Indexing the Windows® Registry for Software Detection

Alex J. Nelson[1,2], Mary T. Laamanen[3],
John Tebbutt[3], Darrell D.E. Long[1]

[1]University of California, Santa Cruz
[2]Prometheus Computing, LLC.
[3]National Institute of Standards and Technology

AAFS, February 20, 2014

DISCLAIMER / DISCLOSURE

Trade names and company products are mentioned in the text or identified. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, Prometheus Computing, LLC., or the University of California, Santa Cruz; nor does it imply that the products are necessarily the best available for the purpose. These products are mentioned for being relevant to the scientific results contained herein.

This research was funded by the National Institute of Standards and Technology Office of Law Enforcement Standards, the Department of Justice National Institute of Justice, the Department of the Navy, the Federal Bureau of Investigation and the National Archives and Records Administration.

What is the Registry to you?

- The Registry is a blend of:
  - Configuration file
  - Log
  - File system

What is the Registry to you?

- The Registry is a blend of:
  - Configuration file
    - Notes currently-live hardware
    - Stores application configs
    - Tells Windows how to start
  - Log
  - File system

What is the Registry to you?

- The Registry is a blend of:
  - Configuration file
  - Log
    - Last applications run, files opened
    - Used wi-fi access points
    - Shutdown time, etc.
    - Has its own set of timestamps
  - File system

What is the Registry to you?

- The Registry is a blend of:
  - Configuration file
  - Log
  - File system
    - Organizes data in a hierarchy
    - Has "Keys" (directories) that store "Values" (files)
    - Has slack space, just like a disk

Most Registry analysis is content analysis.

- Many Registry cells in fresh OS's
  - 100,000 in XP
  - 3-400,000 by Windows 7

- Plenty of analysis niches
  - A Perl library packages ≈300 scripts
  - ≈200 pages of case studies in some books

- The rest of this presentation is none of that: Instead, structural analysis

REGISTRY STRUCTURE

How is the Registry physically stored?

What information can we glean, from just the structure?

