Inferring Past Activity from Partial Digital Artifacts

By

James Jones, Tahir Khan, Kathryn Laskey, Alexander Nelson, Mary Laamanen and Douglas White

Presented At

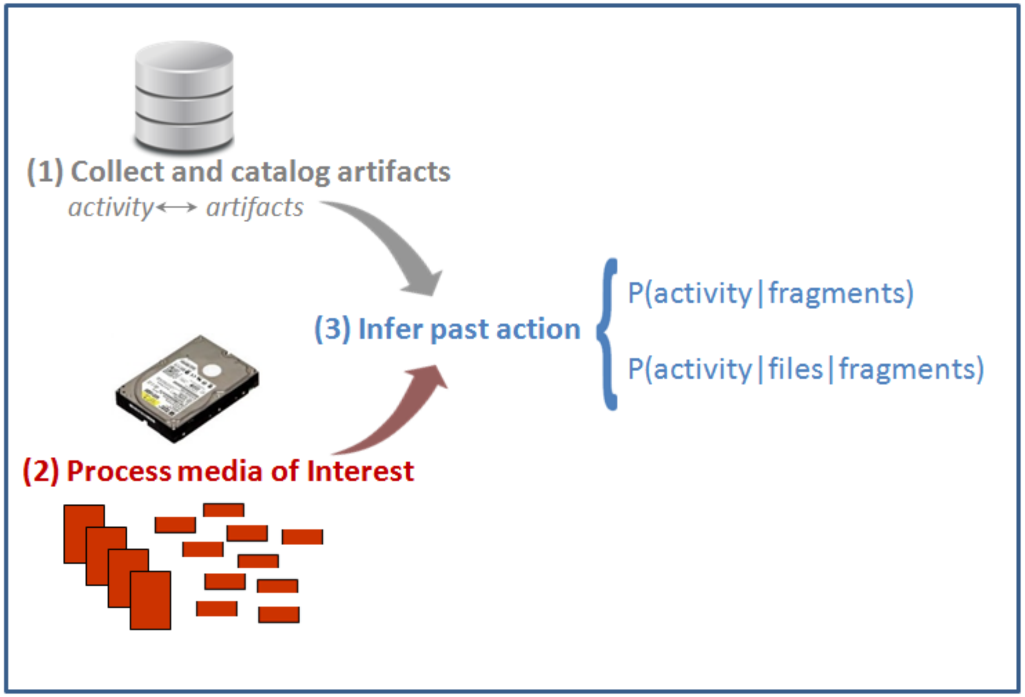
The Digital Forensic Research Conference

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

Inferring Past Activity from Partial Digital Artifacts

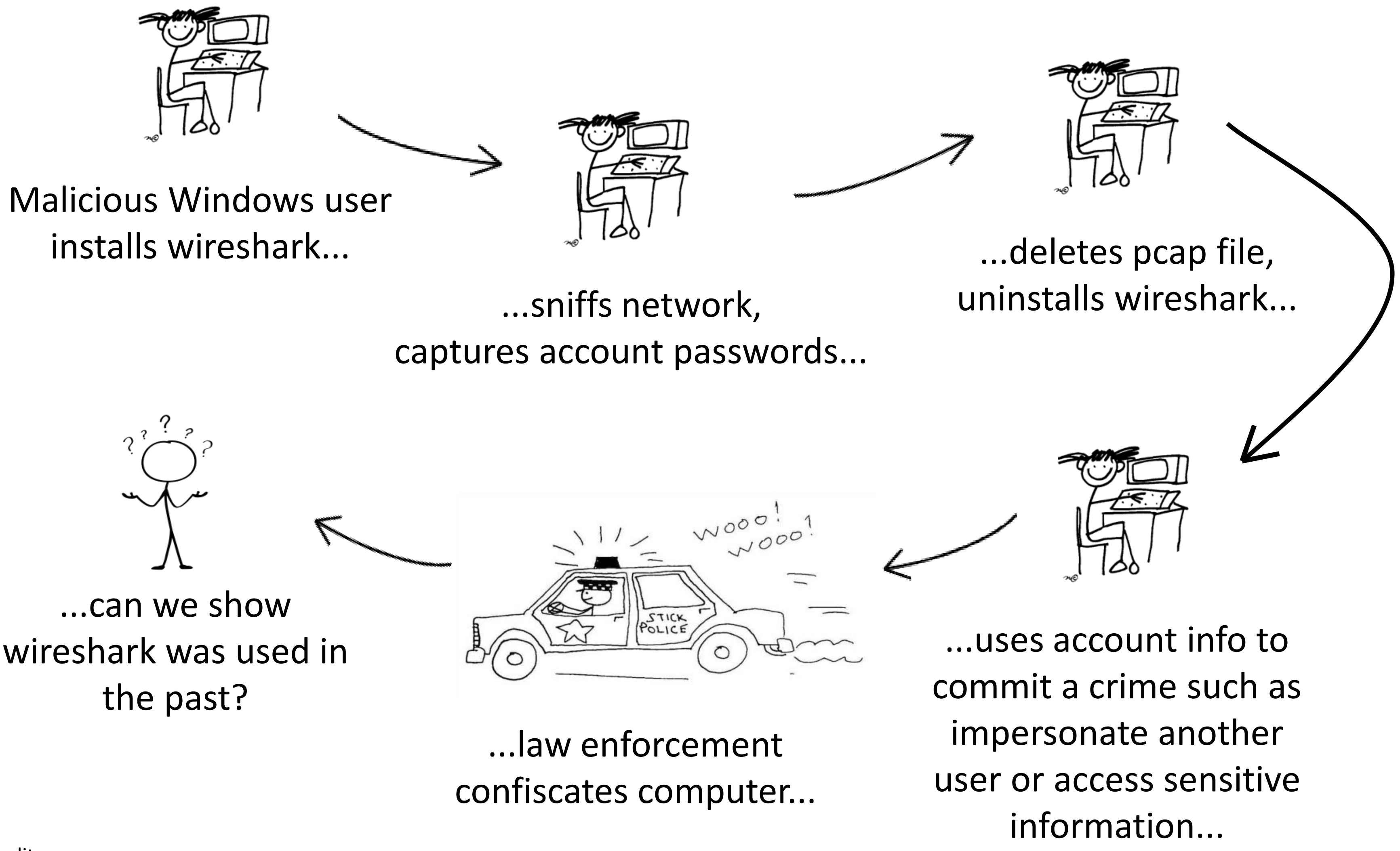
Jim Jones[†], Tahir Khan[†], Kathy Laskey[†], Alex Nelson[‡], Mary Laamanen[‡], Doug White[‡]

[†]George Mason University, [‡]National Institute of Standards and Technology


(1) Collect and catalog artifacts
activity←→ artifacts
(3) Infer past action
P(activity|fragments)
P(activity|files|fragments)
(2) Process media of Interest

This presentation results from research supported by the Naval Postgraduate School Assistance Grant/Agreement No. N00244-13-1-0034 awarded by the NAVSUP Fleet Logistics Center San Diego (NAVSUP FLC San Diego). The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School or National Institute of Standards and Technology, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

A user may uninstall an application to disguise past usage



Malicious Windows user installs wireshark...

...sniffs network, captures account passwords...

...deletes pcap file, uninstalls wireshark...

...uses account info to commit a crime such as impersonate another user or access sensitive information...

...law enforcement confiscates computer...

...can we show wireshark was used in the past?

Image credits:
- http://school.discoveryeducation.com/clipart/clip/stk-fgr6.html
- http://blog.deming.org/2014/10/the-target-is-irrelevant-without-a-method/
- http://www.presentermedia.com/index.php?target=closeup&maincat=clipart&id=9771

We reason over partial file artifacts to infer past application usage

Situation:

Uninstalling an application deletes files associated with the application. These deleted files decay over time, i.e., pieces (sectors) of the deleted files are overwritten. Current forensic techniques rely on finding whole and intact deleted files, which may not be available.

