
# Rapid Forensic Imaging of Large Disks with Sifting Collectors

*By*

**Jonathan Grier and Golden Richard**

*Presented At*

The Digital Forensic Research Conference

**DFRWS 2015 USA** Philadelphia, PA (Aug 9th - 13th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

**http:/dfrws.org**

# Rapid Forensic Imaging

## with Sifting Collectors

***or, "How we can image 99% of the evidence in 8% of the time."***

Jonathan Grier, Grier Forensics
Dr. Golden G. Richard III, University of New Orleans

GRIER FORENSICS

# Legal Disclaimer

*Some technologies presented here may be patent pending*

# The Volume Challenge

| Quote | Source |
|---|---|
| *"**Today's Golden Age of computer forensics is quickly coming to an end**… The growing size of storage devices means that there is frequently **insufficient time to create a forensic image**… investigations are becoming slower… Today a 2TB hard drive can be purchased for $120 but takes more than 7 hours to image"* | Forensics researcher Dr. Simson L. Garfinkel of the Naval Postgraduate School (Garfinkel 2010) |
| *"The **leading challenge for digital forensic investigations is that of scale**… the mere acquisition, extraction and pre-processing of the data sources create a long list of technical problems and… long turnaround times."* | Forensics examiner Andrew Case (Case et al. 2008) |
| *"It took **two days to consolidate the various target media onto a pair of 2TB drives** and thirteen hours to clone and hash each drive using the very latest drive-to-drive tools."* | Forensics attorney Craig Ball, *The End of Digital Forensics?* (Ball 2011) |
| *"By far we think **the biggest challenge is the sheer volume of data**. The existing acquisition methodologies and software tools will not be able to scale as quickly as the datasets which are being acquired. Unfortunately, there are no easy solutions to these problems…. new paradigms for… these huge datasets will be… heroes in the forensics community…"* | Robert Botchek, founder and former president of Tableau (Botchek 2008) |
| *"**Police managers must find a way to examine an increasing number of digital devices**, each containing an immense volume of data… **There is an unacceptable backlog…waiting for examination**."* | First Sergeant Charles L. Cohen of the Indiana State Police (Cohen 2007) |

GRIER FORENSICS

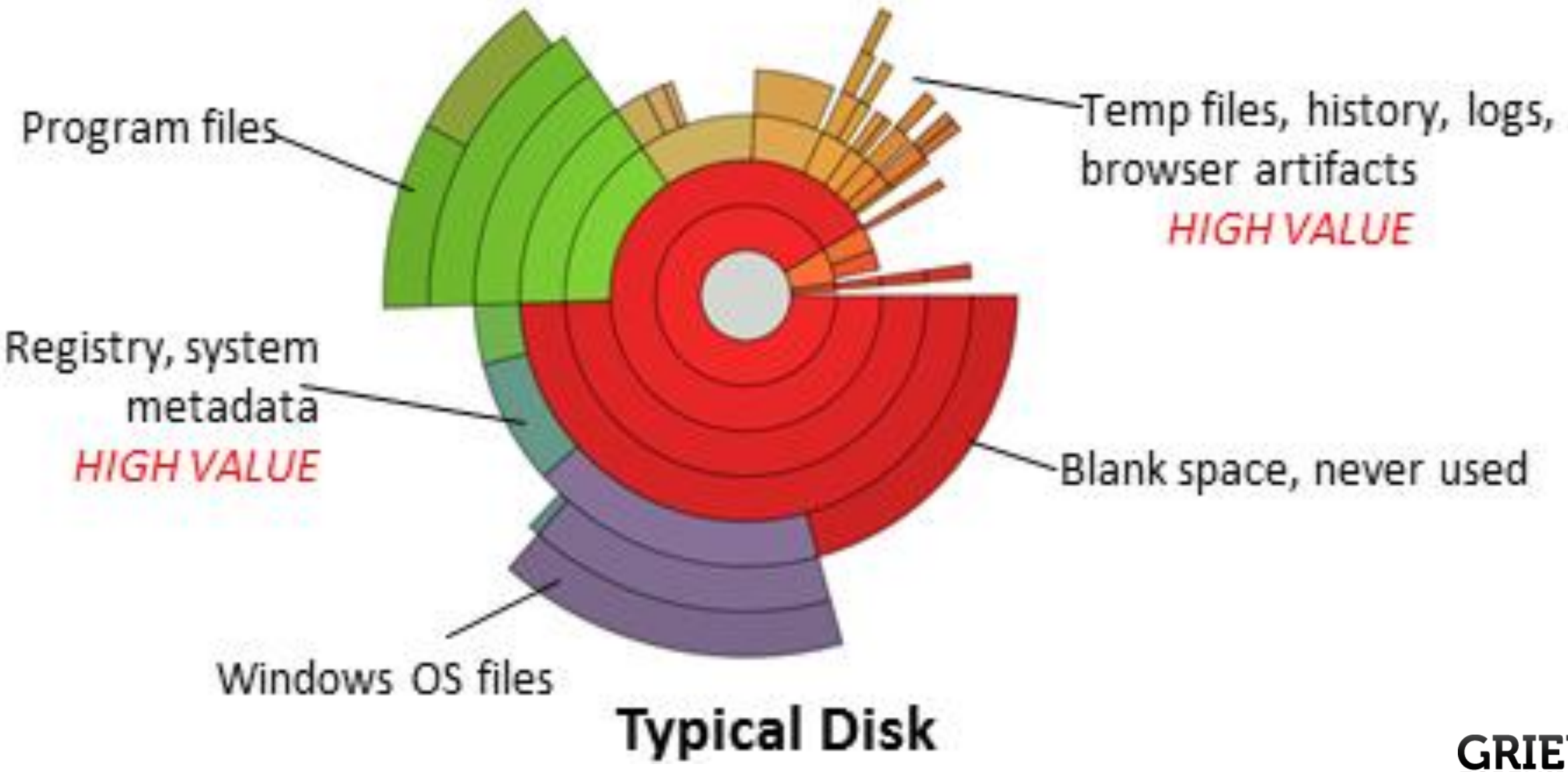
# Existing options

| | Full Disk Imaging | Live Forensics & Triage |
|---|---|---|
| Which evidence is preserved: | Entire disk | Reports, results, and files the examiner chooses to copy |
| Adequate speed | ✘ *Up to ten hours per disk* | ✓ |
| Reproducible | ✓ | ✘ |
| Verifiable | ✓ | ✘ |
| Duplicates evidence automatically | ✓ | ✘ |
| Duplicates at device level | ✓ | ✘ |
| Duplicates entire disk | ✓ | ✘ |
| Alternate lines of investigation are possible in the future | ✓ | ✘ |

