Digital Forensics: Essentials and Data Acquisition

Thematic CERN School of Computing, 2026

Daniel Kouřil
kouril@cesnet.cz

MUNI

EGI CSIRT

cesnet

Digital Forensics

- Methods to collect, preserve and analyze *digital artifacts and evidence*
- Three main phases
  - Acquiring the primary data
  - Analysis and evaluation (establishing the evidence)
  - Reporting
- The first phase is most crucial
  - Must make sure the data is complete, authentic and its integrity can be checked
  - The other phases can be repeated/corrected/scrutinized any time

Examples of investigations

Burst of a ransomware campaign, a self-propagating worm exploiting a zero-day vulnerability in the operating system. After it is executed, the worm blocks access to the data on the disk and asks for ransom.

- Micro Enterprise
  - Can data be recovered?
  - Have data been modified, tempered with?
- Large Organization
  - How the attack is spreading and how to spot it?
  - Have any sensitive data leaked?
- Law Enforcement
  - Is it possible to identify the attacker?
  - Is the determined evidence admissible?

Digital Evidence

- Evidence is determined based on acquired artifacts
  - Digital artifacts are data
- Can be literally anything
  - Files on the storage, memory contents, metadata
  - Computer isn't the only source of data
- Data is digital
  - Potentially hard to get
  - Easy to distort

Today's focus

- We focus on system administrators who want to secure forensically sound data
  - Not aiming at acting in "hostile environment" (like LEA, etc.)
  - The environment is supposed to be known and cooperative (mostly)
    - Architecture details can be (easily) established
    - System passwords and keys are known / available, etc.
- We focus on demonstrating principles using common tools

Data Acquisition

Basic principles

- Common principles
  - The collected data should be:
    - complete (for subsequent analysis)
    - accurate (not altered)
  - “We want to get the most evidence we can with the least amount of alteration”
- Every investigation should be scoped (questions formulated, at least internally)
- Only start the acquisition process if you’re authorized!
- Prepare in advance
- The quality of the data and the soundness of the acquisition process determines the utility of the evidence

Risks of low quality of evidence

- If the evidence is incomplete, it cannot yield relevant outcome
- If the acquisition process doesn't guarantee the integrity and authenticity of the data, results can be disputed
  - Never know when the data will need to be defended, e.g. an internal process with an employee can end up in court
- If you don't know what to do, do not interact with the system at all
  - No commands typed, no programs started, no new logins, …

Getting data from computer

Machine is switched off

- The off-line approach is straightforward
  - The only evidence is on permanent storage
  - The device can be dismounted and processed outside the computer
    - Works for virtual machines as well
- Or the computer can be booted from a trusted media
  - The computer must never boot other than trusted OS (e.g. USB), esp. the OS installed in the host!

Never boot the machine OS

- Windows 10 changes or creates a lot of system files, e.g.:

/Windows/System32/LogFiles/WMI/NetCore.etl
/Windows/System32/LogFiles/WMI/NtfsLog.etl
/Windows/System32/LogFiles/WMI/Wifi.etl
/Windows/System32/LogFiles/WMI/RtBackup/EtwRTEventLog-Application.etl
/Windows/System32/LogFiles/WMI/LwtNetLog.etl
/Windows/System32/LogFiles/WMI/Microsoft-Windows-Rdp-Graphics-RdpIdd-Trace.etl
/Windows/System32/WDI/LogFiles/WdiContextLog.etl.001
/hiberfil.sys
/pagefile.sys
/swapfile.sys
/Program Files/AMD/atikmdag_dce.log
/Windows/System32/LogFiles/WMI/RtBackup/EtwRTUBPM.etl
/Windows/debug/PASSWD.LOG
/Windows/bootstat.dat
/Windows/System32/SleepStudy/UserNotPresentSession.etl
/Windows/ServiceProfiles/NetworkService/NTUSER.DAT{fd9a35da-49fe-11e9-aa2c-248a07783950}.TxR.blf
/Windows/ServiceProfiles/NetworkService/NTUSER.DAT{fd9a35da-49fe-11e9-aa2c-248a07783950}.TxR.0.regtrans-ms
/ProgramData/Microsoft/Windows Defender/Support/MpWppTracing-20201023-170636-00000003-ffffffff.bin
/ProgramData/Microsoft/Windows Defender/Scans/History/Results/Resource/{4862B78F-8B86-4B07-B4CB-254796EFB69D}
/ProgramData/Microsoft/Windows/AppRepository/Packages/Microsoft.Windows.StartMenuExperienceHost_10.0.18362.449_neutral_neutral_cw5n1h2txyewy/ActivationStore.dat
/Windows/System32/winevt/Logs/Microsoft-Windows-LiveId%4Operational.evtx

What if the computer is turned on?

