Transparency and Information Sharing in Digital Forensics

Johan Berggren - Google
Incident Response / Forensics

First, Who am I?

- Johan Berggren
- Incident Response / Forensics at Google
- Background from R&E networks
- Open source enthusiast

The plot for today

Imaginary incident: You need to triage and investigate 42 computers (laptops, servers, Windows, Linux, MacOSX) across 16 countries with a team of 8 investigators working in multiple timezones.. Oh, and in one of the Windows boxes we suspect that there can be evidence hiding in a VSS volume.. and we also need memory dumps.

This is a complex case. We need good tooling, effective information sharing and solid collaboration in order to solve this quickly.

Collection

- Does my tooling support this?
- Do I need a dongle in every remote location?
- Does the license cover this?
- 16 countries you say.. Maybe call in support?
- Windows, Mac and Linux..?
- What about memory?
- I really need this data as soon as possible..

Result of collection

After some (long) time, involving 20+ people I have managed to collect some artifacts from some of the Windows boxes. I also managed to get full disk images of a couple of computers (Windows, Macs and Linux). None of the laptops though, we have to wait until they come back to the office..

Processing

- My tool can extract timestamp information!
- But only for some file formats, and only for Windows..
- No support for encrypted (BitLocker) Windows disk images. No VSS support.
- I need an extra license for Mac support.
- No automation, so we have to do it by hand. This is gonna take time..
- We only have 2 licenses for 2 workstations..

Analyzing

- Ok, we got some data processed. Let’s start working!
- But we only have 2 workstations with our software. One in each timezone.. and we have 8 analysts..
- Ok, one will do analysis and one will keep track of notes. Then we rotate..
- How can we collaborate and share information/knowledge about the case within the team?

Result

- We got some data to analyze, but it took some time and effort to coordinate.
- No memory dumps
- We could only process the Windows artifacts.
- It took a long time because we had to do it by hand.
- We didn’t really utilize all analysts.
- Information sharing within the team was not great.

My ideal tooling

- The suite versus the toolbox e.g. SIFT
- Does not get in the way of the analysis!
- Cross platform support
- Supports one-off scripts and automation.
  - Shouldn't be tied to a vendor's product
  - No dongle!
- Easily adaptable and extendable.
- Support collaboration.
- Be transparent all the way.

Let’s try the toolbox approach

Same imaginary incident, different approach.

GRR Rapid Response for collection and triage

- Open source Incident Response Framework
- Fully fledged response capabilities handling most incident response and forensics tasks
- Remote Live Forensics
- Support for Linux, Mac OS X and Windows clients
- Secure communication infrastructure designed for Internet deployment (HTTP)
- Scalable back-end to handle very large deployments

Why GRR?