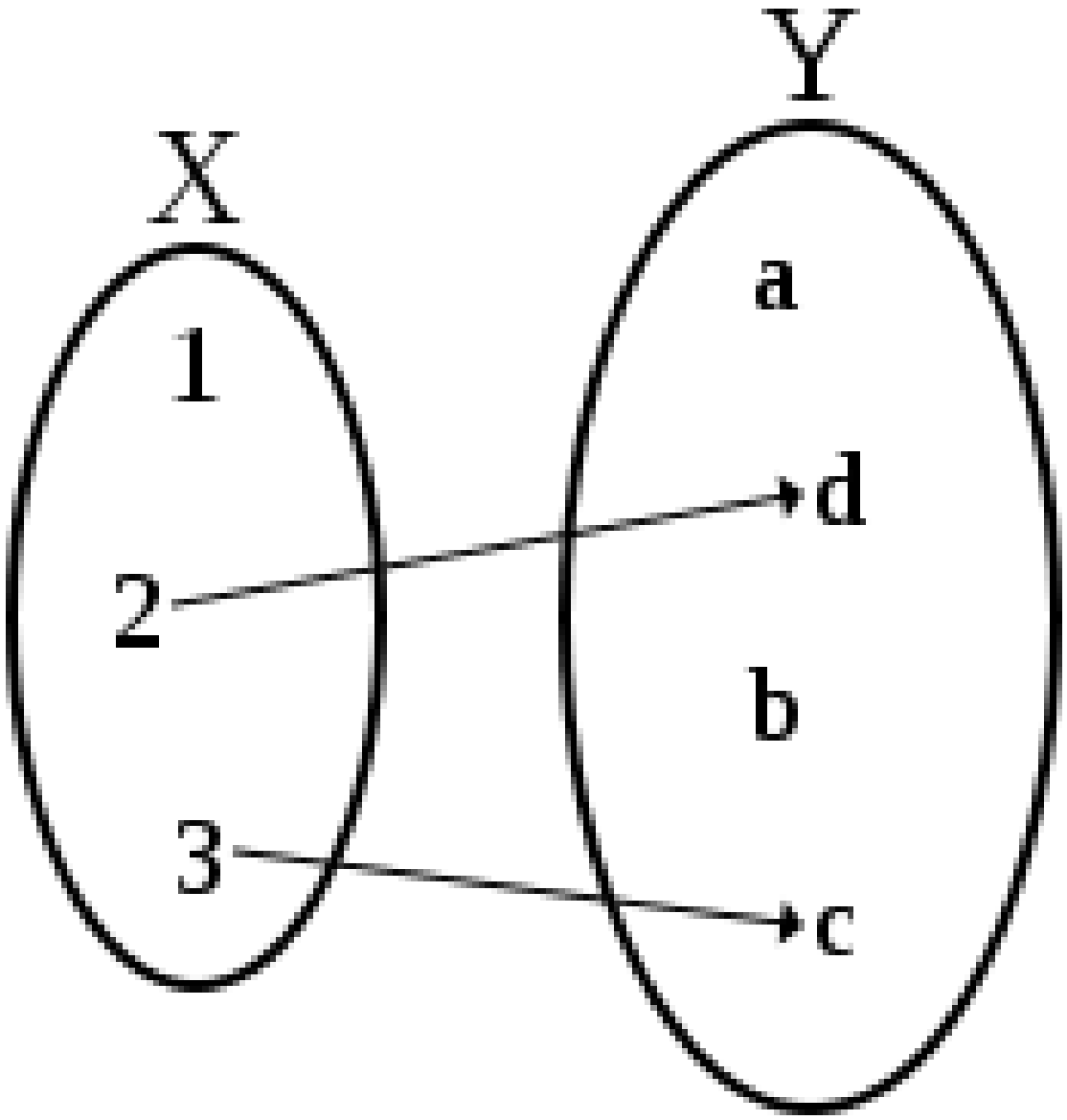
GRIER FORENSICS

# Are there other options??

# Are there other options??



# Are there other options??



# Are there other options??

**Goals:**

- Automatically **locate and collect** the forensically relevant sectors
- Produce a forensically sound, **verifiable, device level, bit-for-bit duplicate** of these sectors
- Be analyzed **using existing forensic tools** (with minor adaptations)
- At **5-10x the speed** and 10-20% the size of traditional imaging.

# Are there other options??

| | Full Disk Imaging | Live Forensics & Triage | Sifting |
|---|---|---|---|
| Which evidence is preserved: | Entire disk | Reports, results, and files the examiner chooses to copy | Identified sectors & regions |
| Adequate speed | ✘ *Up to ten hours per disk* | ✓ | ✓ |
| Reproducible | ✓ | ✘ | ✓ |
| Verifiable | ✓ | ✘ | ✓ |
| Duplicates evidence automatically | ✓ | ✘ | ✓ |
| Duplicates at device level | ✓ | ✘ | ✓ |
| Duplicates entire disk | ✓ | ✘ | ✘ |
| Alternate lines of investigation are possible in the future | ✓ | ✘ | ? |

GRIER FORENSICS

# How do we get there?

1. Identify “relevant” sectors

2. Produce a **device level**,
   sector-by-sector, bit-by-bit,
   **verifiable image**,
   compatible with existing tools

# Finding relevant sectors

**Definition 1.** A region is *forensically relevant* if and only if the conclusions of an associated investigation are substantially altered if the region's contents are replaced with random values.