Question:

Can we infer past application installation and use when the application has been uninstalled and activity such as reboots and normal usage have continued?

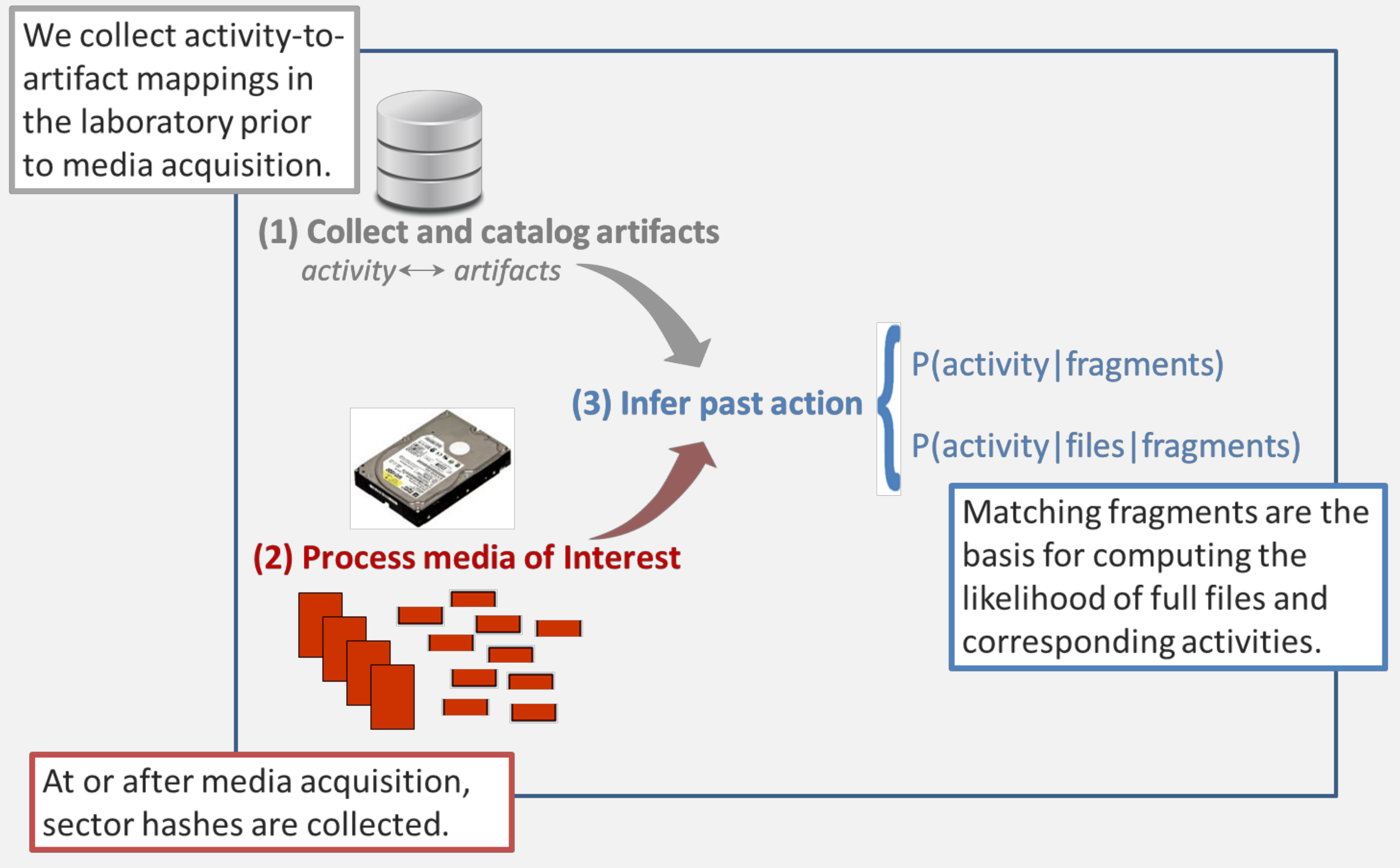
Answer:

Yes, by reasoning over the artifact fragments (file sectors) that remain.

Innovation:

Reasoning over *weighted collections* of artifact fragments.

Our approach reasons over media sectors that match a database associating sectors with application activity



We collect activity-to-artifact mappings in the laboratory prior to media acquisition.

(1) Collect and catalog artifacts

activity ←→ artifacts

(2) Process media of Interest

(3) Infer past action

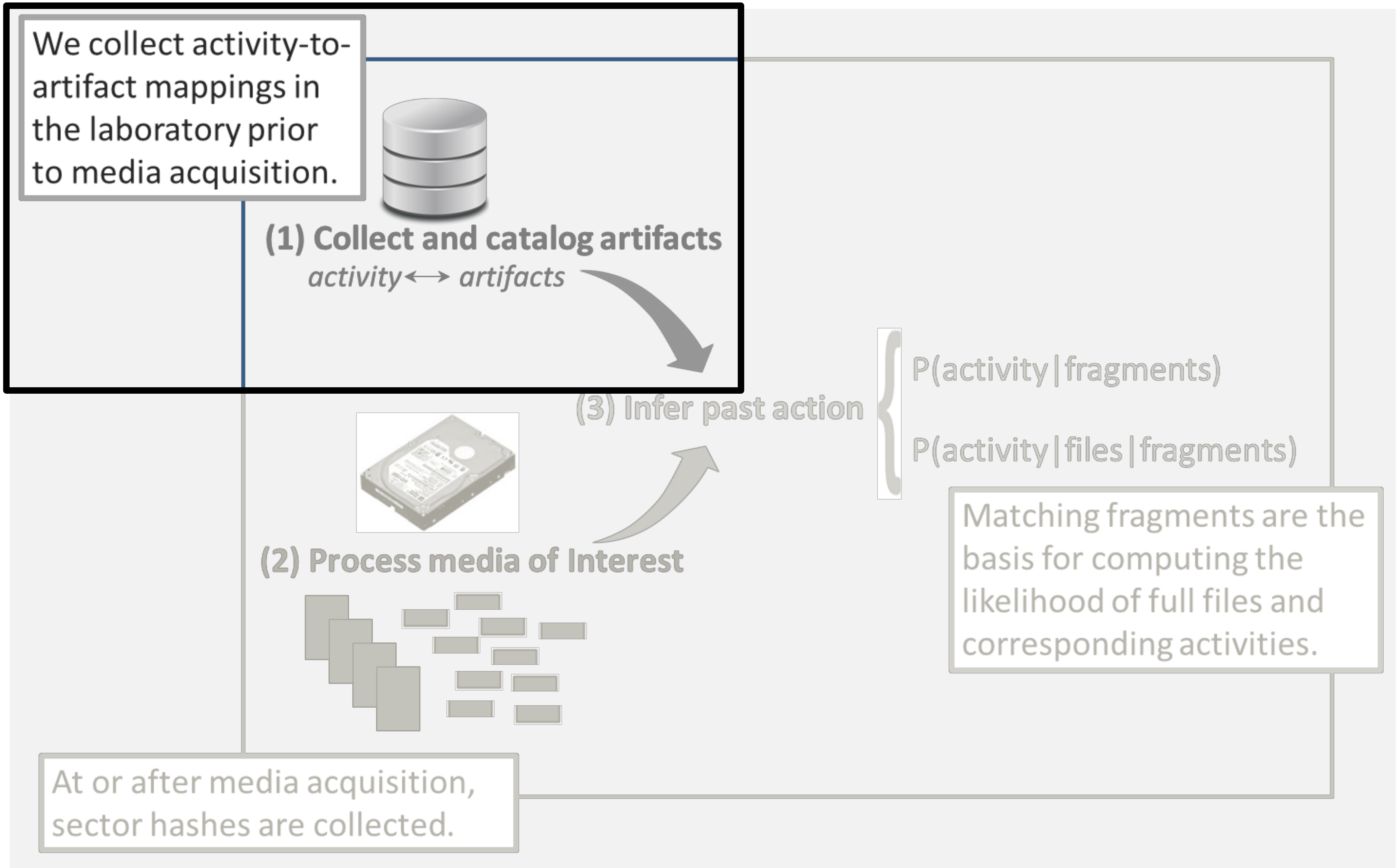
P(activity|fragments)

P(activity|files|fragments)

Matching fragments are the basis for computing the likelihood of full files and corresponding activities.

