RSA CONFERENCE 2014
FEBRUARY 24 – 28 | MOSCONE CENTER | SAN FRANCISCO

Share.
Learn.
Secure.
Capitalizing on
Collective Intelligence

How to Catch an Insider Data Thief

SESSION ID: HUM-R03

Jonathan Grier

Principal
Grier Forensics
jdgrier@grierforensics.com

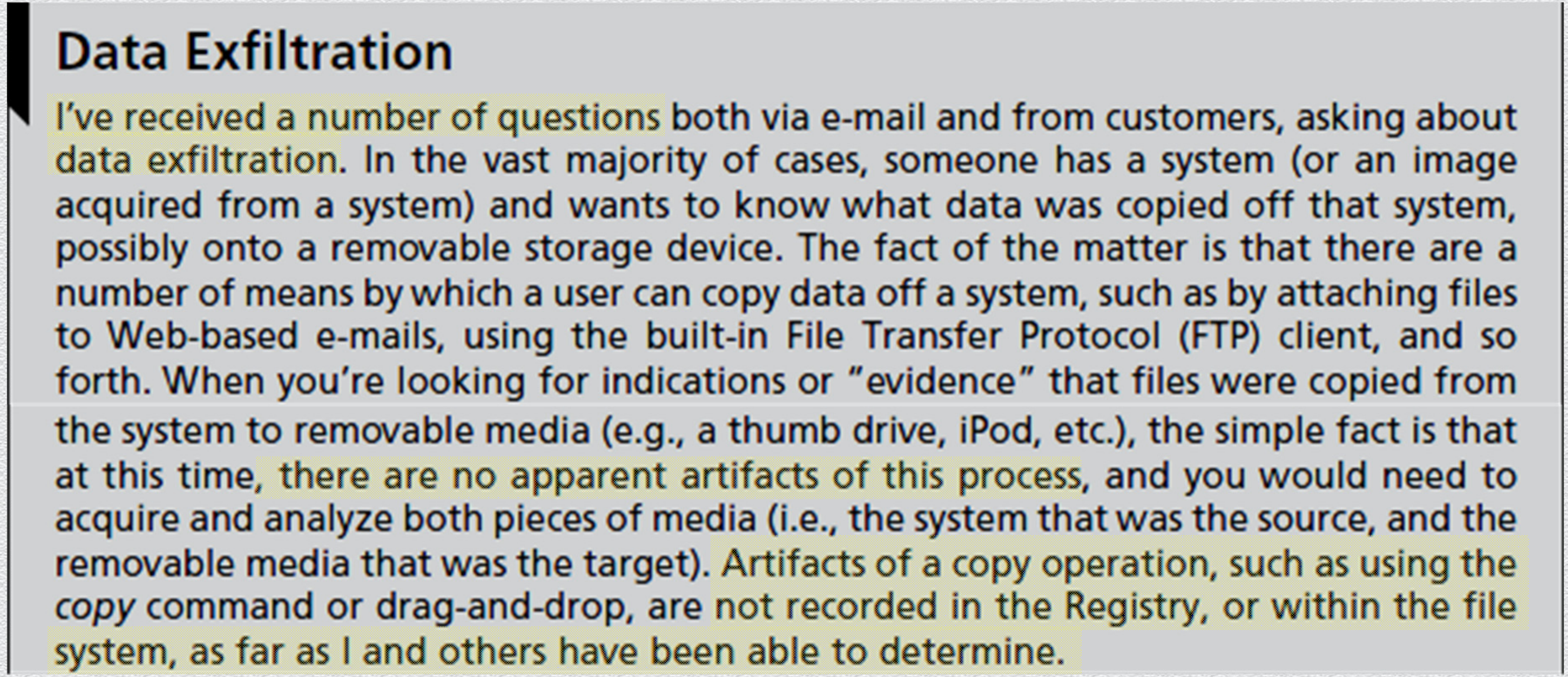
Concerning Confidentiality

To preserve client confidentiality, case information (names, places, dates and settings) has been omitted or altered.