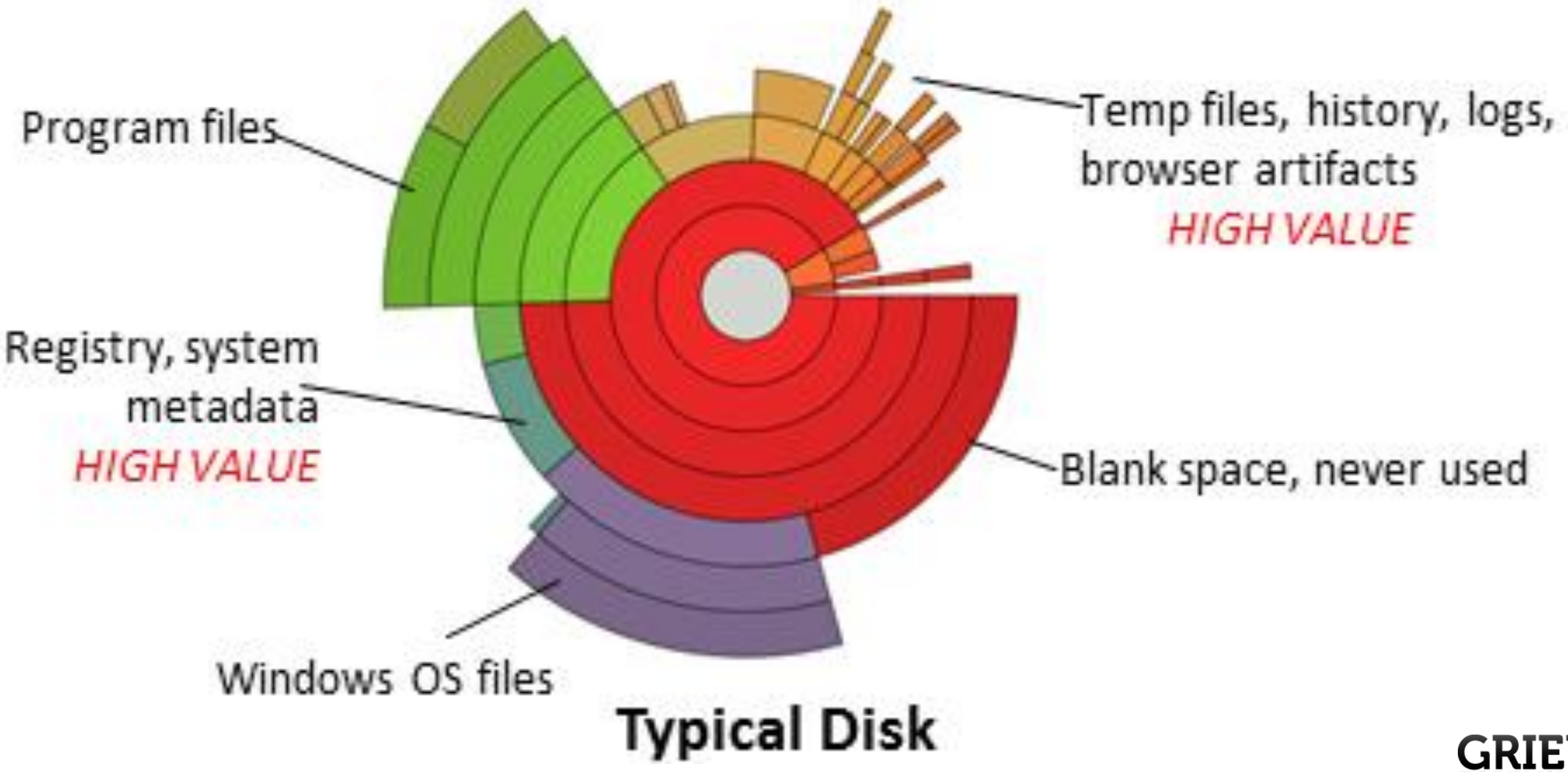
Relevance is therefore not a property of a region alone, but a *region in the context of a forensic investigation.*

1. Perform a **forensic investigation** in a **fully reproducible manner**
    That is, the procedure must be an **effective procedure** (using Turing's classic definition: a fully defined procedure, which results in a quantifiable, definite conclusion).

2. **Randomize** a particular region

3. **Repeat** the investigation

The **region is relevant**, in the context of this investigation, if and only if, **the conclusions differ**.

