Persistent Storage Acquisition – Part I

The Basics of Copying Data

Tobias Dussa
WP8-T1

Webinar, January 2022

Public

www.geant.org

Game Plan

- The general approach to grabbing a hold of persistent data.
- Discuss easy cases:

  Desktops, notebooks, USB sticks, flash cards, and the like.
- … and some things to watch out for.
- Questions/discussion/open mike session.

STOP! A Word of Warning

We cannot and do not provide any legal counseling!

- If you know or suspect that there will be legal steps taken, talk to a lawyer first.
- Depending on your local legislation, there is a very real possibility that you inadvertently destroy evidence.

Preparatory Remarks and Intro

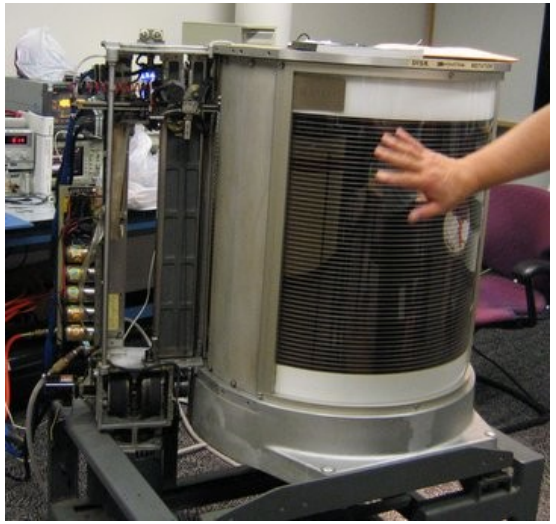
The Bigger Picture

So there is some sort of persistent storage that you would like to analyze.

Some parameters to consider:

- What kind of storage? What size?
- Where is the storage located?
- What is the objective of the analysis?
- How “safe” do you want to play this out?

Plenty of Potential Places of Persistency



General remark: Image attribution on the final slides.

General Observations – Structure

- Mass storage devices typically come with some sort of structure, for example:
  - Partitions,
  - slices,
  - logical/RAID volumes,
  - file systems,
  - virtual images.
- Data can be hidden in any of these layers (on purpose or by happenstance).

General Observations – Structure

- Each layer generally contains “slack space” that is not accessible from “higher” layers.
- Examples:
  - Deleted files,
  - unallocated space in volume groups,
  - space between non-aligned partitions boundaries.
- Also, there may be interesting metadata in “lower” layers.
- All in all, it is advisable to get the data from as low a layer as possible.

General Observations – Size

- Mass storage comes in a huge range of sizes, from (realistically) just a few gigabytes up to several petabytes or more.
- This raises two potential problems:
  - How long it takes to acquire data, and
  - where to store the acquired data.

General Observations – Location

- The location (relative to you) of the desired mass storage device is crucial.
- Can range significantly: You can hold a USB drive in your hand, or you can want to analyze a machine on a different continent half-way around the world.
- (Of course, the issue is not only the geographic, but also the network distance.)

General Observations – Objective of the Analysis

- There are a number of potential objectives:
  - “Just” to find out what happened,
  - to demonstrate and practice due diligence,
  - to aid or guide external investigators,
  - to secure evidence for legal action.
- The more formal the objective, the more restricted your options.

General Observations – Safety Level

- Somewhat related to the objective.
- Questions to ask:
  - Is it acceptable to lose (access to) the acquired data?
    Depends.
  - Is it acceptable to alter (parts of) the acquired data?
    Depends.
  - Is it acceptable to give others access to the acquired data (unintentionally)?
    Almost certainly not.
- Usually, forensics call for a fairly high degree of safety.

So How to Approach The Data

- The advantage of **persistant** storage acquisition: It is persistent. In other words, there is (comparatively) plenty of time.
- (Possible exception: Encryption keys!)
- (Well, and things like **`tmpfs`**.)
- If possible, **physically** get the device.
- If impossible or not feasible, get a master copy of the data, as low-level as possible.
- Then create a working copy.

