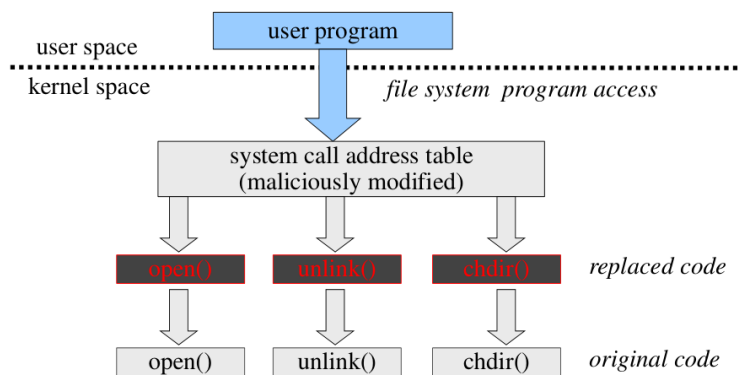


Figure 6.1: Example of System Call Interception

6.3.2 Examples of Linux system modification

A common method of modifying a Linux system is through the use of **Loadable Kernel Modules (LKM)**. This concept is not unique to Linux and can be found in many other operating systems, such as kernel extensions in macOS or kernel-mode drivers in Windows.

An LKM can override the original system call (*syscall*) function. The typical steps to achieve this include:

- **Develop** a modified version of the system call function.
- **Modify** the system call table, which is an array of function pointers, to point to the new version of the function.
- If you want to **modify behavior**, you can re-implement the function with the desired changes.
- If you want to **add functionalities**, enrich the function with additional features and then call the original one to preserve its behavior.

This allows for either enhancing the system with new capabilities or subtly altering existing functionalities without being easily detected.

```
#include <linux/kernel.h>
#include <linux/module.h>
#include <linux/syscalls.h>
#include <linux/uaccess.h>
asmlinkage int (*original_open)(const char __user *filename, int flags,
mode_t mode);
asmlinkage int custom_open(const char __user *filename, int flags, mode_t
mode) {
    printk(KERN_INFO "Intercepted file open: %s\n", filename);
    return original_open(filename, flags, mode);
}
static int __init syscall_init(void) {
    original_open = (void *)sys_call_table[__NR_open];
    sys_call_table[__NR_open] = custom_open;
    return 0;}
static void __exit syscall_cleanup(void) {
    sys_call_table[__NR_open] = original_open;}
module_init(syscall_init); module_exit(syscall_cleanup);
MODULE_LICENSE("GPL");
```

Figure 6.2: Examples of Linux system modification