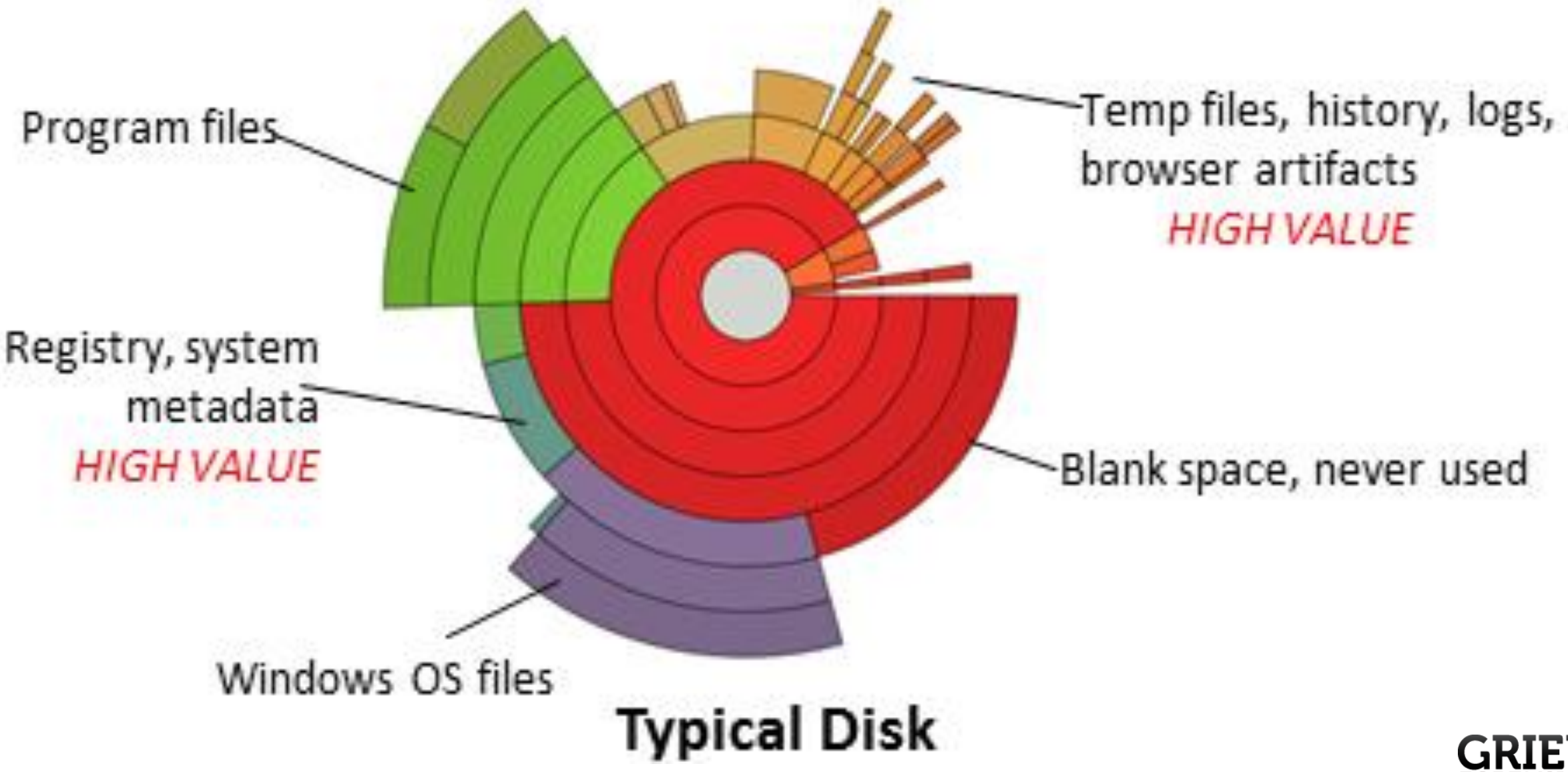



**Typical Disk**

GRIER FORENSICS

# ***Expected relevance*:**

The a priori **probability measure**
(between 0 and 1)
that an ***unread*** **region**
known to have certain **properties**
is **relevant**,
in the context of a particular investigation.




**Typical Disk**

GRIER FORENSICS

# How do we get there?

1. Identify “relevant” sectors

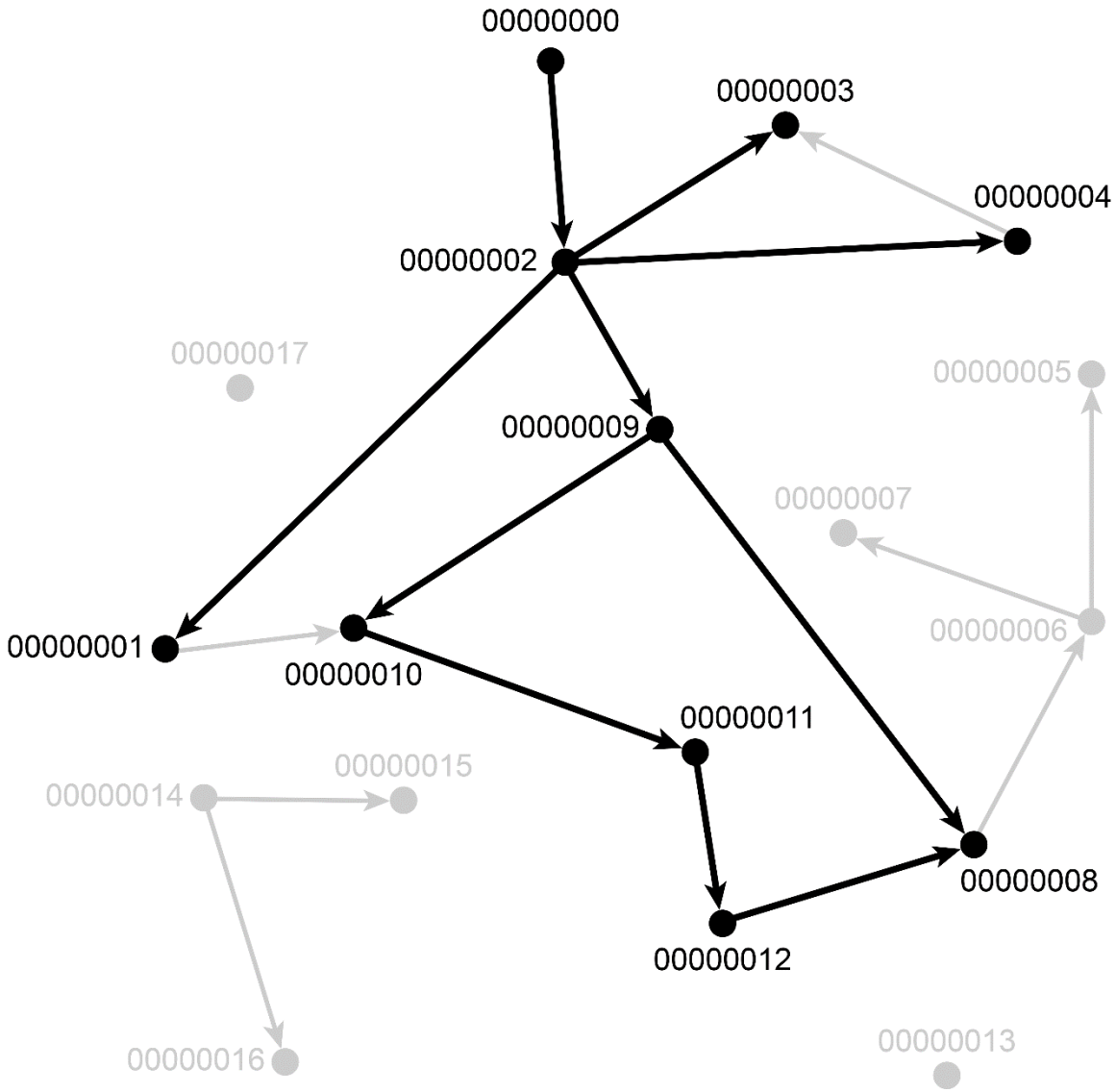
2. Produce a **device level**,
   sector-by-sector, bit-by-bit,
   **verifiable image**,
   compatible with existing tools

# How do we find them?

# How do *examiners* find them?

# Focusing procedure

# Focusing procedure




GRIER FORENSICS

# Focusing procedure

compressible

vs.