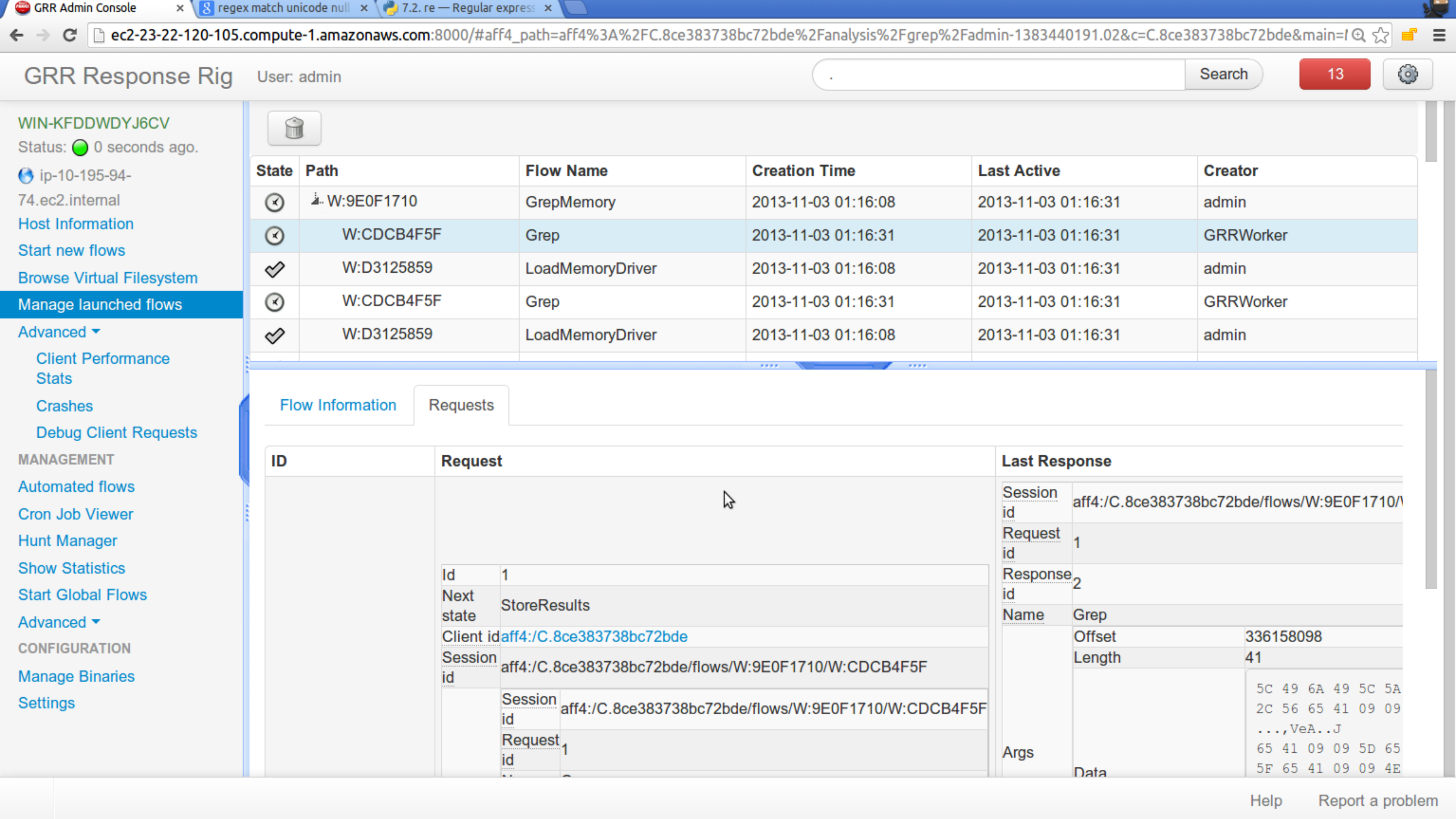
- ***Tell me if this machine is compromised***
    - (while you're at it, check 20000 of them)
- ***Joe saw something weird, check his machine***
    - (p.s. Joe is on holiday in Sweden and on 3G)
- ***Forensically acquire 42 machines for analysis***
    - (p.s. they're in 5 continents and only 2 are Windows)

GRR Flows

- To run an analysis on the client, we run flows
    - e.g. GetFile, ListDirectory, ListProcesses, GetMemory
- Requests and Responses
- State machine
- Do not take up server resources while waiting for the client
- Scales well. The individual states in the flow can be made by different machines

GRR Hunts

- Run flows on many clients
    - Or subset of the fleet, e.g. only Windows machines
- Find malicious code and abnormal behavior amongst the entire fleet of clients
- Fast triage
    - Look for Indicators of Compromise



State	Path	Flow Name	Creation Time	Last Active	Creator
◷	W:9E0F1710	GrepMemory	2013-11-03 01:16:08	2013-11-03 01:16:31	admin
◷	W:CDCB4F5F	Grep	2013-11-03 01:16:31	2013-11-03 01:16:31	GRRWorker
✓	W:D3125859	LoadMemoryDriver	2013-11-03 01:16:08	2013-11-03 01:16:31	admin
◷	W:CDCB4F5F	Grep	2013-11-03 01:16:31	2013-11-03 01:16:31	GRRWorker
✓	W:D3125859	LoadMemoryDriver	2013-11-03 01:16:08	2013-11-03 01:16:31	admin