At or after media acquisition, sector hashes are collected.

Step 1: Collect and catalog artifacts



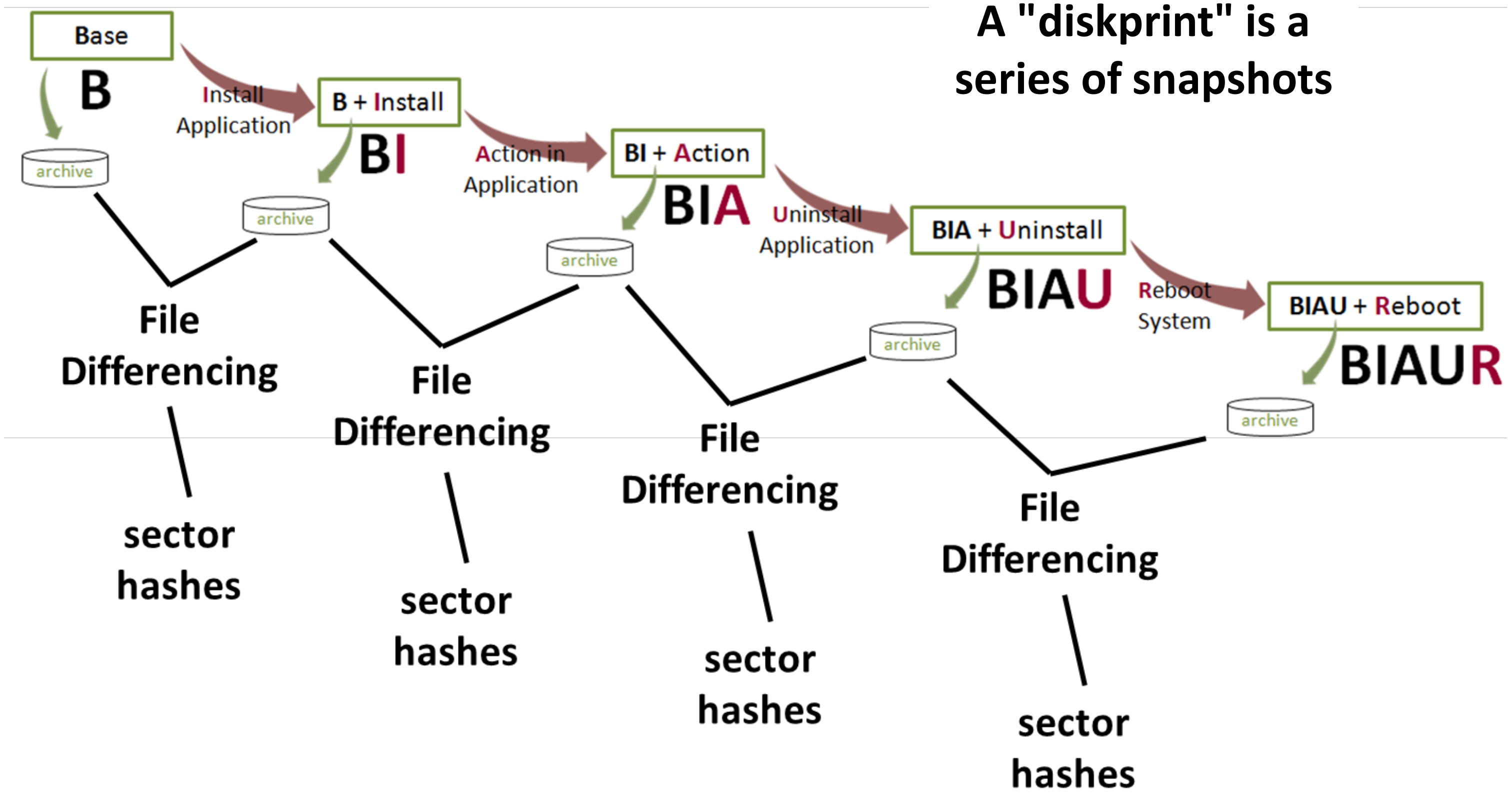
We collect activity-to-artifact mappings in the laboratory prior to media acquisition.

(1) Collect and catalog artifacts
activity⟷ artifacts
(3) Infer past action
P(activity|fragments)
P(activity|files|fragments)
(2) Process media of Interest

Matching fragments are the basis for computing the likelihood of full files and corresponding activities.

At or after media acquisition, sector hashes are collected.

We repeated the file differencing process to collect artifacts (files) over a sequence of related activities



A "diskprint" is a series of snapshots

Base

B

archive

Install Application

B + Install

BI

archive

Action in Application

BI + Action

BIA

archive

Uninstall Application

BIA + Uninstall

BIAU

archive

Reboot System

BIAU + Reboot

BIAUR

archive

File Differencing

sector hashes

File Differencing

sector hashes

File Differencing

sector hashes

File Differencing

sector hashes

Initial diskprinting generated 93M sector hashes from 66k files

16 applications:

- Adv Keylogger
- Chrome
- Eraser
- Firefox
- HxD hex editor
- Invisible Secrets
- MS Office
- Python
- Safari
- Sdelete
- Thunderbird
- TrueCrypt
- UPX
- WinRar
- WinZip
- Wireshark

3 operating systems:

- Windows XP (32 bit)
- Windows 7-32bit
- Windows 7-64bit

5 actions:

- Install
- Open
- Close
- Uninstall
- Reboot

Data set:

- 29 diskprints
- 186 slices
- 167 difference sets
- ~66k files
- ~93M hashes
    - f < 100

	WinXP	Win7x32	Win7x64
Adv Keylogger	✓		
Chrome	✓	✓	✓
Eraser		✓	
Firefox	✓	✓	✓
HxD hex editor		✓	
Invisible Secrets	✓		
MS Office	✓	✓	✓
Python	✓		
Safari	✓	✓	✓
Sdelete		✓	✓
Thunderbird	✓		
TrueCrypt	✓		
UPX		✓	✓
WinRar		✓	✓
WinZip		✓	✓
Wireshark		✓	✓

We remove file differencing noise and non-probative artifacts

Three categories of artifacts are collected:

A. spurious

B. positively attributed but not probative

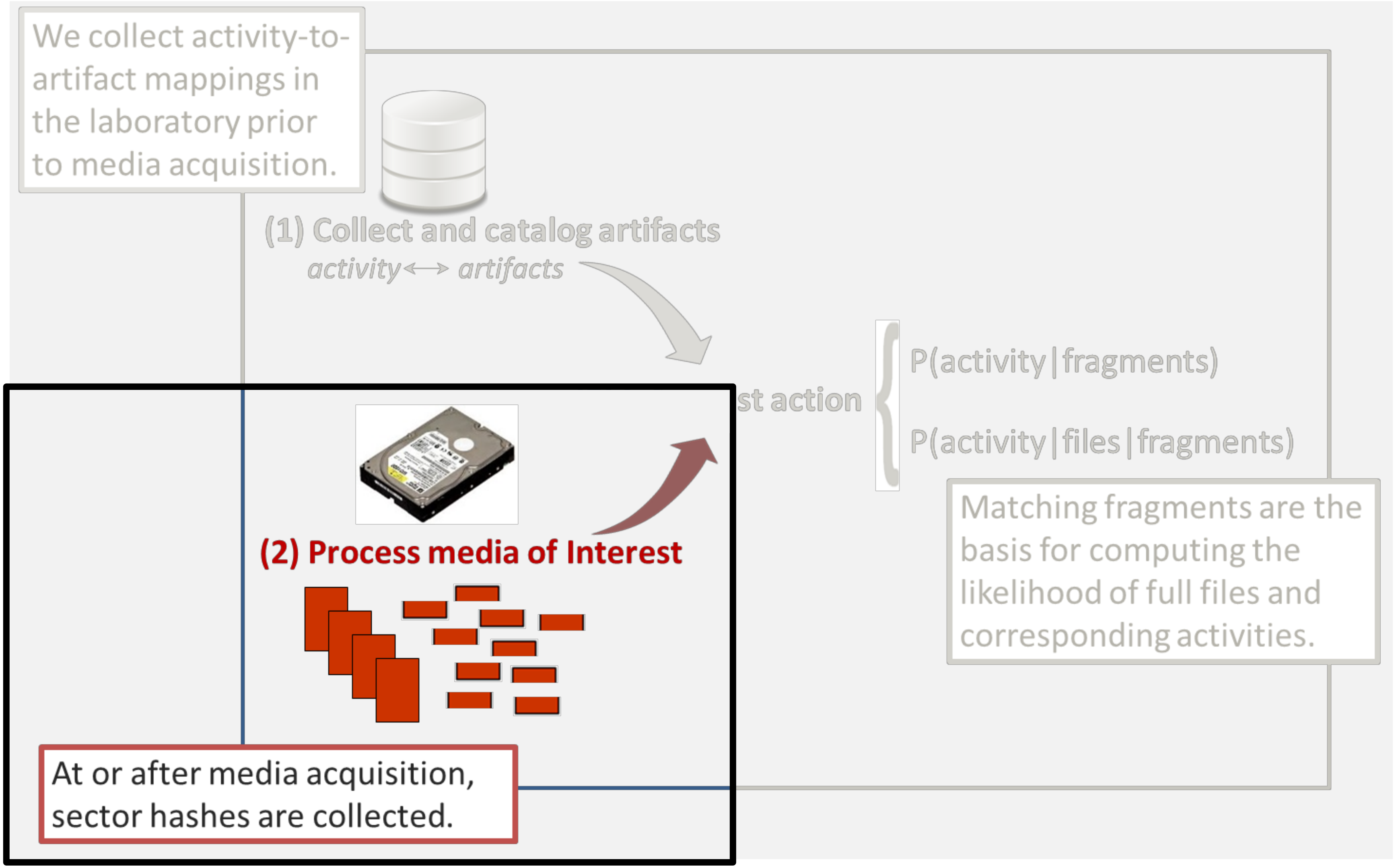
C. positively attributed and possibly probative

Select category C by post-processing:

- include by keyword (owning file's path and filename)
- exclude by OS image comparison
- exclude if low entropy
- include if frequency < 100

RESULT: ~8M hashes from ~20k files

Step 2: Process media of interest



We collect activity-to-artifact mappings in the laboratory prior to media acquisition.

(1) Collect and catalog artifacts

activity⟵⟶ artifacts

st action

P(activity|fragments)

P(activity|files|fragments)

Matching fragments are the basis for computing the likelihood of full files and corresponding activities.

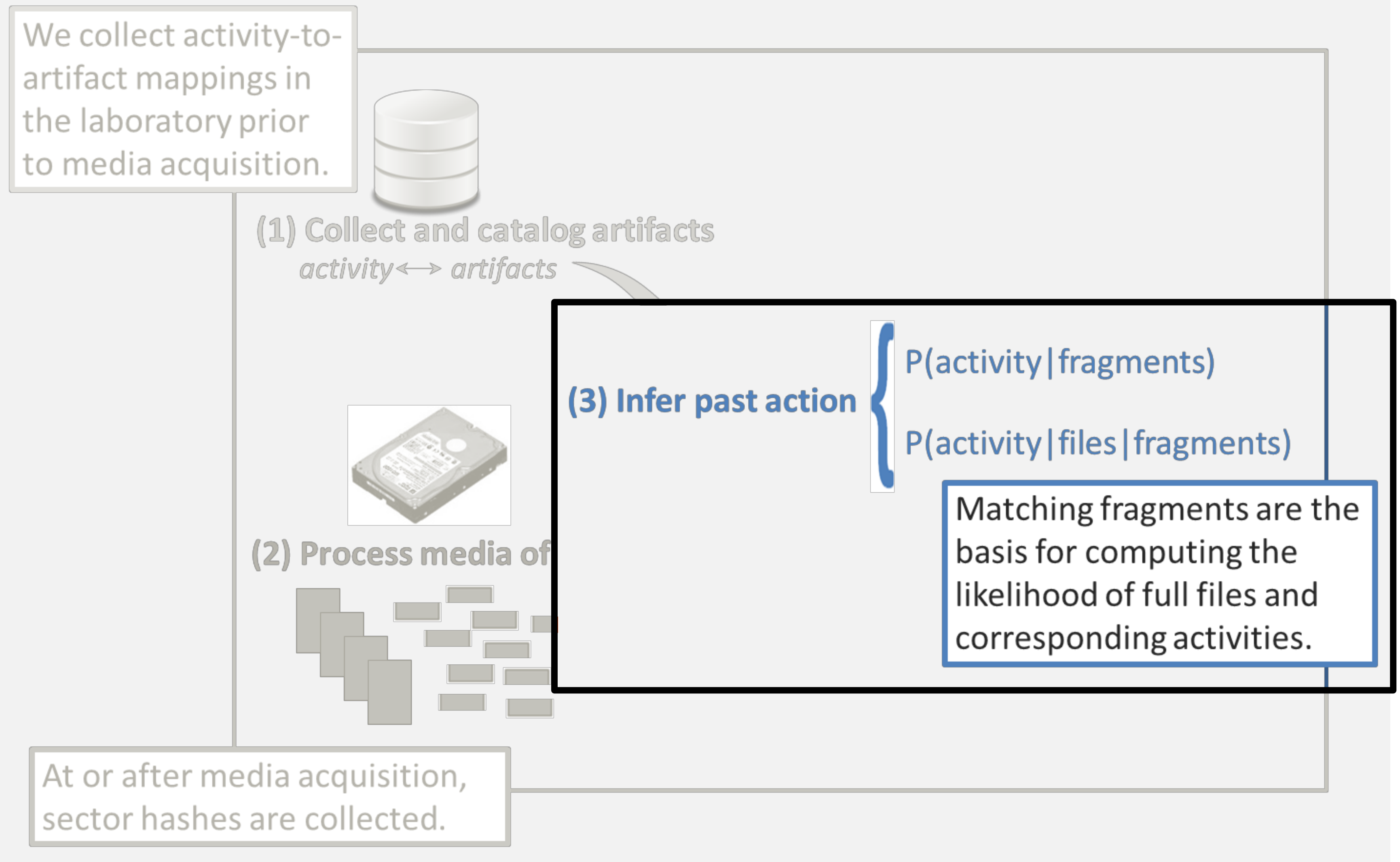
(2) Process media of Interest

At or after media acquisition, sector hashes are collected.

We hash the sectors on media of interest

- md5deep
- sector-aligned piecewise hashing
- 512-byte sectors

Step 3: Infer past action



We collect activity-to-artifact mappings in the laboratory prior to media acquisition.

(1) Collect and catalog artifacts

activity⟷ artifacts

(3) Infer past action

P(activity|fragments)

P(activity|files|fragments)

Matching fragments are the basis for computing the likelihood of full files and corresponding activities.