The data and techniques presented have not been modified.

Part I: Understanding

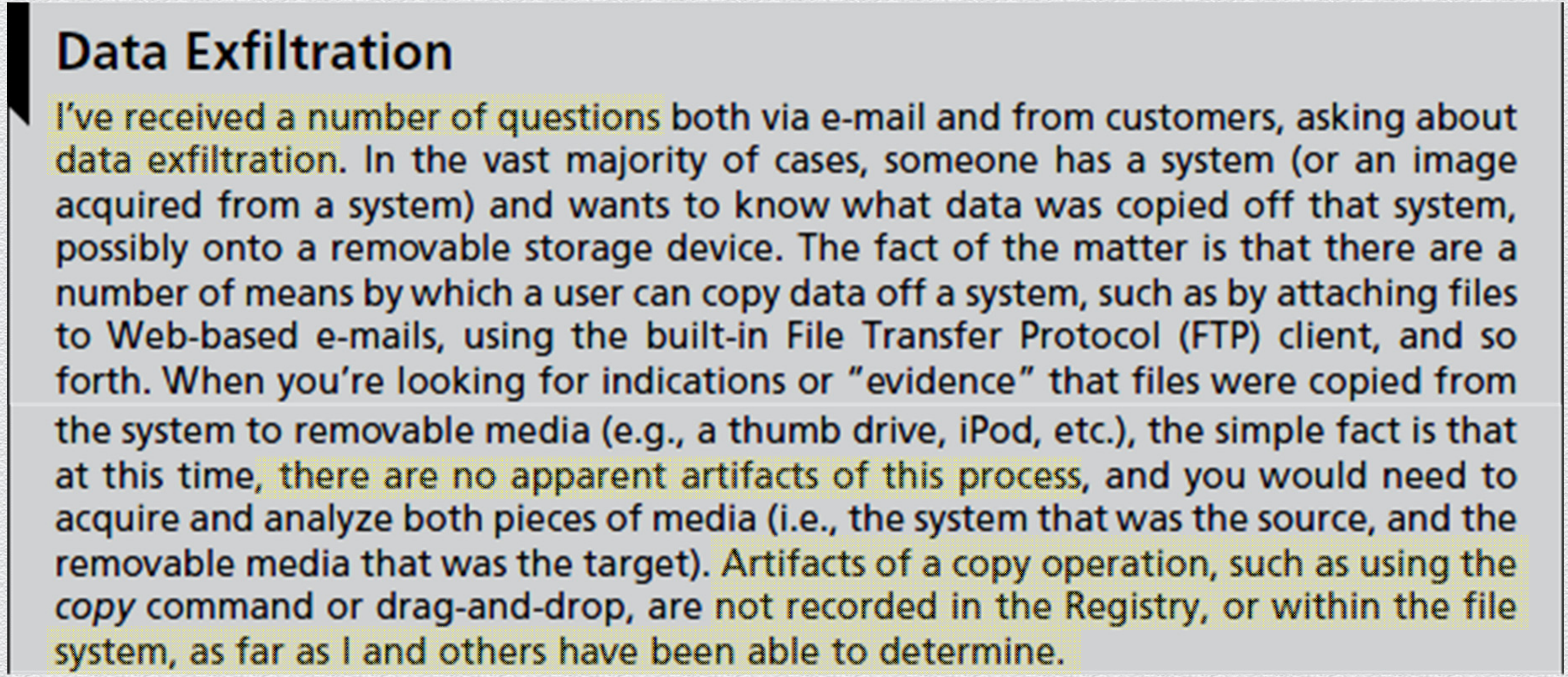
Can you find the data thief?