Registry logical layout: The RegEdit view

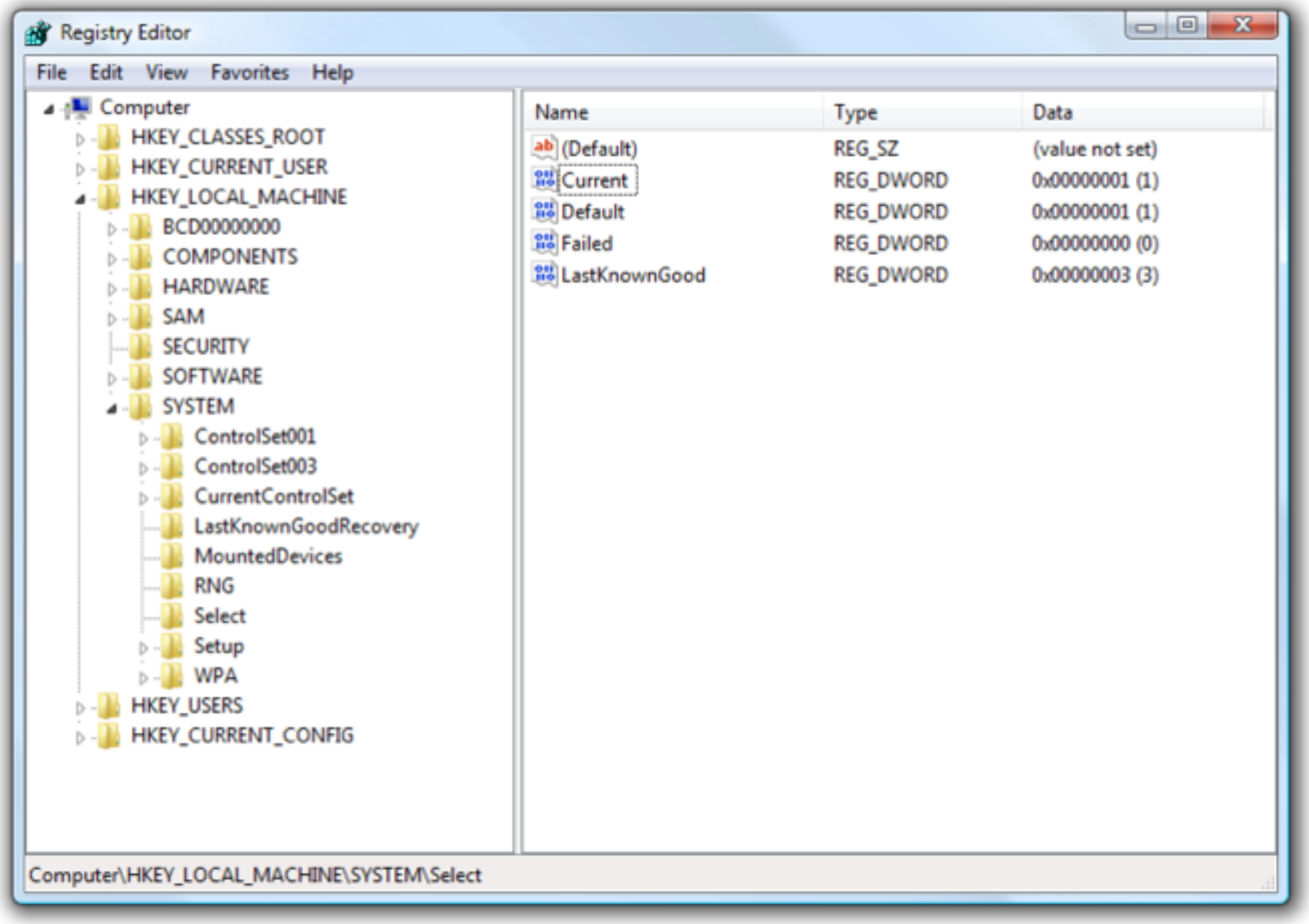


Image source: http://en.wikipedia.org/wiki/File:Registry_Editor_Vista.png

Registry logical layout: The RegEdit view

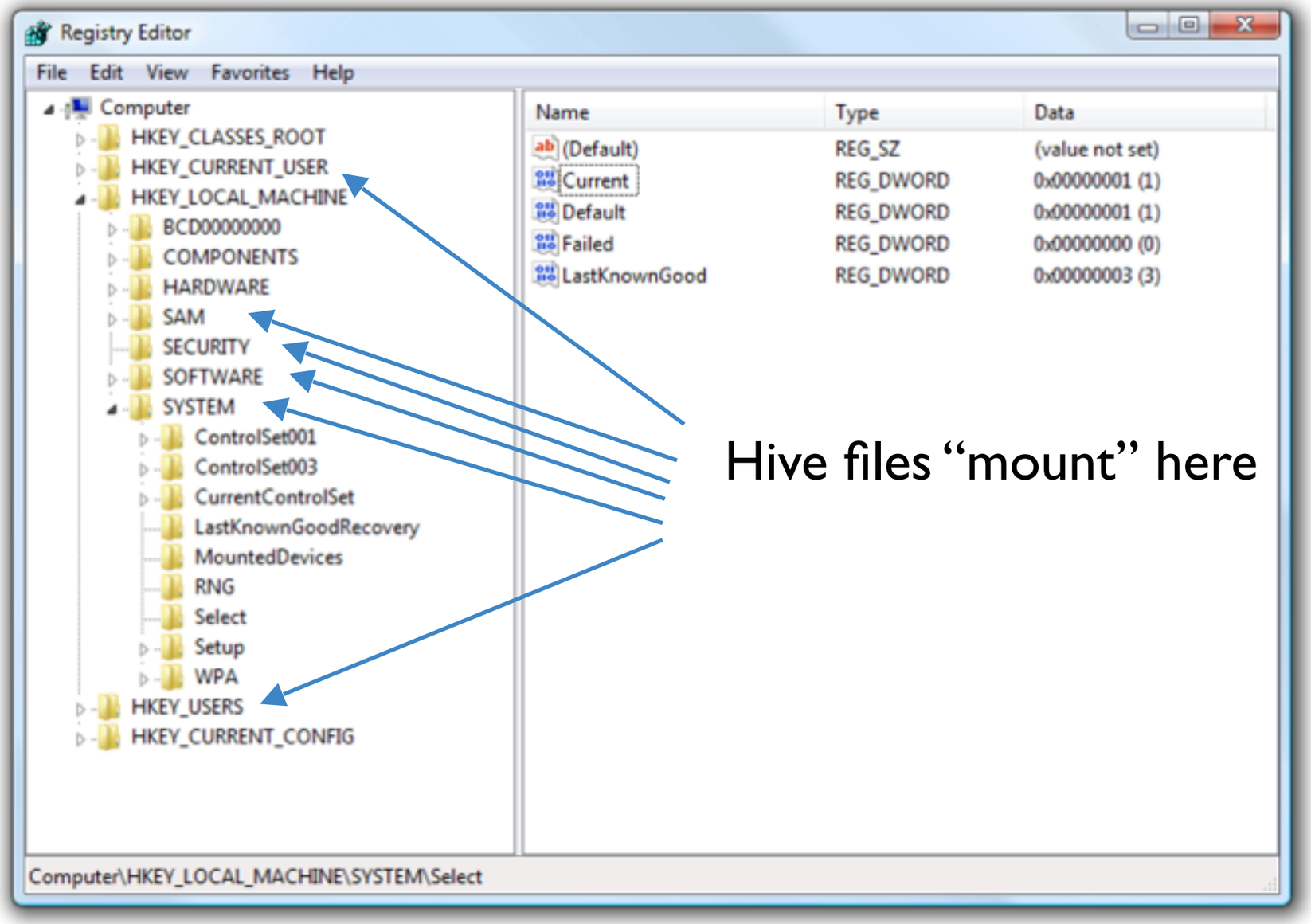


Image source: http://en.wikipedia.org/wiki/File:Registry_Editor_Vista.png

Where are hive files?

- System hives:
  - C:\WINDOWS\system32\config\SAM
  - C:\WINDOWS\system32\config\Security
  - C:\WINDOWS\system32\config\Software
  - C:\WINDOWS\system32\config\System

- User files:
  - C:\Users\<user>\ntuser.dat
  - C:\Users\<user>\local settings\application data\microsoft\windows\usrclass.dat

- Other hives:
  - C:\WINDOWS\system32\config\Default
  - C:\WINDOWS\system32\config\Components

- Backups are stored in various places.

What's in hive files?