(2) Process media of

At or after media acquisition, sector hashes are collected.

We created and processed 11 test images

8 applications:

- Adv Keylogger
- **Chrome**
- Eraser
- **Firefox**
- HxD hex editor
- Invisible Secrets
- MS Office
- Python
- **Safari**
- **Sdelete**
- Thunderbird
- TrueCrypt
- **UPX**
- **WinRar**
- **WinZip**
- **Wireshark**

1 operating system:

- Windows XP (32 bit)
- Windows 7-32bit
- **Windows 7-64bit**

5 actions:

- **Install**
- **Open**
- **Close**
- **Uninstall**
- **Reboot**

Data set:

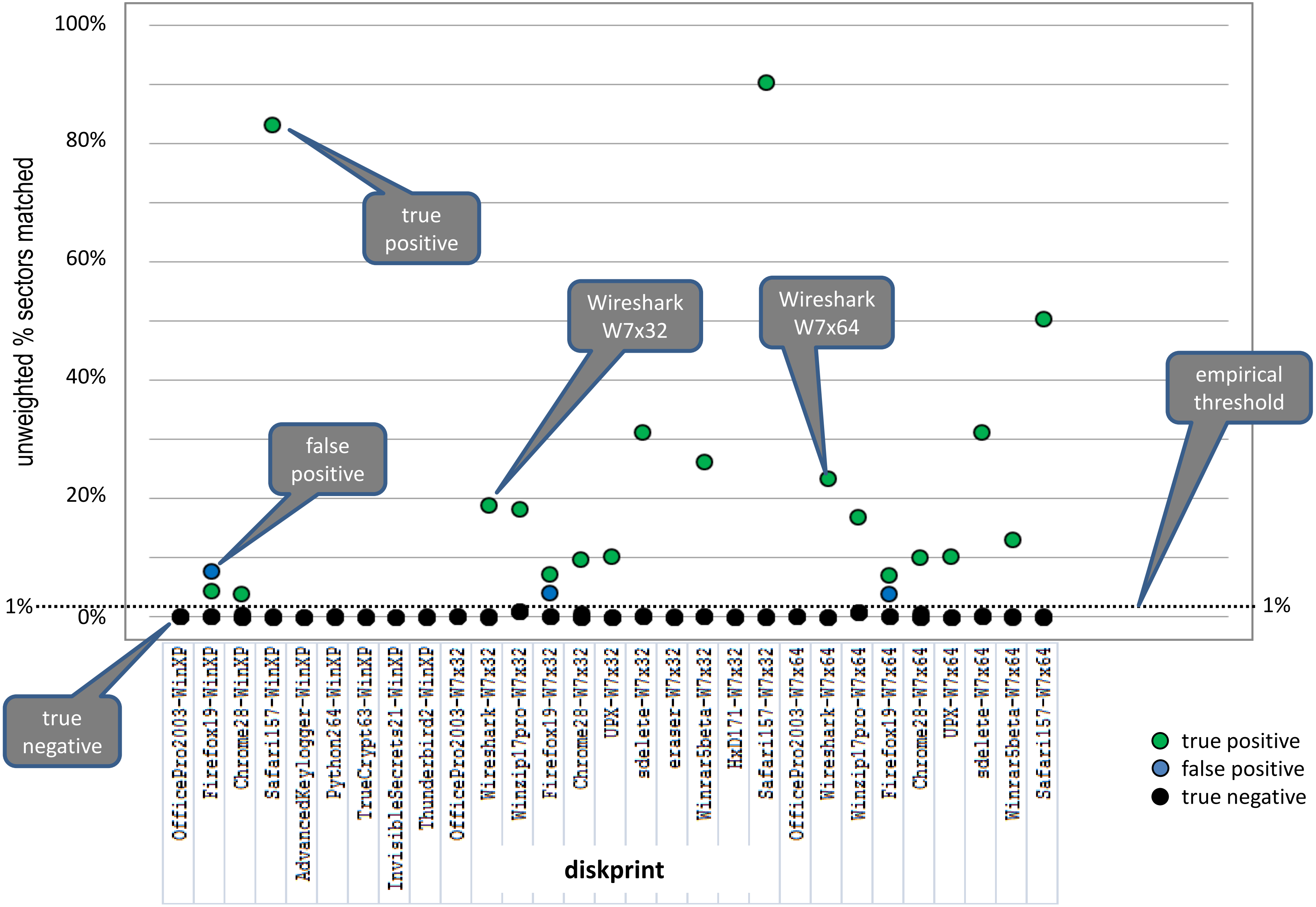
- 11 image sequences
  - 8 single and 3 multi-app
- 55 snapshots
  - 64 GB disks
- 55 hash sets
  - ~125M hashes each

	WinXP	Win7x32	Win7x64
Adv Keylogger	✓		
Chrome	✓	✓	✔
Eraser		✓	
Firefox	✓	✓	✔
HxD hex editor		✓	
Invisible Secrets	✓		
MS Office	✓	✓	✓
Python	✓		
Safari	✓	✓	✔
Sdelete		✓	✔
Thunderbird	✓		
TrueCrypt	✓		
UPX		✓	✔
WinRar		✓	✔
WinZip		✓	✔
Wireshark		✓	✔