Data Exfiltration

I've received a number of questions both via e-mail and from customers, asking about data exfiltration. In the vast majority of cases, someone has a system (or an image acquired from a system) and wants to know what data was copied off that system, possibly onto a removable storage device. The fact of the matter is that there are a number of means by which a user can copy data off a system, such as by attaching files to Web-based e-mails, using the built-in File Transfer Protocol (FTP) client, and so forth. When you're looking for indications or "evidence" that files were copied from the system to removable media (e.g., a thumb drive, iPod, etc.), the simple fact is that at this time, there are no apparent artifacts of this process, and you would need to acquire and analyze both pieces of media (i.e., the system that was the source, and the removable media that was the target). Artifacts of a copy operation, such as using the *copy* command or drag-and-drop, are not recorded in the Registry, or within the file system, as far as I and others have been able to determine.

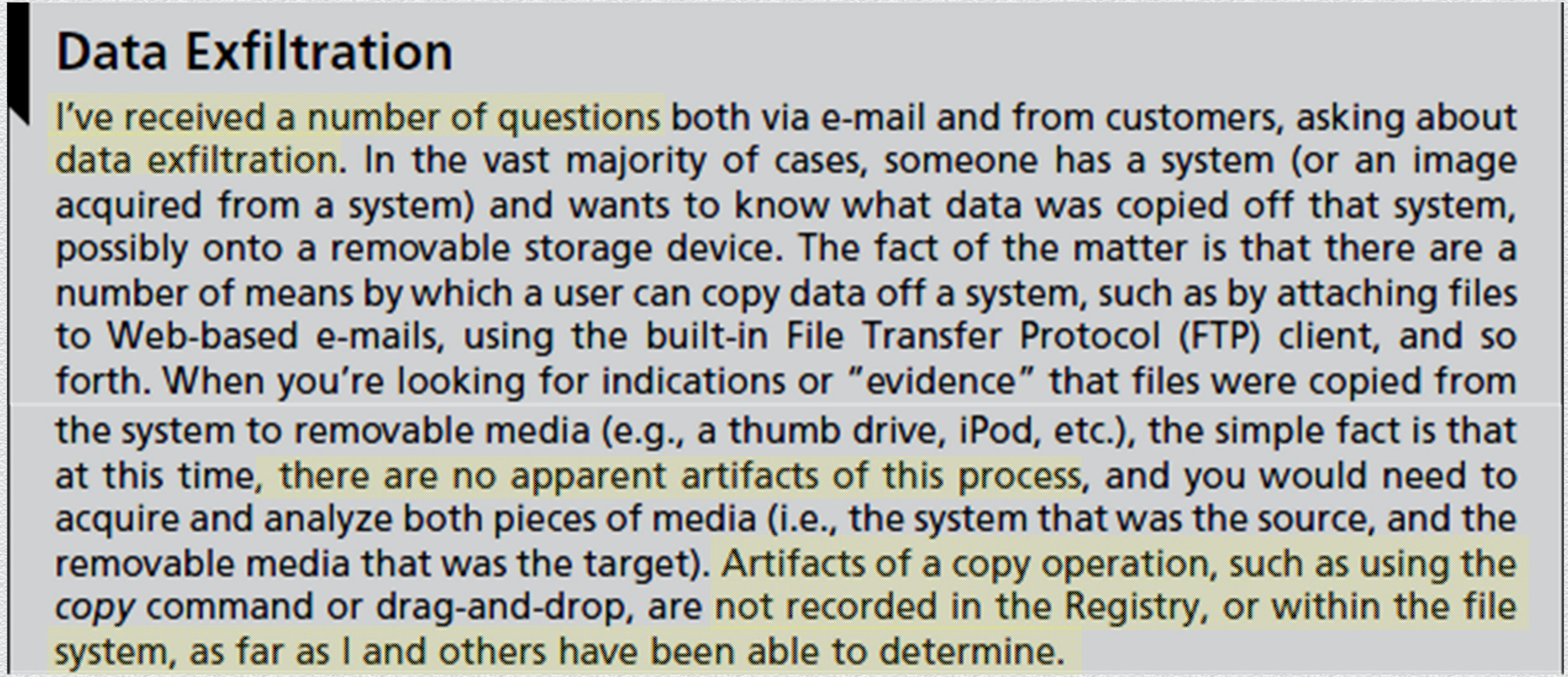
Harlan Carvey, *Windows Forensic Analysis*, 2009