File systems.

- *Hives* (≈ partitions) have a header, including:
  - Most of hive file name
  - Last-modified time

- Hives contain basic file systems:
  - *Keys* (≈ directories) have last-modified time
  - *Values* (≈ files) have typed data
    - (int, string, string list, raw binary, etc.)
  - *Cells* are keys and values
  - *Deleting* content is marking it unallocated

ANALYSIS @NIST

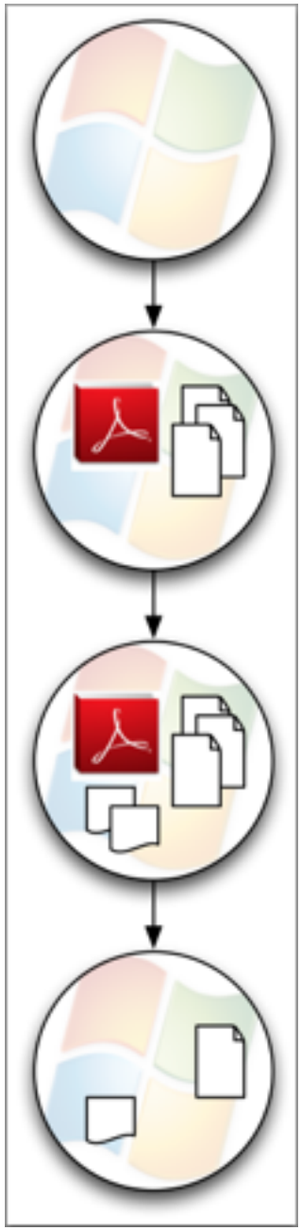
Fingerprinting applications, using Diskprints and Document Search

Fingerprinting applications opens computer software histories.

- With only Registry data, we can automatically determine:
  - If a computer ran "Hidden" software
    - *E.g.* thumb drive Firefox
  - Use of anti-forensics software
- ...just from Registry paths.

These "Fingerprints" come from Diskprints.