ID	Request	Last Response
	Id: 1 Next state: StoreResults Client id: aff4:/C.8ce383738bc72bde Session id: aff4:/C.8ce383738bc72bde/flows/W:9E0F1710/W:CDCB4F5F Session id: aff4:/C.8ce383738bc72bde/flows/W:9E0F1710/W:CDCB4F5F Request id: 1	Session id: aff4:/C.8ce383738bc72bde/flows/W:9E0F1710/\ Request id: 1 Response id: 2 Name: Grep Args: Offset 336158098 Length 41 Data

```
5C 49 6A 49 5C 5A
2C 56 65 41 09 09
...,VeA..J
65 41 09 09 5D 65
5F 65 41 09 09 4E
```

Plaso for processing

- Open source timelining tool.
- Modular and flexible
- Targeted analysis or the kitchen sink approach
- Easy to automate and script

Plaso architecture

- Preprocessing
  - Collect information about the image.
    - e.g. timezone, hostname, users etc..
- Collection
  - Find all the files to process
- Extraction
  - Parse the files and store all the events
  - Community effort
- Storage & Output

Information sharing

- Different shapes and forms
  - Within team
  - Within organisation
  - Between organisations
  - Between tools
- Let our tools work for us
  - Encourage information sharing and collaboration
  - Make information sharing part of the design

Timesketch

- Open source collaborative forensic timeline analysis
- Web based tool to analyse timeline data
- Modelled around collaboration and information sharing
    - Users can work simultaneously on the same data
    - Annotate
    - Share findings

Timesketch architecture

- WebUI
  - Focuses on collaboration
  - You share information while you are analyzing
- HTTP RESTful API
  - Add authn and authz
- Backend storage and search
  - Fast
  - Search across indexes

Sketch

Screenshot..

Multiple timelines

Screenshot..

Annotations

Screenshot..

Share views

Screenshot..

Result

- We were able to quickly triage.
- We collected the data we needed fast.
- We processed all the data.
- Most of the collection and processing was automated.
- All analysts worked in parallel and shared their findings with timesketch.

Information sharing, moving forward

- Even more central in the tools design
- Stories
    - Mix data with narrative
    - Let the data explain the story
    - Build context around events
- Knowledge sharing
    - forensicwiki.org
    - Artifacts to glue tools together

Artifacts

- Artifacts (examples)
  - Windows Application Event Log
  - A (Windows Registry) Run Key
  - A process
  - A mutex
  - Browser history
- Artifact definitions
  - Share artifact knowledge with the community
  - Integrate with tools
  - Data driven

Artifacts in our toolbox

- Collection based on artifacts (e.g. GRR)
- Extraction and processing with artifacts (e.g. Plaso)
- Overlay your data with artifact descriptions to aid in analysis (e.g. Timesketch)

What about transparency?

- Open source
  - Verify the result from our tools
  - Understand why the data is presented to you
  - Add transparency to the process
  - **Keep your team motivated**
    - Developing open source software can be a motivator!
    - “Free” education.

Conclusion

- Incident Response at scale is hard.
- Relying on a single monolithic product can sometimes be a limiting factor.
- Open source forensics have come a long way.
- Open source drives motivation and innovation.
- Open source adds transparency.
- Collaboration and information sharing should be part of the tools design.

Questions?

References

swiss army knife (Creative Commons)

http://en.wikipedia.org/wiki/File:Wenger_EvoGrip_S17.JPG

Plaso logo (Used with permission)

https://lh6.googleusercontent.com/Imix4Wnn8v__wXcv4vXdXwzOzlFuiV6i5uVvUm2_8F6FMY7Qjze-qcHLiugFjwsOdNn9s5aVrk94diS2kRumQPPPZZHLzNq1VdSk8vSuoHrqPwCot1RoifA6UMU