Data Exfiltration

I've received a number of questions both via e-mail and from customers, asking about data exfiltration. In the vast majority of cases, someone has a system (or an image acquired from a system) and wants to know what data was copied off that system, possibly onto a removable storage device. The fact of the matter is that there are a number of means by which a user can copy data off a system, such as by attaching files to Web-based e-mails, using the built-in File Transfer Protocol (FTP) client, and so forth. When you're looking for indications or "evidence" that files were copied from the system to removable media (e.g., a thumb drive, iPod, etc.), the simple fact is that at this time, there are no apparent artifacts of this process, and you would need to acquire and analyze both pieces of media (i.e., the system that was the source, and the removable media that was the target). Artifacts of a copy operation, such as using the *copy* command or drag-and-drop, are not recorded in the Registry, or within the file system, as far as I and others have been able to determine.

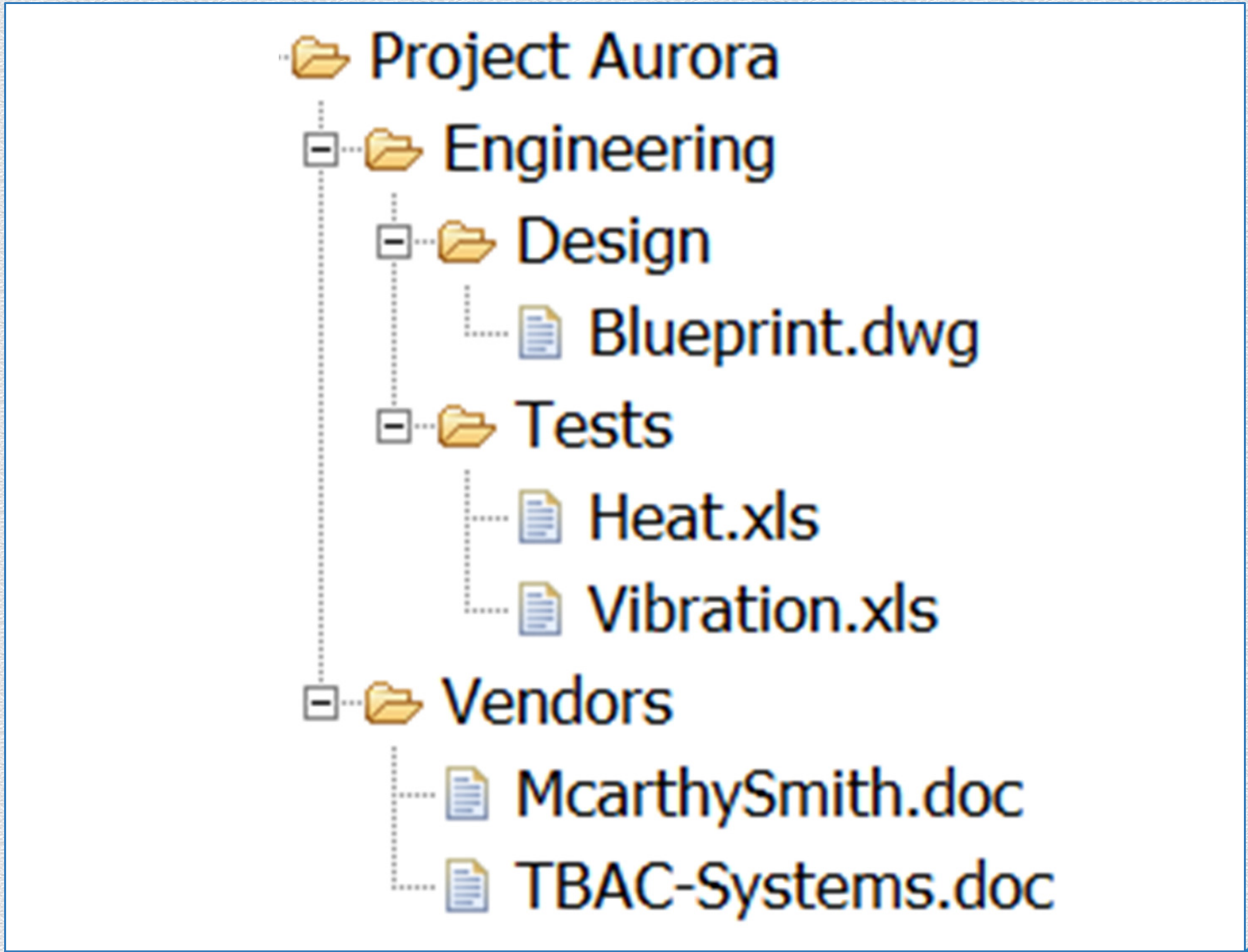
No Artifacts = No Forensics