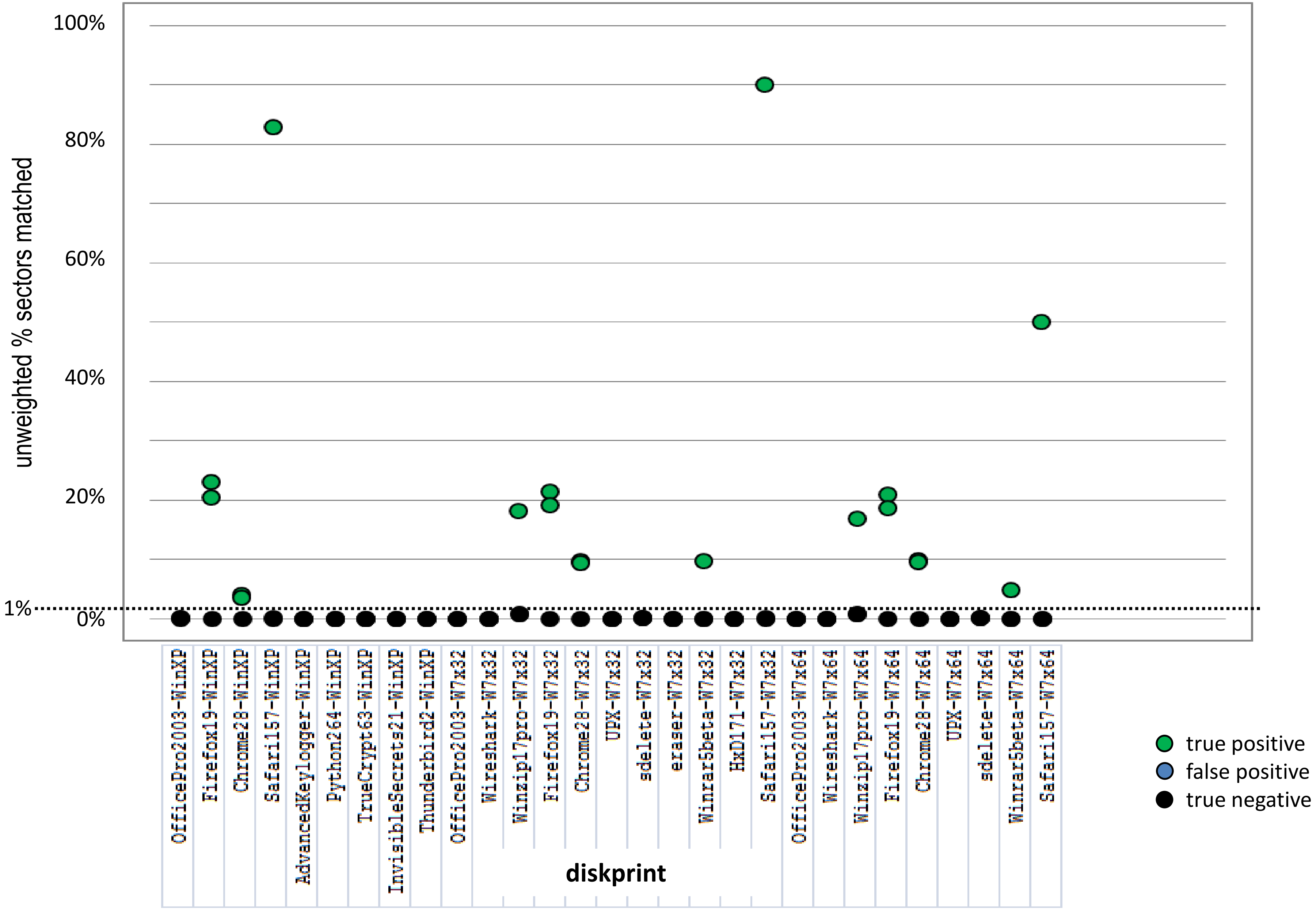
The algorithm computes % of sectors (hashes) matched between source image and each diskprint in the catalog

```
$ python3 ~/hashdb/process_img.py 9480-2-14416-1-50 -p
Processing matches...

Source Image: 9480-2-14416-1-50 (Wireshark-W7x64)

Results:
diskprintID        diskprintName             sectors_found sectors_total sector%
9480-2-14416-1     Wireshark-W7x64                   48995        209666  23.37%
9480-1-14417-1     Wireshark-W7x32                   32484        171515  18.94%
9480-1-14782-1     Winzip17pro-W7x32                  2135        240229   0.89%
9480-2-14782-1     Winzip17pro-W7x64                  2162        262854   0.82%
9480-1-15142-1     sdelete-W7x32                         1           642   0.16%
9480-2-15142-1     sdelete-W7x64                         1           642   0.16%
234-1-14351-1      OfficePro2003-WinXP                1004        656354   0.15%
9480-2-14351-1     OfficePro2003-W7x64                1004       1077126   0.09%
9480-1-14351-1     OfficePro2003-W7x32                1004       1090216   0.09%
9480-1-15149-1     Winrar5beta-W7x32                     8          9196   0.09%
234-1-14887-1      Firefox19-WinXP                      45         96377   0.05%
9480-2-15149-1     Winrar5beta-W7x64                     8         18328   0.04%
9480-1-14887-1     Firefox19-W7x32                      44        103341   0.04%
9480-1-15150-1     HxD171-W7x32                          2          4774   0.04%
9480-2-14887-1     Firefox19-W7x64                      44        106270   0.04%
234-1-7959-1       Thunderbird2-WinXP                   16         68102   0.02%
234-1-15487-1      Python264-WinXP                      20         86287   0.02%
9480-1-15146-1     eraser-W7x32                         13         69984   0.02%
9480-2-15137-1     Chrome28-W7x64                       92        670051   0.01%
234-1-15137-1      Chrome28-WinXP                      139       1035098   0.01%
9480-1-15137-1     Chrome28-W7x32                       92        686986   0.01%
9480-1-15151-1     Safari157-W7x32                      29        316224   0.01%
234-1-15151-1      Safari157-WinXP                      29        343824   0.01%
9480-2-15151-1     Safari157-W7x64                      32        569645   0.01%
234-1-15488-1      TrueCrypt63-WinXP                     1         24520   0.00%
234-1-15485-1      AdvancedKeylogger-WinXP               0          4716   0.00%
234-1-15489-1      InvisibleSecrets21-WinXP              0          6689   0.00%
9480-1-15141-1     UPX-W7x32                             0          1796   0.00%
9480-2-15141-1     UPX-W7x64                             0          1813   0.00%
```

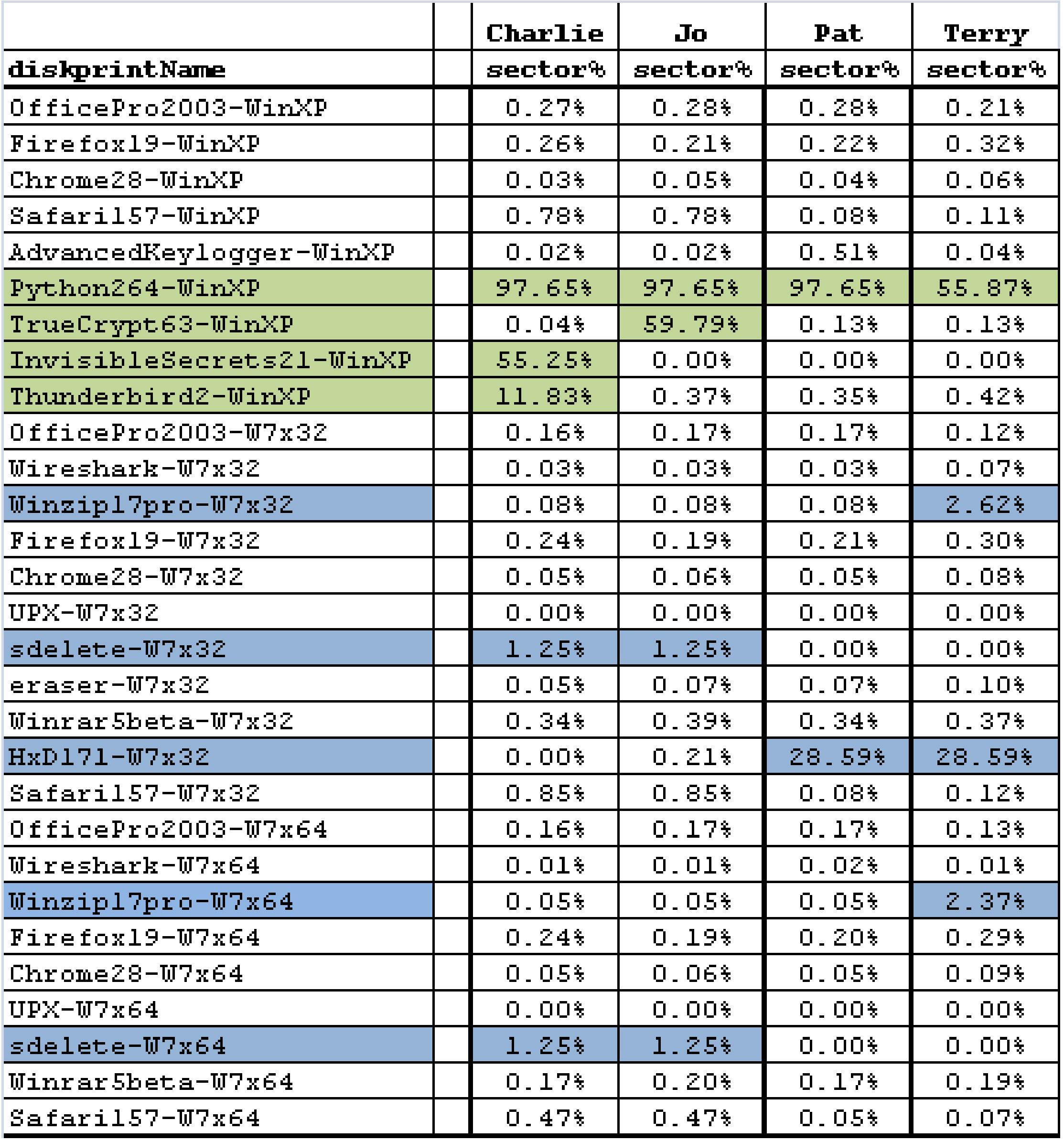
We processed 8 single-application test images