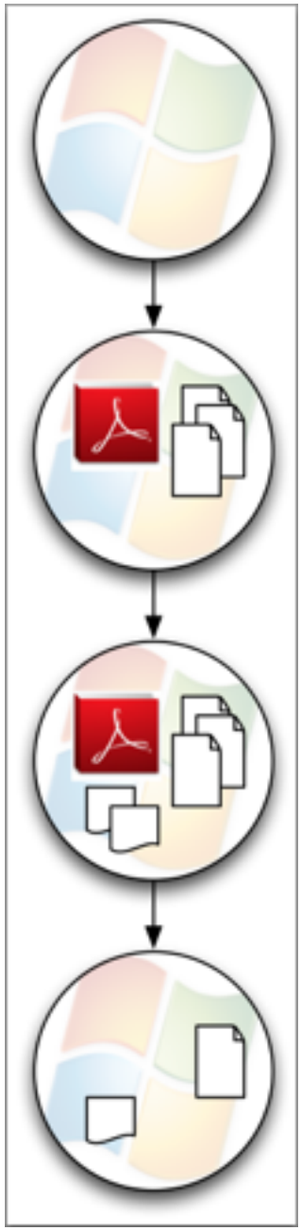
- Diskprints capture changes from:



  - Baseline
  - Installing
  - Running
  - Uninstalling

These "Fingerprints" come from Diskprints.

- Diskprints capture changes from:



  - Baseline

    → Δ (A change set)

  - Installing

    → Δ

  - Running

    → Δ

  - Uninstalling

Diskprint data are open for arbitrary post-mortem analysis.

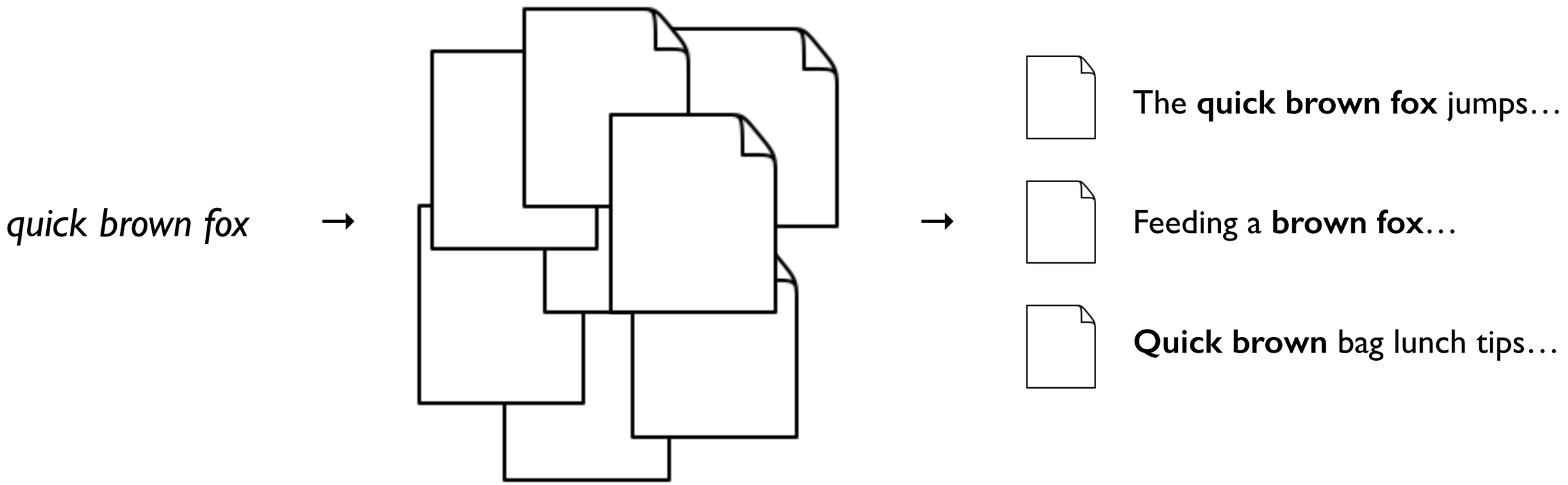
- Generated data include:
  - Network captures
  - RAM
  - Disk state
  - Registries

(Any analyses you want to run?)

(Apps you want printed?)

Document search overview

- Given a *corpus* of *documents*...
- Issue a *query*...
- And *rank* documents by *relevancy*.



Applying fingerprints

- Call changes to a Registry a *document*;
- Call added/removed/modified cells *terms*;
- Treat a computer's *Registry* as a *query*;

- Document search returns likely software used.

- First, must determine distinctness of our terms.

WHAT WE'VE LEARNED

Registry cell paths are distinct to operating system and application, but can be normalized.