Data Exfiltration

I've received a number of questions both via e-mail and from customers, asking about data exfiltration. In the vast majority of cases, someone has a system (or an image acquired from a system) and wants to know what data was copied off that system, possibly onto a removable storage device. The fact of the matter is that there are a number of means by which a user can copy data off a system, such as by attaching files to Web-based e-mails, using the built-in File Transfer Protocol (FTP) client, and so forth. When you're looking for indications or "evidence" that files were copied from the system to removable media (e.g., a thumb drive, iPod, etc.), the simple fact is that at this time, there are no apparent artifacts of this process, and you would need to acquire and analyze both pieces of media (i.e., the system that was the source, and the removable media that was the target). Artifacts of a copy operation, such as using the *copy* command or drag-and-drop, are not recorded in the Registry, or within the file system, as far as I and others have been able to determine.

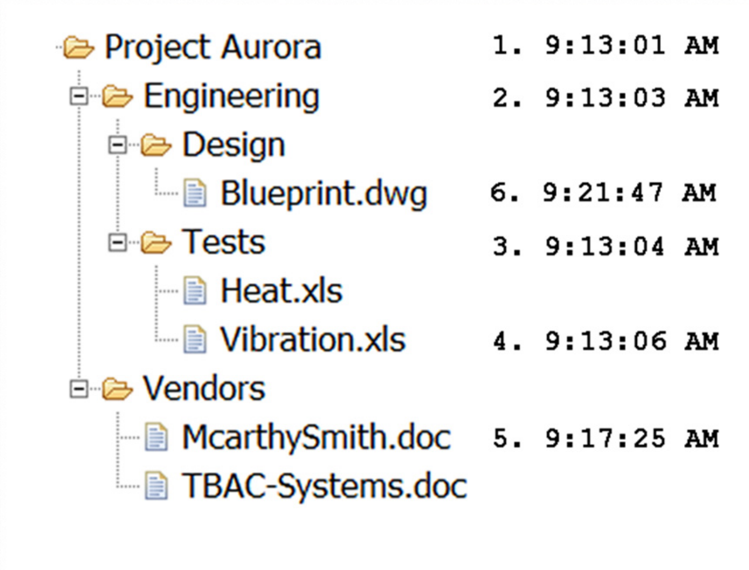
No Artifacts = No Forensics ???