incompressible

# Compressible focusing procedure

1. Definable in advance (not requiring spontaneous judgment)

# Compressible focusing procedure

1. Definable in advance (not requiring spontaneous judgment)

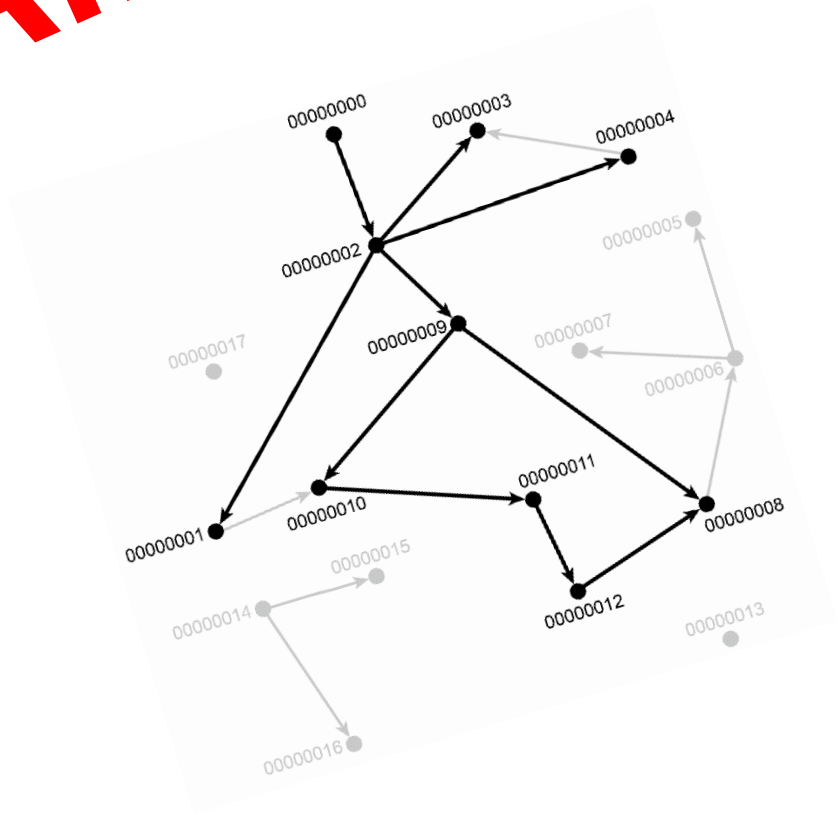
2. Metadata, not exhaustive search

| Focusing procedure | Example | Can it be defined in advance? | Can it be performed with minimal disk reads? |
|---|---|---|---|
| Known artifact | Examine known artifacts, such as Firefox's history, stored in known folders | Y | Y |
| MAC timestamps | Examine only files created within a certain timeframe | Y | Y |
| Anomalous file name or location | Examine files in the `WINDOWS` directory that are anomalous or misspelled | Y | Y |
| Autoruns | Examine files that run automatically on boot | Y | Y |
| Hash based file search | Examine files with hashes corresponding to known malware or contraband | Y | N |
| Keyword based sector search | Search disk for keywords of interest | N | N |