Either way, they're indexable.

The NSRL printed these programs by June '13.

Operating system	Architecture	Application	Slice tally
XP Professional	32-bit	(OS)	7
Vista Ultimate with S.P. 1	32- & 64-bit	(OS)	9 (x2)
7 Ultimate	32- & 64-bit	(OS)	7 (x2)
Vista Ultimate with S.P. 1	32-bit	Adobe Acrobat Reader 3.0	4
Vista Ultimate with S.P. 1	32-bit	Adobe Dynamic Media Solutions	11
Vista Ultimate with S.P. 1	64-bit	Adobe Photoshop Lightroom 4	5
Vista Ultimate with S.P. 1	32- & 64-bit	Limewire Basic	5 (x2)
Vista Ultimate with S.P. 1	64-bit	Microsoft Office Home and Student 2010	5
Vista Ultimate with S.P. 1	64-bit	Norton AntiVirus 2012 with Antispyware	4
Vista Ultimate with S.P. 1	32- & 64-bit	StreetFinder Travel Navigation Software	5 (x2)
Vista Ultimate with S.P. 1	32- & 64-bit	TurboTax Deluxe Plus State	9 (x2)
Vista Ultimate with S.P. 1	32- & 64-bit	mozilla Firefox 2	4 (x2)

Registry paths are distinct to major Windows versions.

Overlap of raw cell paths in baseline images: (all hives)

	xp-32 (1)	xp-32 (2)	vista-32	vista-64	7-32	7-64	8-32	8-64
xp-32 (1)	115474	111770	0	0	0	0	0	0
xp-32 (2)	111770	116066	0	0	0	0	0	0
vista-32	0	0	228702	62	0	0	60	60
vista-64	0	0	62	359314	0	0	60	60
7-32	0	0	0	0	307251	63	0	0
7-64	0	0	0	0	63	443806	0	0
8-32	0	0	60	60	0	0	318396	250095
8-64	0	0	60	60	0	0	250095	466628

Registry paths can be normalized across Windows versions.

Overlap of normalized cell paths in baseline images (System hive normalized):

	xp-32 (1)	xp-32 (2)	vista-32	vista-64	7-32	7-64	8-32	8-64
xp-32 (1)	115474	111770	10585	10491	8351	8114	46610	44212
xp-32 (2)	111770	116066	10739	10633	8560	8243	47320	44905
vista-32	10585	10739	228702	62136	31499	30936	11565	11473
vista-64	10491	10633	62136	359314	31030	31278	11414	11583
7-32	8351	8560	31499	31030	307251	84728	13627	13482
7-64	8114	8243	30936	31278	84728	443806	13489	13520
8-32	46610	47320	11565	11414	13627	13489	318396	250095
8-64	44212	44905	11473	11583	13482	13520	250095	466628

An app printed in one OS should be detectable in other OS's.

Added Registry paths are distinct to the application.

Most added keys only appear in one application's prints (Note: using raw paths):

Applications with a given key added	Count
1	2171
2	237
3	24
4	6
5	3
6	31
7	39

Early results: A Vista machine's software profile

- Disk image: M57-Patents "Terry"
- Querying the to-date Disk Printed apps, using cells added, raw paths.

Application	Step	Similarity
Adobe Acrobat Reader 3.0	Close	0.630023
mozilla Firefox 2	Install	0.338983
mozilla Firefox 2	Uninstall	0.169255
Adobe Dynamic Media Solutions	Uninstall (3 of 4)	0.166305
mozilla Firefox 2	Close	0.158567
StreetFinder Travel Navigation	Open	0.116993
Limewire Basic	Uninstall	0.108174
...		

Next steps

- Print more apps

- Run ground-truth app searches on M57-patents corpus (including malware)

- Apply to file system namespaces

- Look up software used in other forensic corpora

Conclusion

- The Registry is:
  - A configuration store
  - A log
  - A file system

- Hundreds of thousands of cells on any computer

- Software leaves fingerprints in the Registry – distinct, and detectable

Questions?

- My email:
  - a.nelson@prometheuscomputing.com

- Registry analysis code is available.
  - Diskprint analytic workflow:
    - https://github.com/ajnelson/diskprint_workflow
  - RegXML Extractor:
    - https://github.com/ajnelson/regxml_extractor

- Image credits:
  - Wikipedia, Amazon, OmniGraffle, Adobe