We processed 3 multi-application test cases



We found indications of additional applications in the M57 data set (final snapshots)

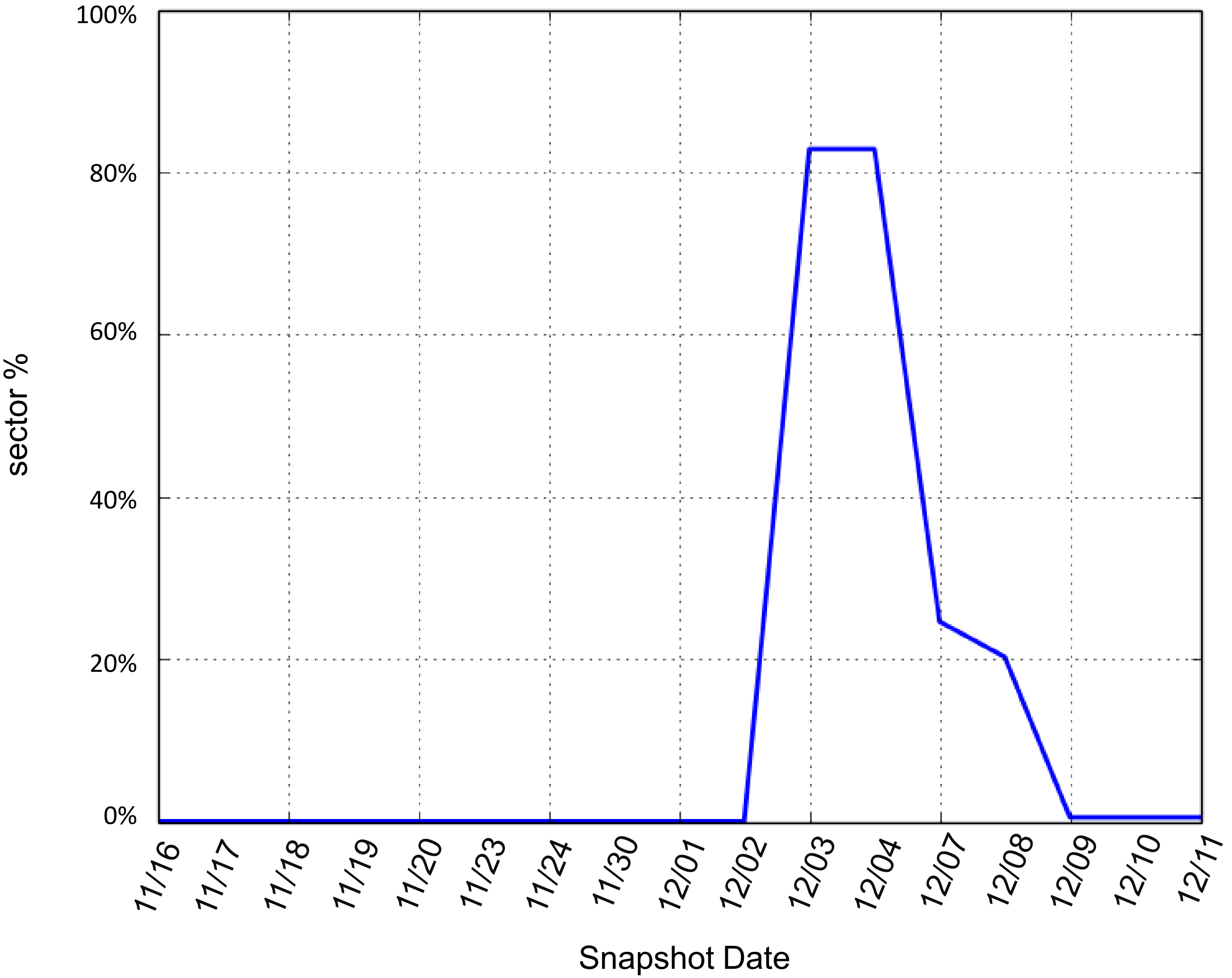


		Charlie	Jo	Pat	Terry
diskprintName		sector%	sector%	sector%	sector%
OfficePro2003-WinXP		0.27%	0.28%	0.28%	0.21%
Firefox19-WinXP		0.26%	0.21%	0.22%	0.32%
Chrome28-WinXP		0.03%	0.05%	0.04%	0.06%
Safari157-WinXP		0.78%	0.78%	0.08%	0.11%
AdvancedKeylogger-WinXP		0.02%	0.02%	0.51%	0.04%
Python264-WinXP		97.65%	97.65%	97.65%	55.87%
TrueCrypt63-WinXP		0.04%	59.79%	0.13%	0.13%
InvisibleSecrets21-WinXP		55.25%	0.00%	0.00%	0.00%
Thunderbird2-WinXP		11.83%	0.37%	0.35%	0.42%
OfficePro2003-W7x32		0.16%	0.17%	0.17%	0.12%
Wireshark-W7x32		0.03%	0.03%	0.03%	0.07%
Winzip17pro-W7x32		0.08%	0.08%	0.08%	2.62%
Firefox19-W7x32		0.24%	0.19%	0.21%	0.30%
Chrome28-W7x32		0.05%	0.06%	0.05%	0.08%
UPX-W7x32		0.00%	0.00%	0.00%	0.00%
sdelete-W7x32		1.25%	1.25%	0.00%	0.00%
eraser-W7x32		0.05%	0.07%	0.07%	0.10%
Winrar5beta-W7x32		0.34%	0.39%	0.34%	0.37%
HxD171-W7x32		0.00%	0.21%	28.59%	28.59%
Safari157-W7x32		0.85%	0.85%	0.08%	0.12%
OfficePro2003-W7x64		0.16%	0.17%	0.17%	0.13%
Wireshark-W7x64		0.01%	0.01%	0.02%	0.01%
Winzip17pro-W7x64		0.05%	0.05%	0.05%	2.37%
Firefox19-W7x64		0.24%	0.19%	0.20%	0.29%
Chrome28-W7x64		0.05%	0.06%	0.05%	0.09%
UPX-W7x64		0.00%	0.00%	0.00%	0.00%
sdelete-W7x64		1.25%	1.25%	0.00%	0.00%
Winrar5beta-W7x64		0.17%	0.20%	0.17%	0.19%
Safari157-W7x64		0.47%	0.47%	0.05%	0.07%