Let’s Dive In:
How to Grab Data

Get a Hold of the Physical Storage Device

- Getting the physical is immensely helpful.
- Guarantees that you get all the data there is.
- This is not always possible, e. g. because of:
  - Size,
  - distance,
  - importance.
- (Make sure you can then also *read* the device! Must have correct controllers, interfaces, crypto keys, ...)
- Easiest case: USB stick.

Alternatively: Create a Master Copy

- If you cannot *get* the physical media, hopefully you can get **to** the physical media.
- Bring a suitable mobile storage medium.
- … and a suitable boot medium with your favorite toolbox.
- Boot into your toolbox (again: beware of crypto keys!), attach your storage, clone the data onto your mobile storage.
- Treat this clone like the original.

Back Home: Create a Working Copy

- Take the original or the master copy.
- Very carefully clone the data to create a working copy.
- Store the original or master copy *safely* and *securely*.
- Analyze the working copy.
- If you run into problems with the working copy later on, repeat and create a *new* working copy.

“Clone the Data”

- The above steps have used the term “clone the data” without further ado.
- Generally speaking, this means “create a bitwise copy of the raw data”. This can be achieved by a number of ways, e. g.:
  - **cat** (most versatile, almost always available, cannot handle errors)
  - **dd** (can be very much parametrized, almost always available, cannot handle errors)
  - **ddrescue** or **dd_rescue** (not always available, handles errors gracefully, not suitable for pipelining)

“Clone the Data” – With Hardware

There is hardware to clone HDDs that generally works well. Usually, this means “copy a SATA drive onto another SATA drive.” If this is what is needed and you have such a device, excellent. (Make absolutely sure you put the HDDs in the **correct** slots though or you will clone an empty HDD onto your original!)

“Clone the Data” – Getting an Overview

Many times, “cloning the data” really comes down to “finding the right devices to put into the **dd** command line.” Some good sources of information:

- **blkid**
- **lsblk**
- **/dev/disk** directory tree

“Cloning the Data”: Things to Remember

- Triple-check the source and the target of the write operation!
- Make sure the target storage is large enough.
- If you clone into a file (rather than directly onto a block device), make sure the file system supports sufficiently large files.
- Verify that the copy is accurate (run checksums – on the fly, if possible).
- Setting the fresh copy read-only and/or immutable is generally a good idea.

Wrap-Up

Recap

- Many times, the circumstances will directly or indirectly dictate how to go about acquiring persistent data.
- In general, it is preferable to grab physical devices rather than their logical block devices, whole block devices rather than partitions, whole partitions rather than the files in the file system, all files in a file system rather than individual files.
- Obviously, size and/or location of the data may make options impractical or impossible.

Dangerous Pitfalls

Beware of

- incomplete copies,
- insufficient storage,
- wrong argument order, wrong mount points, wrong device names,
- automounting/auto-repairing,
- missing crypto keys.

“Plenty of Potential Places of Persistency” Images

- DAT cartridge:
  Hades2k, CC BY-SA 2.0, via Wikimedia Commons
- Flash memory cards:
  CC BY-SA 2.0, via Wikimedia Commons
- Hard drive stack:
  Ashley Pomeroy, CC BY-SA 4.0, via Wikimedia Commons
- HDDs:
  vnunet.com, CC BY-SA 2.5, via Wikimedia Commons
- RAID enclosure:
  Tophost, CC BY-SA 2.0, via Wikimedia Commons
- USB sticks:
  Usbmemorydirect.com, CC BY-SA 3.0, via Wikimedia Commons

“General Observations – Objective of the Analysis” Images

- Evidence bag:
  craig mack, CC BY-SA 3.0, via Wikimedia Commons
- Write blocker:
  ErrantX, Public domain, via Wikimedia Commons
- Safe:
  Binarysequence, CC BY-SA 4.0, via Wikimedia Commons
- Bank vault:
  Renaud d'Avout, CC BY-SA 3.0, via Wikimedia Commons

GÉANT
Networks • Services • People

Thank you

Any questions?

www.geant.org

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 731122 (GN4-2).