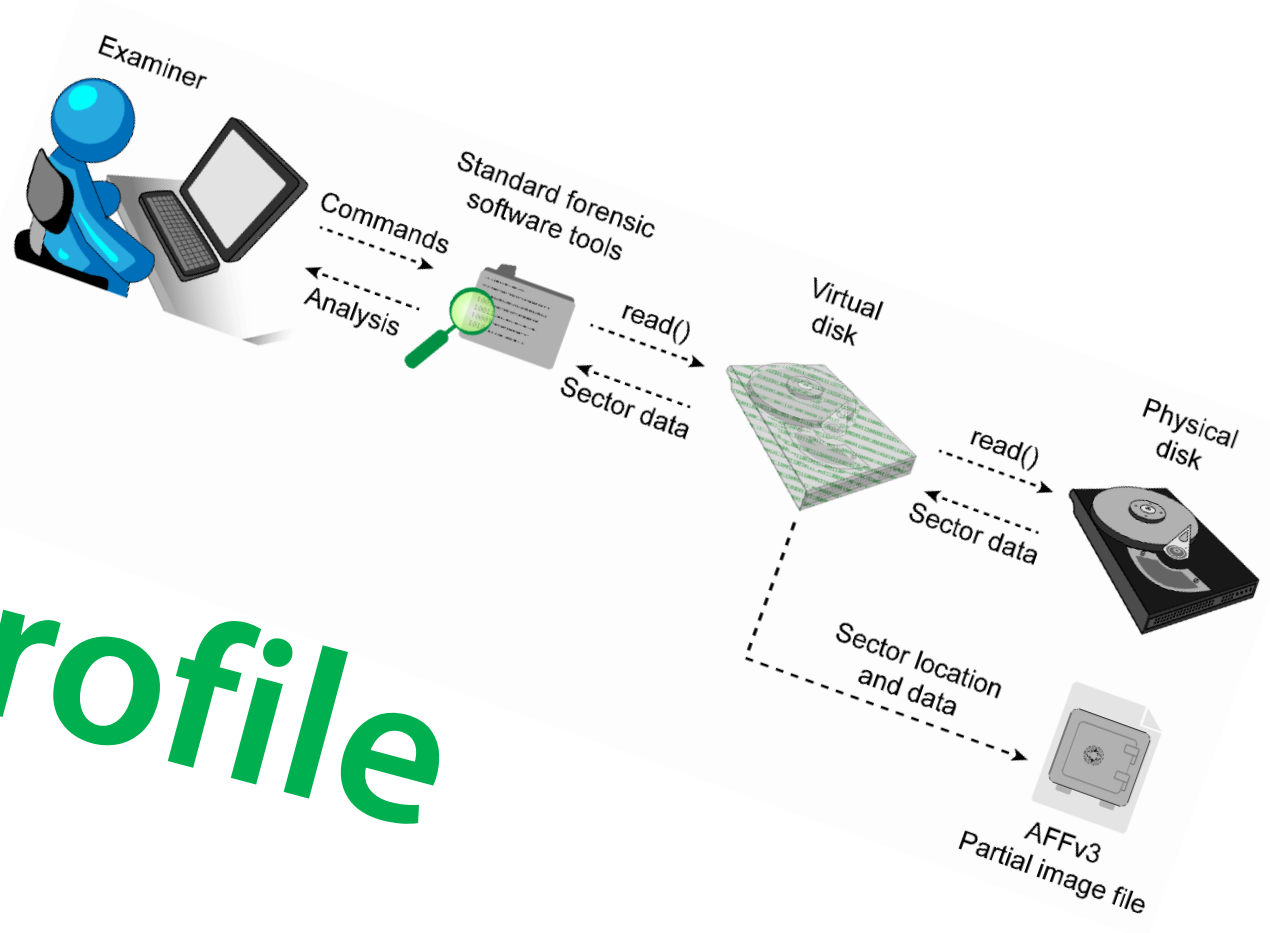
# approximating superset

# locality

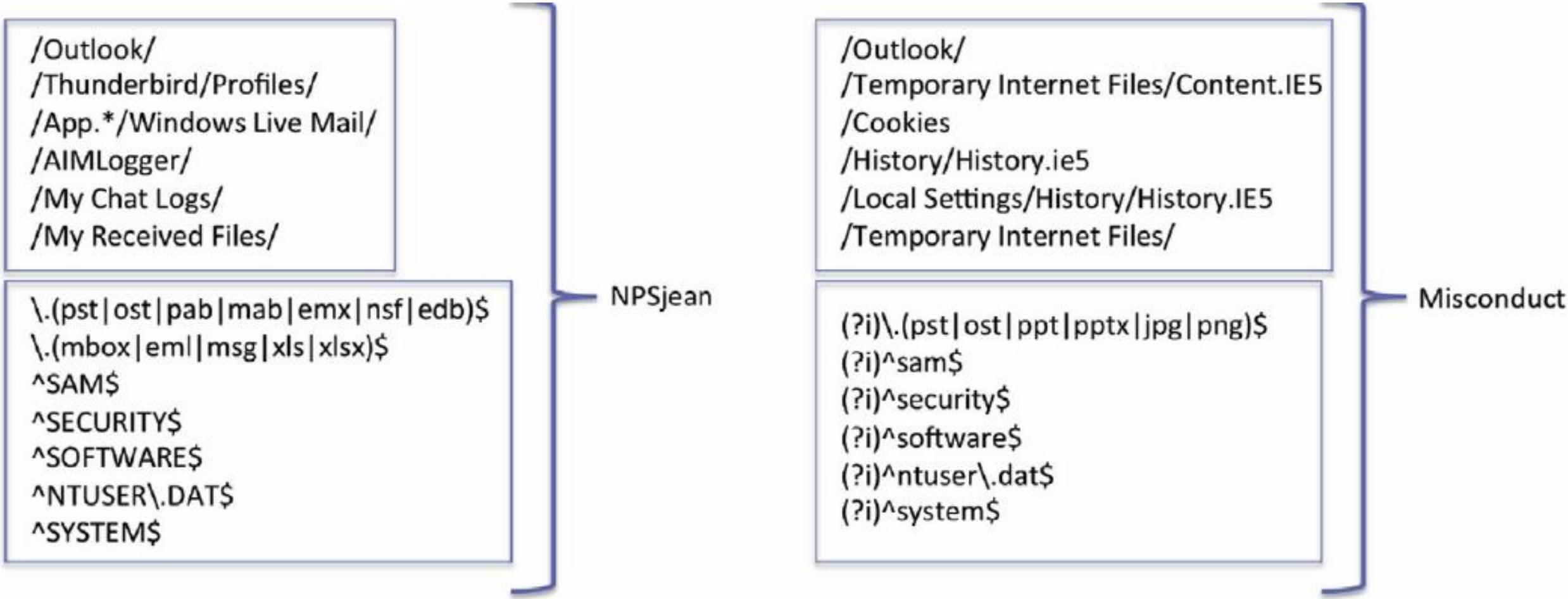
# Allocation




# Profile




GRIER FORENSICS



# approximating superset

# testing

# results

*Figure 