NSRL Next Generation – Diskprinting

Mary T. Laamanen¹, Alex J. Nelson²,³

¹ NIST
² Prometheus Computing
³ University of California, Santa Cruz

December 3, 2014
Disclaimer

• This talk mentions several software products.

• No mentions are or should be construed as endorsements of that software.

• In this research, they are test subjects.
Diskprints show when artifacts appear.

- Baseline
- Installation
- Running
- Uninstallation
- Rebooting
Diskprints show when artifacts appear.

- Baseline
- Installation
- Running
- Uninstallation
- Rebooting
Diskprints show when artifacts appear.

- Baseline
- Installation
- Running
- Uninstallation
- Rebooting

Registry entries: 460,000
Diskprints show when artifacts appear.

- Baseline
- Installation
- Running
- Uninstallation
- Rebooting

Registry entries: 460,000

+ 10 – 10,000
Diskprints show when artifacts appear.

- Baseline
- Installation
- Running
- Uninstallation
- Rebooting

Registry entries: 460,000

+ 10 – 10,000

+ more
Diskprints show when artifacts appear.

- Baseline
  - Registry entries: 460,000

- Installation
  - + 10 – 10,000

- Running
  - + more

- Uninstallation
  - + less & more

- Rebooting
We need to understand artifact origins.

• Files, Registry cells – mostly unknown origins.
  – Most created by software.
  – Some recognized from malware signatures.
  – Most just *in the way of finding relevant data.*
Diskprints help recognize *artifacts* and *behaviors*.

- Whole virtual machine states are available.

- We compute changes between states, making:
  - Catalogues of system behavior
  - Known-file lists
  - Software signatures
Diskprint data are being made from *forensic differencing*.

- New NSRL data sets based on diskprint sequences.
  - Using forensic differential analysis [Garfinkel *et al.*, DFRWS 2012]
  - Extension: *Forensic sequence analysis*
Outline: Data set production

• File system analysis language
• Diskprint lineage analysis workflow
• Results (with URL)
• Research on software signatures
• Conclusions
File system analysis language

Digital Forensics XML
File system analysis with DFXML

- Digital Forensics XML describes storage system metadata.
  - Currently hosted by NIST.
    - Originally by Garfinkel [SADFE, 2009; DI, 2012].
  - Document language (with XML schema).
  - Python bindings available.
  - In use by forensic researchers, digital archivists.
DFXML describes storage, and changes.

- Simple annotations for files.
  - New, removed, modified.

- New analytics on *reduced data*.
  - *E.g.* timeline of changes, instead of whole system.
The structure of diskprint data

Lineage graph
The diskprint lineage graph

A machine’s state is related to its ancestors.
The diskprint lineage graph

A machine’s state is related to its ancestors.

The history can fork.

The tree is rooted at the baseline OS.
The diskprint lineage graph

A machine’s state is related to its ancestors.

The history can fork.

The tree is rooted at the baseline OS.
The diskprint lineage graph

A machine’s state is related to its ancestors.

The history can fork.

The tree is rooted at the baseline OS.

The lineage graph is all of the trees.
The diskprint analysis workflow

*Lineage-based differencing*
The diskprint analysis workflow

Some results can be derived from a single snapshot.
The diskprint analysis workflow

Some results can be derived from a single snapshot.
The diskprint analysis workflow

Some results come from two snapshots.
Results

*New-content data sets*
Now available: File system difference data

- File system changes available in:
  - Differential DFXML
  - NSRL RDS format (CSV)
  - CybOX

- Sector hashes of new and modified files

- [http://www.nsrl.nist.gov/dskprt/sequence.html](http://www.nsrl.nist.gov/dskprt/sequence.html)
Research

Registry-based software signatures
Developing software signatures

• What artifacts are distinct to an application?
  – Or, have sufficient affinity?

• Can the Windows Registry show the software history of a computer?
  – A boon to triage.
Methodology: “Document” search

1. Observe the sets of Registry artifacts created by a snapshot.

2. Assemble those sets into “Fingerprint documents”

3. Query with a Registry.
Signature challenges

• Some indistinct artifacts confuse signatures.
  – Need “Background noise” identification.

• (See me at poster session for more.)
Summary

Data in use,
research on horizon.
Community

- Forensic standards
  - MITRE
- Archival applications of Digital Forensics
  - BitCurator
- Academia
  - George Mason University
  - San Jose State University
  - University of California, Santa Cruz
Conclusion

• Diskprints are a record of system states.

• The workflow extracts artifacts and behaviors.

• Artifact attribution tells a computer’s software story.