- Never shutdown the computer from the system
  - Files get changed during the process (similarly to the booting process)
- If you want to proceed with off-line approach, pull the power plug
  - But think twice before you do

Why (not) to pull the power plug?

- If you pull the plug
  - you don’t risk any change of the evidence
  - you immediately stop any malicious activities
- There is a big disadvantage, though
  - information may be lost immediately by virtue of the volatility of digital data
  - The contents of RAM is lost when power is off, couldn’t be recovered
  - Examples:
    - Data is lost – a mail being composed
    - Crucial information that never hits disk (encryption keys, passwords)
    - System structures – list of running processes, open connections

Live capture

- Accessing a live, running system and collecting volatile information
- Data capture should follow the order of volatility
  - Data is volatile (either frequently changing and/or available only for a limited time)
  - There are different levels
  - Most volatile data needs to be captured first
    - e.g., list of open connections is more volatile than disk contents

Things to remember during live capture

- Minimize all activities on the system
  - (every action leaves/modifies traces)
- A running system persistently modifies itself even without investigation activities
  - keeps producing logs, performing SW updates, …
- Remember the capturing is always mediated by the system that is being investigated
  - Don’t trust the programs on the system
  - Be prepared the collected information might be distorted/hidden
    - Imagine a kernel rootkit hiding certain processes or connections

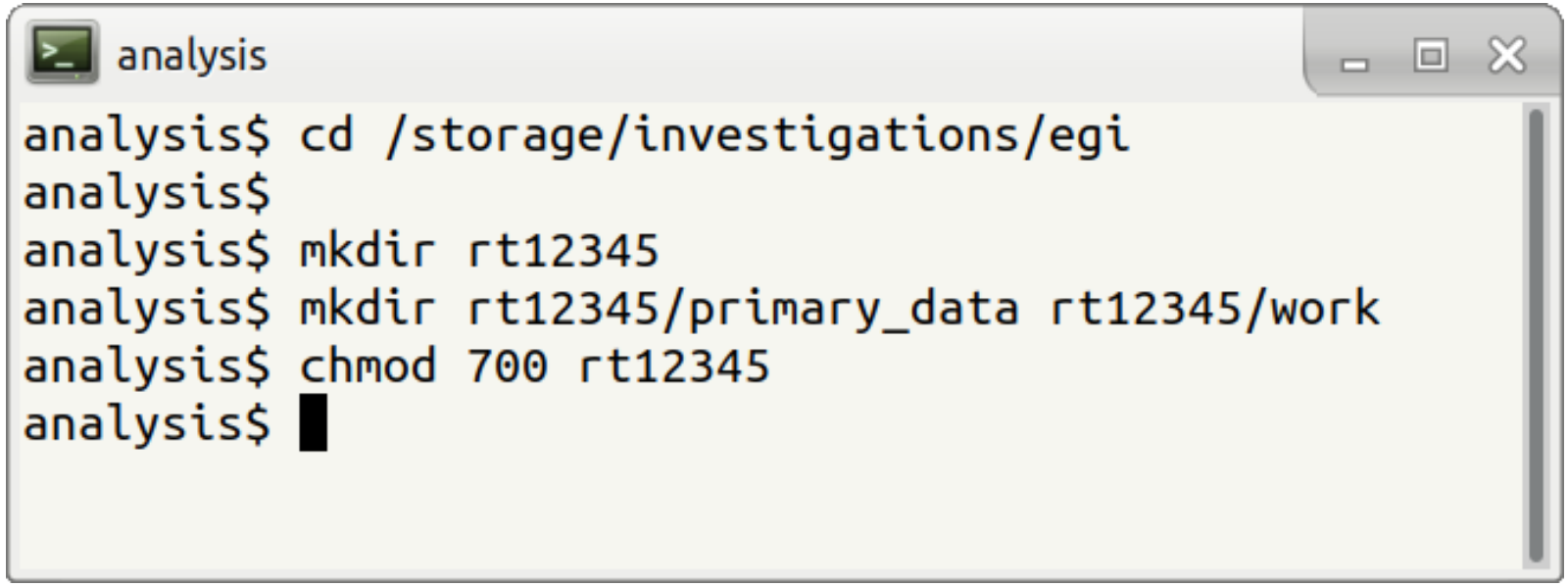
Perils of live acquisition

- Inherently a thin line between investigation and acquisition
  - The system needs to be investigated to establish potential sources of evidence
- It's important to document the process
  - E.g., using video/image records, a log of performed actions, etc.
  - In serious investigations work in a pair (investigating + documenting steps)