1: Conventional image – SAM file is present*

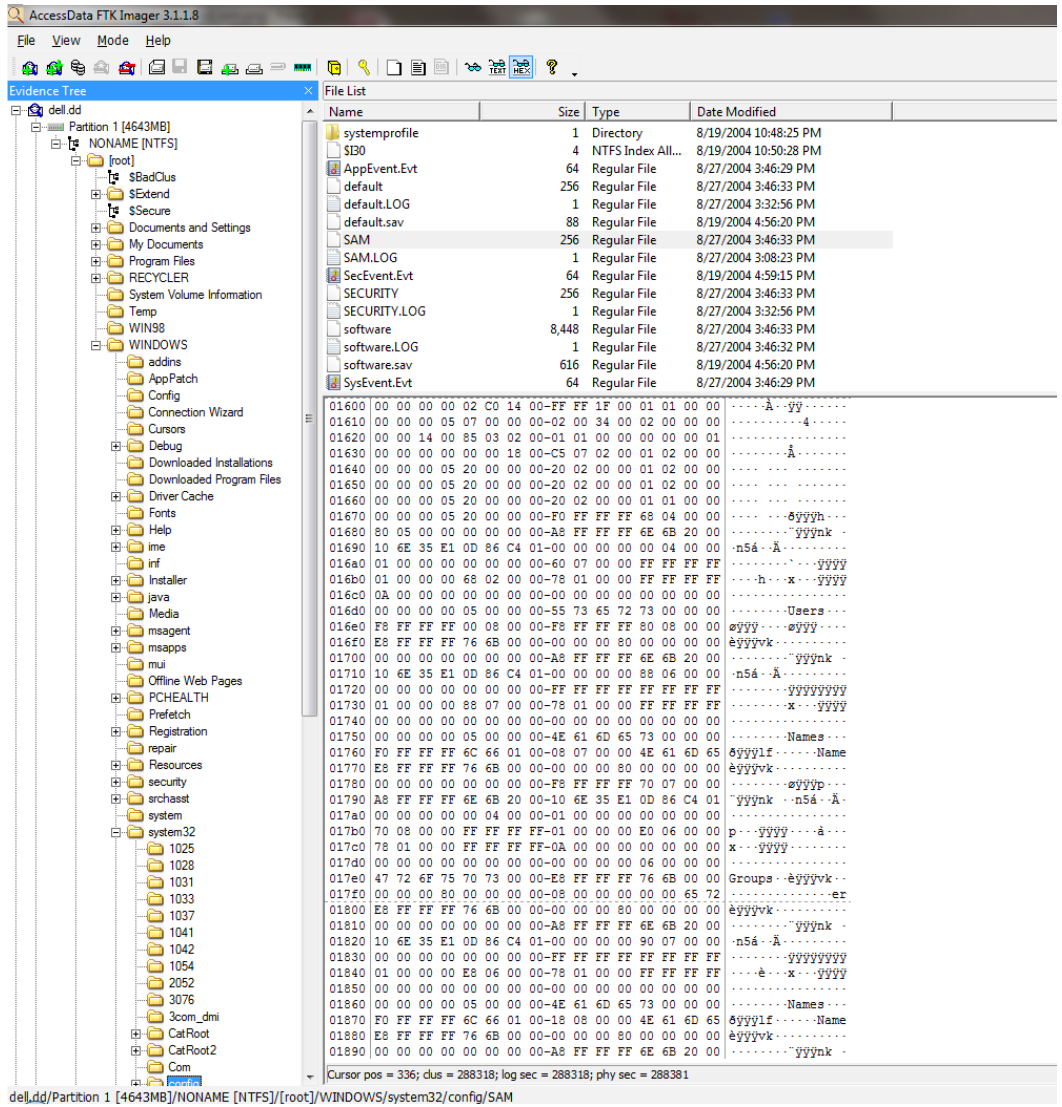
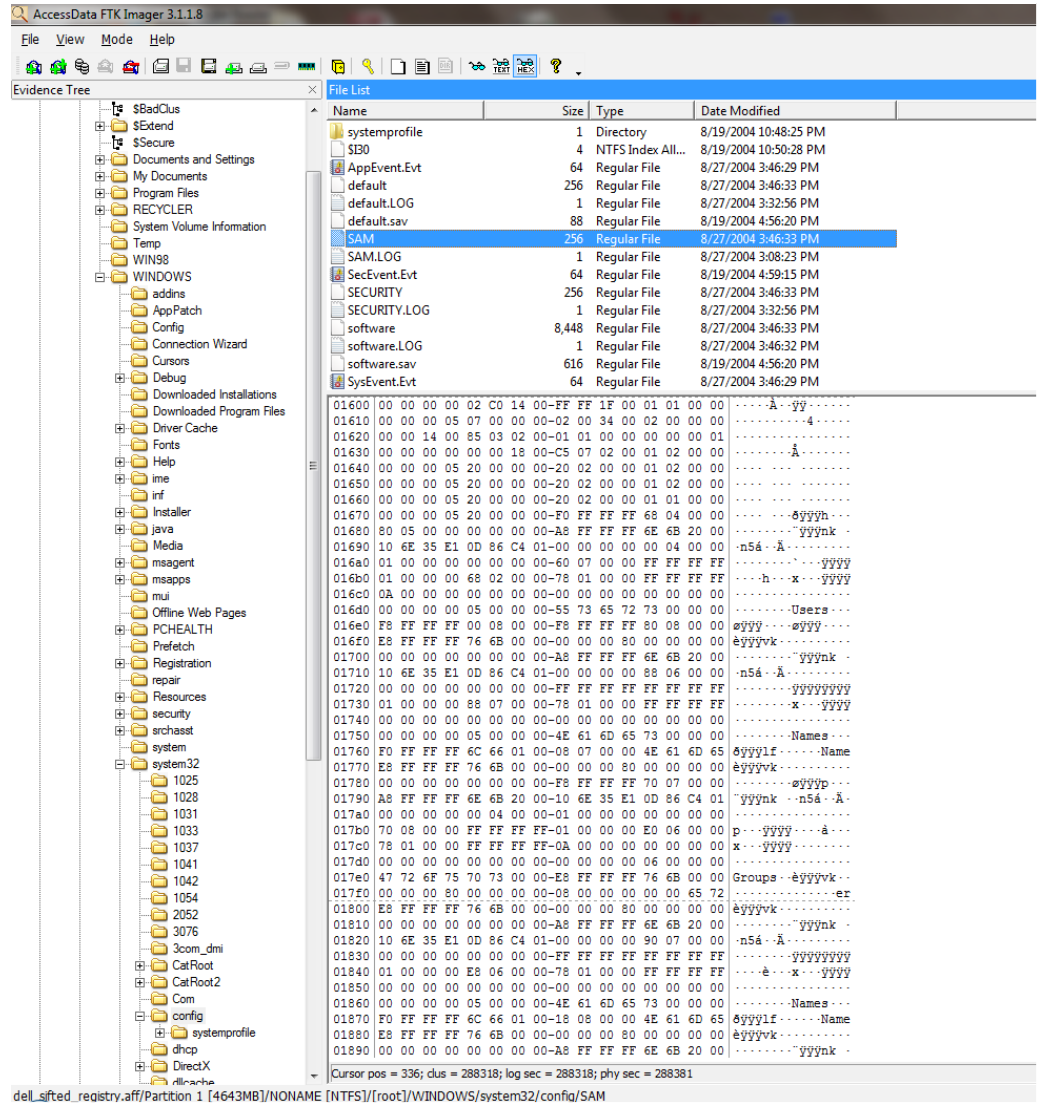



*Figure 2: Sifted Image – SAM file is fully extent. Data matches conventional image.*




