Enterprise-Scale Digital Forensics With Autopsy

Richard Cordevane
Brian Carrier
Motivation

Same story each year:

○ Cases are getting bigger
○ Devices are getting bigger
○ Labs are not getting bigger at same rate
○ Examiners leaving for private sector

Problem: Your lab needs to be able to scale and get consistent results even with high turn over.
Agenda

We’re going to focus on three problems (in 30 min):

- How to scale
- How to get consistency
- How to transfer knowledge
Problem #1: Scaling Large Cases
Typical Scenario

Large case comes in with many devices.
2 or more examiners are assigned to the case.
Each is assigned a device (or data source)
They analyze it on their desktop computer
Somehow communicate about their findings.
Somehow merge reports at the end
It Works, But….

This approach is not effective or time efficient.

Each person is working in isolation:

- Knowing what is on other devices helps to provide context for the current device.
- Examiners can’t see results from their colleagues.

Time Efficient:

- Need to repeatedly merge so that everyone knows what has been found.
- Merging results is tedious and/or manual.
A Collaborative Environment is Better

In a collaborative system...

○ Everyone can see all of the results in real time
○ No merging of results required
○ Single, unified report generated at any time

Collaborative systems exist but often they cost a lot of money...

Autopsy is a Collaborative Environment.
What Do You Need?

Hardware for 2 servers
Shared storage
Download Autopsy
  - http://www.sleuthkit.org/autopsy/
Download other open source packages:
  - PostgreSQL (Central database)
  - Apache SOLR (Indexed keyword search)
  - ActiveMQ (Messaging server)
Architecture
Create a Multi-User Case

Enter New Case Information:

Case Name: myCase
Base Directory: \myServer\cases\output

Case data will be stored in the following directory:
\myServer\cases\output\myCase

Single-user  Multi-user
That's It. No other notable changes.
Examiners Have Visibility

What are they analyzing:

Their tags:
Generate Unified Report

Select and Configure Report Modules

- HTML Report
- Excel Report
- Add Tagged Hashes
- Files - Text
- Google Earth KML
- STIX
- TSK Body File

A report about results and tagged items in HTML format.

This report will be configured on the next screen.
Group By Data Source

Unified cases can make it harder to focus on a single data source.

Autopsy allows you to group and filter by data source.
Automated Ingest: Motivation

When big cases come in, a surge in processing is needed.

You’ll want to know which devices to start with.

You don’t have someone watching at 2AM to start processing a new image.

You don’t want to waste examiner time waiting for processing to complete.
Automated Ingest: Solution

Autopsy can have “Auto Ingest” nodes that constantly scan folders for new data sources.

The data sources are analyzed using a preconfigured setup (hash sets, keywords, etc.)

Analysis is done 24x7.

Dashboard allows for prioritization and to review progress.
**Pending Jobs**

<table>
<thead>
<tr>
<th>Case</th>
<th>Data Source</th>
<th>Job Created</th>
</tr>
</thead>
<tbody>
<tr>
<td>epsilon</td>
<td>kw_r96.img</td>
<td>2016/08/26 15:04:35</td>
</tr>
<tr>
<td>alpha</td>
<td>mtd1_system.bin</td>
<td>2016/08/26 15:04:35</td>
</tr>
<tr>
<td>alpha</td>
<td>mtd3_userdata.bin</td>
<td>2016/08/26 15:04:35</td>
</tr>
<tr>
<td>theta</td>
<td>thunderbird_smallmage.dd</td>
<td>2016/08/26 15:04:35</td>
</tr>
</tbody>
</table>

**Running Jobs**

<table>
<thead>
<tr>
<th>Case</th>
<th>Data Source</th>
<th>Host Name</th>
<th>Stage</th>
<th>Time in Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma</td>
<td>dump.bin</td>
<td>win-4913</td>
<td>Opening case</td>
<td>9s</td>
</tr>
</tbody>
</table>

**Completed Jobs**

<table>
<thead>
<tr>
<th>Case</th>
<th>Data Source</th>
<th>Job Created</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi</td>
<td>small.img</td>
<td>2016/08/26 15:04:35</td>
<td><img src="https://example.com/status_icon.png" alt="Status" /></td>
</tr>
<tr>
<td>alpha</td>
<td>mtd1_cache.bin</td>
<td>2016/08/26 15:04:35</td>
<td><img src="https://example.com/status_icon.png" alt="Status" /></td>
</tr>
<tr>
<td>beta</td>
<td>green_images.img</td>
<td>2016/08/26 15:04:35</td>
<td><img src="https://example.com/status_icon.png" alt="Status" /></td>
</tr>
<tr>
<td>beta</td>
<td>blue_images.img</td>
<td>2016/08/26 15:04:35</td>
<td><img src="https://example.com/status_icon.png" alt="Status" /></td>
</tr>
<tr>
<td>alpha</td>
<td>b00_mmb00.bin</td>
<td>2016/08/26 15:04:35</td>
<td><img src="https://example.com/status_icon.png" alt="Status" /></td>
</tr>
</tbody>
</table>
Case Review With Auto Ingest

Add data sources to folders to be analyzed.

Examiners open cases as they are being analyzed or as they complete to prioritize.

Data is being analyzed ASAP because the auto ingest nodes don’t stop until the data is done.
Problem: Consistency & Speed
Consistency is Critical

Many labs struggle with:

- Having consistent hash sets across all desktops.
- Making sure evidence from past cases is flagged in future cases.
- Ensuring all examiners are looking at all of the same places.
Central Repository

The Central Repository is an optional database that stores data across cases.

- Each case also gets its own database

The Central Repository stores:

- References to where each file/MD5 was seen
- Common configuration data

Can be used for single-user cases.
Central Hash Sets

Problem: Tedious to copy around the latest version of the NIST NSRL or notable hash sets

Solution: Store the hash sets in the Central Repository and each node queries it.

The Autopsy hash lookup module knows about local and remote hash sets and uses them interchangeably.
Flag Previously Notable Files

When an examiner tags a file as “Notable”, that can be stored in the Central Repository.

When a new case contains that same file, it will get flagged as being “Previously Notable”.

Benefits:

- Easier than maintaining hash sets
- Helps to triage / prioritize
Flagging Previously Notable Files
Automating Your Checklist

There are dozens of places to look for possible evidence, that do not often exist:

- Various cloud storage tools
- Phone backups
- Virtual machine containers
- ...

Can be easy to forget to look for one of them
Interesting Files Module

Allows you to enter file name patterns to look for.

- truecrypt.exe
- DropBox folder
- Google\Drive folder

Ensures that the examiner is always alerted to their presence.
Problem: Turnover and Knowledge Loss
Examiners build up a lot of knowledge as they do cases. They learn about what apps do and what files are for. When they leave, that knowledge leaves the lab.
Knowledge Transfer Solution

Store comments about files in the Central Repository. Examiners can comment about what a file is for and what an app does. Future examiners will see that and not have to research them again.
Adding A Comment

Properties
View File in Directory
View in New Window
Open in External Viewer
View File in Timeline...
Extract File(s)
Add File Tag
Remove File Tag
Add/Edit Central Repository Comment
Add File to Hash Set
Seeing a Comment

Central Repository Comments

Case: demo-111222112
Type: Files
Comment: This file is really bad...
Path: /documents and settings/john/local settings/temporary internet files/content.ie5/om25xf75/detection_as3[1].swf
Conclusion

Multi-user cluster allows you to process data more quickly and collaborate more easily.

Central Repository allows you to store historical data and have consistent results.

Try it tomorrow!
brianc <at> basistech.com

Connect on LinkedIn