Handling primary data

Managing captured data

- Integrity protection and authenticity is necessary
  - Cryptographic hashes with them
- Store primary data as read-only and perform analysis on copies of the data
  - Data might be large, you'll need a double space (to store and analyse)
- After the analysis is done, data needs to be archived


analysis
analysis$ cd /storage/investigations/egi
analysis$
analysis$ mkdir rt12345
analysis$ mkdir rt12345/primary_data rt12345/work
analysis$ chmod 700 rt12345
analysis$

Acquiring evidence from storage systems

Strategies

- Byte-stream copies
  - Source is a whole disk or a volume
  - Target is either image file or another disk
  - Copies are exact (byte) replicas of the original
- Logical (sparse) acquisition
  - Capturing only specific files
  - Takes less time and space
  - Omits some information
- Acquisition of contents metadata
  - Quick, with minimal space requirements
  - Reasonable notion of the whole system
  - Inherently misses information from content

Acquisition in virtual environment

- No need to manipulate with HW devices
  - Snapshots or disk copies can be obtained easily
- It can be obtained from a device or can be retrieved from a storage “manager”
  - Virtualization, containers
  - Different formats are used
    - Must be converted to be usable by common tools

Performing full byte-stream copy

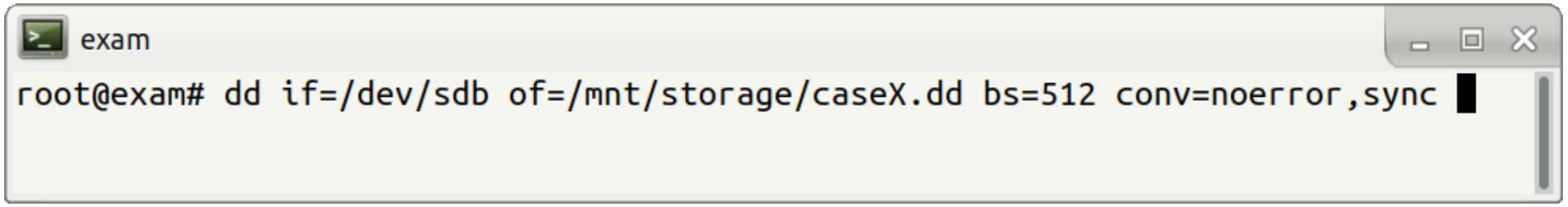
Full byte copy

- The image is exact copy of the source
  - It is a continuous byte stream stored in a single file
- Error handling
  - The source might fail to read some parts of the media
  - The acquisition tool has to handle errors properly, e.g. to fill in failing sectors with zeros (and report the problem)
    - If a sector was skipped, the addresses would change, making it difficult to reconstruct partitions, file systems data etc.

Mirroring

- Connect the source media to a computer and mirror the device directly
  - Where write blocker cannot/should not be used
  - Also usable for other media (USB sticks, memory cards, non-removable storage)
- You can consider to use the existing computer but make sure it boots a trusted OS
- Decide whether to image a whole disk or a particular partition
  - Depends on the goals of investigation and expected sources of evidence

Imaging using dd



- `dd` is a common tool to transfer data between two endpoints (files)
- `if` refers to the source device (/dev/sdb)
- `of` refers to the target (external device)
- `bs` block size (the amount of data to read at once)
- `conv=noerror,sync` makes sure the processing doesn't stop error and failing blocks are filled with 0's

```
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: gpt
Disk identifier: 30C3A121-3D9E-4BAD-BAC5-07E0C31AB47F

Device         Start      End  Sectors  Size Type
/dev/nbd0p1     2048   206847   204800  100M EFI System
/dev/nbd0p2   206848   239615    32768   16M Microsoft reserved
/dev/nbd0p3   239616 15968256 15728641  7.5G Microsoft basic data
/dev/nbd0p4 15970304 16891903   921600  450M Windows recovery environment
forensics# fdisk -l /dev/nbd1
Disk /dev/nbd1: 20 GiB, 21474836480 bytes, 41943040 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x72d129cc

Device      Boot Start      End  Sectors Size Id Type
/dev/nbd1p1       2048 41943039 41940992  20G 83 Linux
forensics# mount /dev/nbd1p1 /mnt
forensics# mkdir /mnt/case
forensics# cd /mnt/case
forensics# mkdir primary_data work
forensics# cd primary_data
forensics# dd if=/dev/nbd0 of=nbd0.dd bs=512 conv=noerror,sync status=progress
8639730176 bytes (8.6 GB, 8.0 GiB) copied, 183 s, 47.2 MB/s
16891937+0 records in
16891937+0 records out
8648671744 bytes (8.6 GB, 8.1 GiB) copied, 183.278 s, 47.2 MB/s
forensics# sha1sum nbd0.dd > nbd0.dd.sha1
forensics# cd /
forensics# sync
forensics# umount /mnt
```