GRIER FORENSICS

*Figure 5: Results for conventional image.*

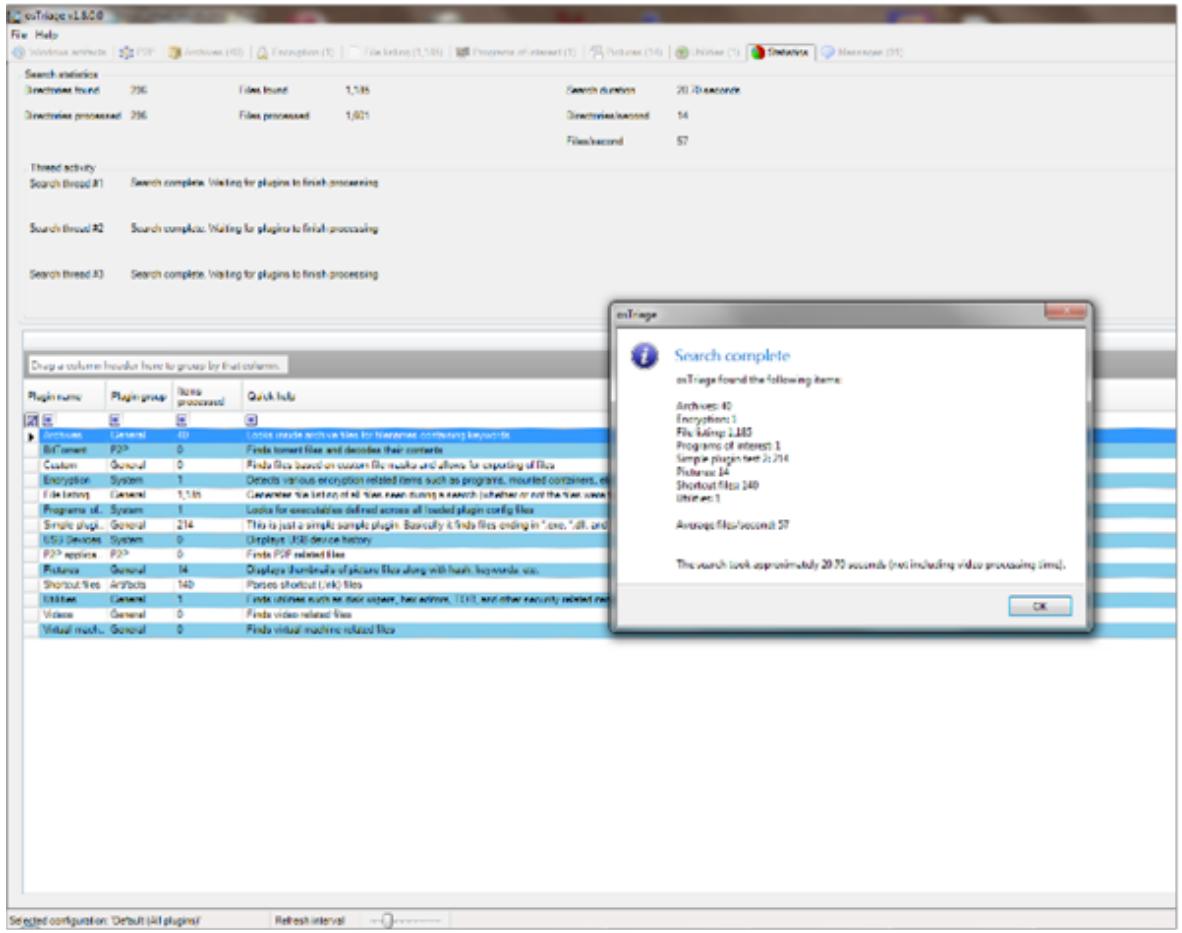
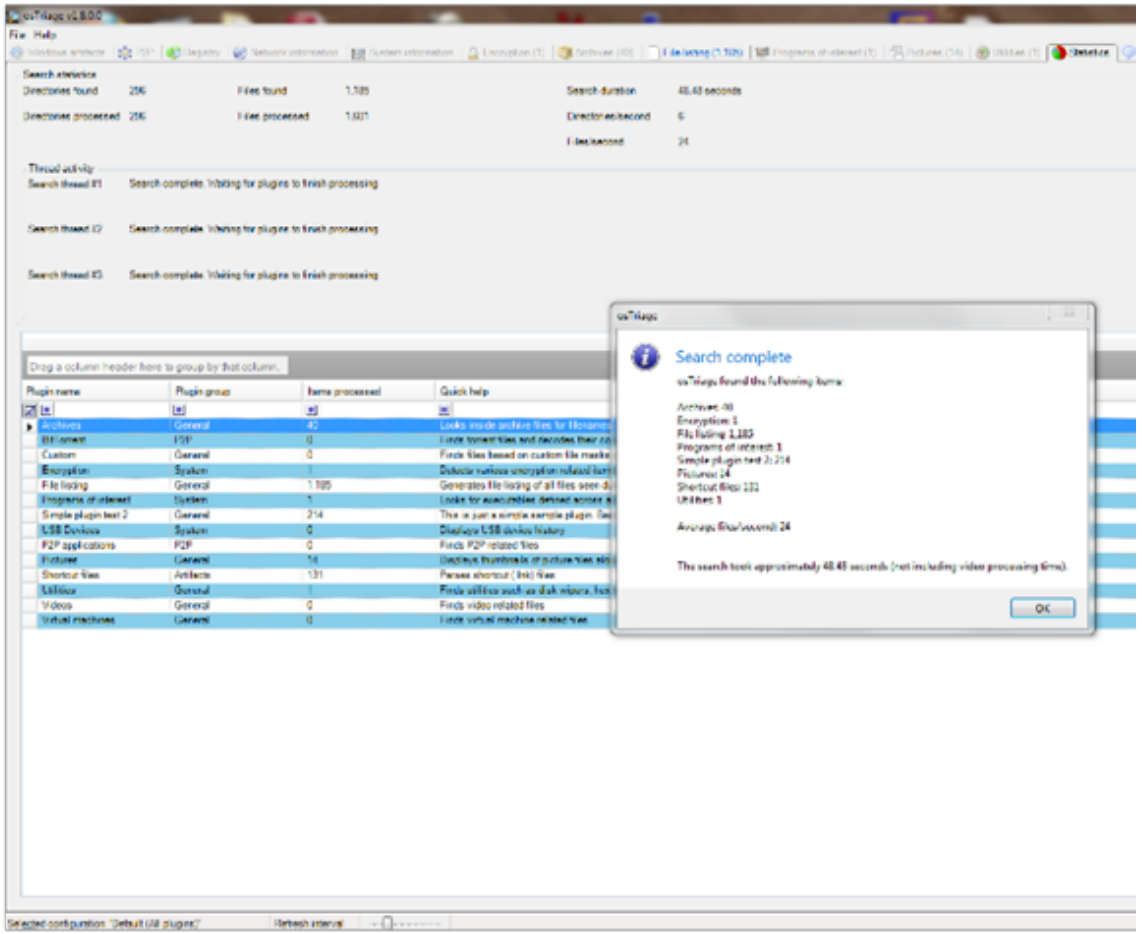



*Figure 6: Results for sifted image. Nearly all of the evidence obtained by osTriage from the conventional image is obtained from the sifted image as well.*




GRIER FORENSICS

# quantitative testing

| Disk | Profile | Acceleration | Tool | Accuracy | Comprehensiveness |
|---|---|---|---|---|---|
| NIST CFReDS Hacking Case (4.6 GB) | Registry | 4.5x | log2timeline | 100% | 100% |
| | Registry | 4.5x | Regripper | 100% | 100% |
| | IEHistory | 5.0x | Pasco | 100% | 100% |
| | Email | 3.7x | Bulk_extractor | 100% | 95% |
| | Registry | 4.5x | Mactime | 100% | 100% |
| NPS DOMEXUSERS (40 GB) | Registry | 13x | log2timeline | 100% | 54% |
| | Registry | 13x | Regripper | 100% | 100% |
| | IEHistory | 13x | Pasco | 100% | 100% |
| | Email | 11x | Bulk_extractor | 100% | 57% |
| | Registry | 13x | Mactime | 100% | 100% |

# investigative testing

GRIER FORENSICS

synthetic case:

3.2x acceleration

100% accuracy

100% comprehensiveness

GRIER FORENSICS

real case, disk 1:

2.9x acceleration

100% accuracy

99% comprehensiveness

# real case, disk 2:

# error

real case, disk 3:

9.6x acceleration

100% accuracy

100% comprehensiveness

# in the real world

GRIER FORENSICS

# Multiple Applications

fast imaging:
bypass blank regions

reproducible triage:
examiner driven

comprehensiveness

acceleration

new niche:
profile driven collection
(?)

GRIER FORENSICS

We're hiring.
If this is the type of project you'd like to work on, please contact me directly.

# Questions?
# Want to pilot?

Jonathan Grier
jgrier@grierforensics.com

GRIER FORENSICS