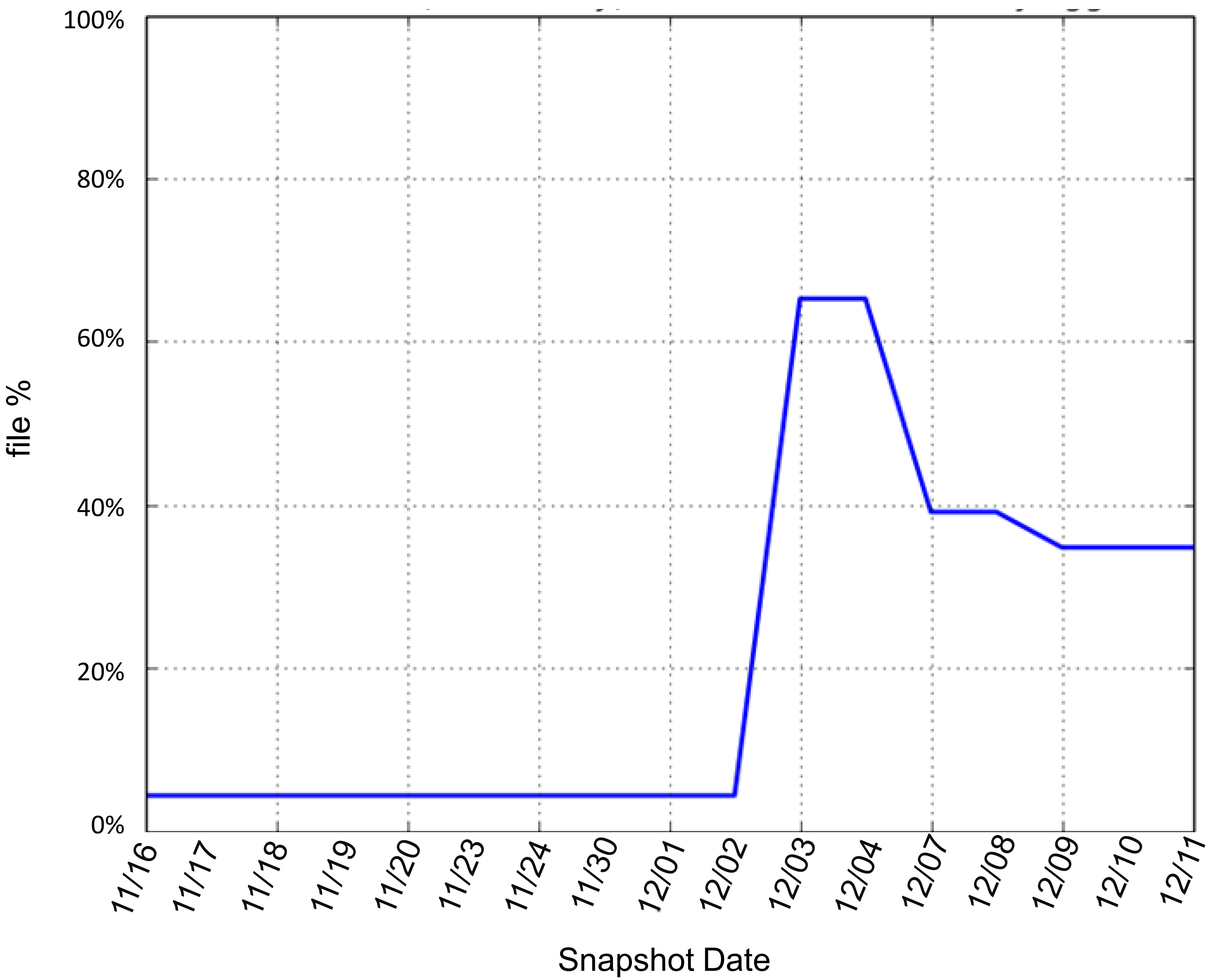
Python: Visible
TrueCrypt: Visible
InvisibleSecrets: Visible
Thunderbird: Visible

Winzip: Not visible
sdelete: Not visible
HxD: Not visible

M57 Pat (AdvancedKeylogger) partial artifact (sector) persistence after uninstall



M57 Pat (AdvancedKeylogger) partial artifact (file) persistence after uninstall



File hits are weighted by % of file matched

- File hits:
  - Original: files_found/ files_total
  - Weighted:

$$\left(\sum_{i=1}^{\text{num\_file\_matches}} \frac{\text{matched\_sectors}_F}{\text{total\_sectors}_F} \right) / \text{ files\_total}_{DP}$$

- Example:
  - Original: (1 + 1)/5 = 40%
  - Weighted: (3/5 + 1/10)/5 = 14%

Sector hits are weighted by catalog frequency

- Sector hits:
  - Original: sectors_found/ sectors_total
  - Weighted:

$$\left(\sum_{i=1}^{\text{num\_sec\_matches}} 1 / \text{freq}_S \right) / \text{sectors\_total}_{DP}$$

- Example:
  - Original: (1 + 1 + 1)/10 = 30%
  - Weighted: (1/1 + 1/4 + 1/2)/10 = 17.5%

Weighted output for M57 Pat (final image)

```
Source Image: pat20091211

Results:
diskprintID        diskprintName              sectors_found sectors_total sector% w_sector% files_found files_total   file% w_file%
234-1-15487-1      Python264-WinXP                    84260         86287  97.65%    97.05%        2341        2355  99.41%  98.91%
234-1-7959-1       Thunderbird2-WinXP                   239         68102   0.35%     0.30%          77         172  44.77%  24.94%
234-1-15485-1      AdvancedKeylogger-WinXP               24          4716   0.51%     0.49%           8          23  34.78%  21.97%
9480-1-15150-1     HxD171-W7x32                        1365          4774  28.59%    28.55%           2          12  16.67%   8.39%
234-1-14887-1      Firefox19-WinXP                      213         96377   0.22%     0.06%          43         115  37.39%   3.17%
9480-2-14887-1     Firefox19-W7x64                      213        106270   0.20%     0.06%          44         146  30.14%   2.93%
9480-1-14887-1     Firefox19-W7x32                      213        103341   0.21%     0.06%          44         132  33.33%   2.78%
9480-2-14782-1     Winzip17pro-W7x64                    137        262854   0.05%     0.03%          29         153  18.95%   2.03%
234-1-15137-1      Chrome28-WinXP                       416       1035098   0.04%     0.02%         118         624  18.91%   1.64%
9480-2-15137-1     Chrome28-W7x64                       314        670051   0.05%     0.02%          36         499   7.21%   1.63%
9480-1-14782-1     Winzip17pro-W7x32                    204        240229   0.08%     0.05%          35         149  23.49%   1.50%
234-1-15488-1      TrueCrypt63-WinXP                     32         24520   0.13%     0.09%           5          16  31.25%   1.22%
9480-1-15137-1     Chrome28-W7x32                       313        686986   0.05%     0.02%          36         669   5.38%   1.22%
9480-2-15149-1     Winrar5beta-W7x64                     31         18328   0.17%     0.04%          14          81  17.28%   0.85%
9480-1-15149-1     Winrar5beta-W7x32                     31          9196   0.34%     0.08%          14          41  34.15%   0.84%
234-1-15151-1      Safari157-WinXP                      264        343824   0.08%     0.02%          33         918   3.59%   0.62%
9480-1-15151-1     Safari157-W7x32                      264        316224   0.08%     0.02%          33         907   3.64%   0.54%
234-1-14351-1      OfficePro2003-WinXP                 1832        656354   0.28%     0.05%          95        2801   3.39%   0.47%
9480-1-14351-1     OfficePro2003-W7x32                 1832       1090216   0.17%     0.03%          95        3800   2.50%   0.45%
9480-2-14351-1     OfficePro2003-W7x64                 1832       1077126   0.17%     0.03%          95        3804   2.50%   0.42%
9480-2-15151-1     Safari157-W7x64                      266        569645   0.05%     0.01%          37        1504   2.46%   0.39%
9480-1-14417-1     Wireshark-W7x32                       46        171515   0.03%     0.01%          11         617   1.78%   0.10%
9480-1-15146-1     eraser-W7x32                          51         69984   0.07%     0.07%           3          24  12.50%   0.02%
9480-2-14416-1     Wireshark-W7x64                       34        209666   0.02%     0.01%           8         611   1.31%   0.02%
234-1-15489-1      InvisibleSecrets21-WinXP               0          6689   0.00%     0.00%           0          19   0.00%   0.00%
9480-1-15141-1     UPX-W7x32                              0          1796   0.00%     0.00%           0          19   0.00%   0.00%
9480-1-15142-1     sdelete-W7x32                          0           642   0.00%     0.00%           0           5   0.00%   0.00%
9480-2-15141-1     UPX-W7x64                              0          1813   0.00%     0.00%           0          19   0.00%   0.00%
9480-2-15142-1     sdelete-W7x64                          0           642   0.00%     0.00%           0           4   0.00%   0.00%
```

Future research will extend our approach and apply it to other domains

- Extensions of this work:
  - enhance computation
  - sector differencing
  - instrumented collection
  - noise reduction at collection
- Apply approach to malware
- Apply approach to mobile platforms
- Model artifact persistence
- Apply to memory artifacts

Questions?