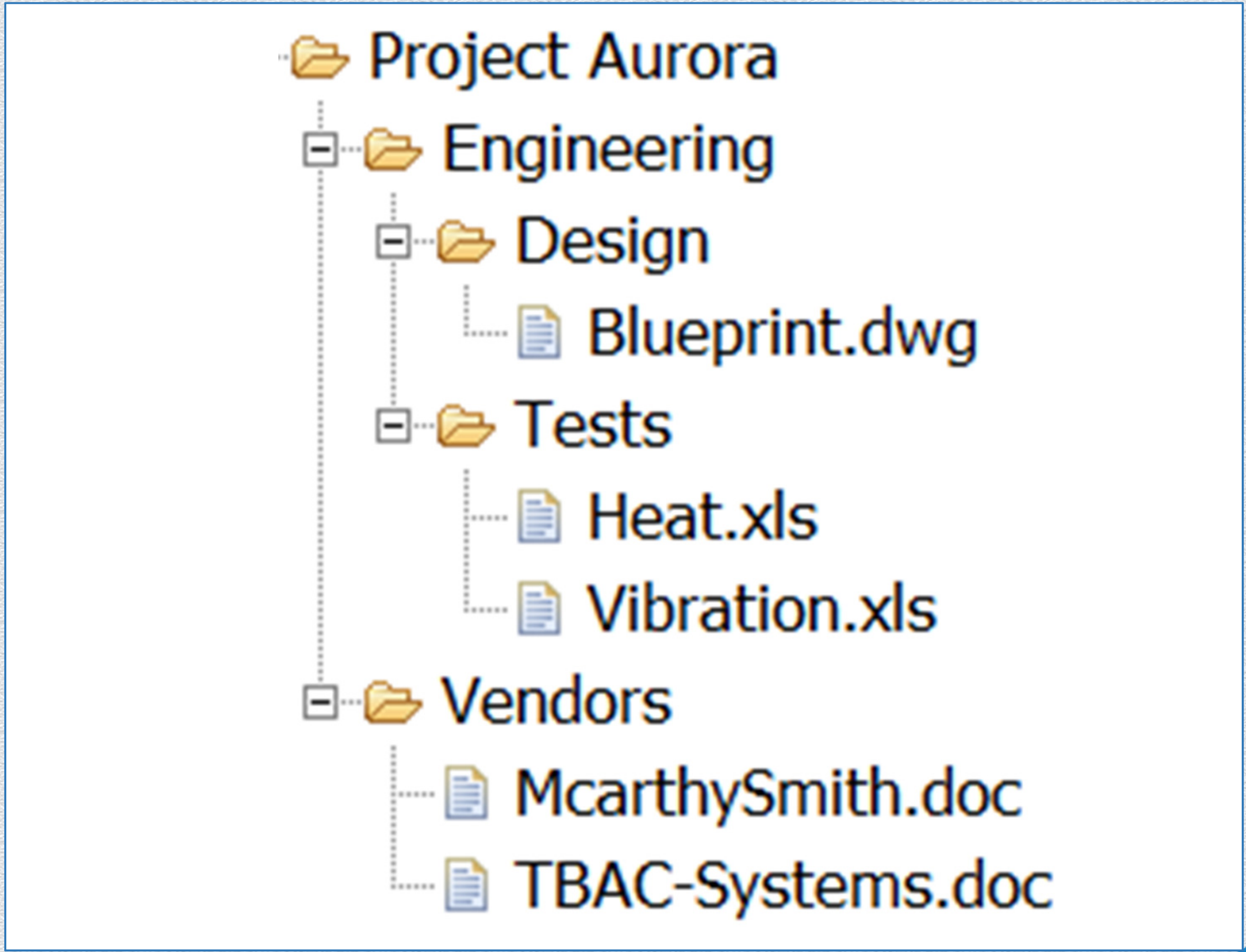
- Project Aurora
  - Engineering
    - Design
      - Blueprint.dwg
    - Tests
      - Heat.xls
      - Vibration.xls
  - Vendors
    - McarthySmith.doc
    - TBAC-Systems.doc

Access timestamps updates during:

Routine access



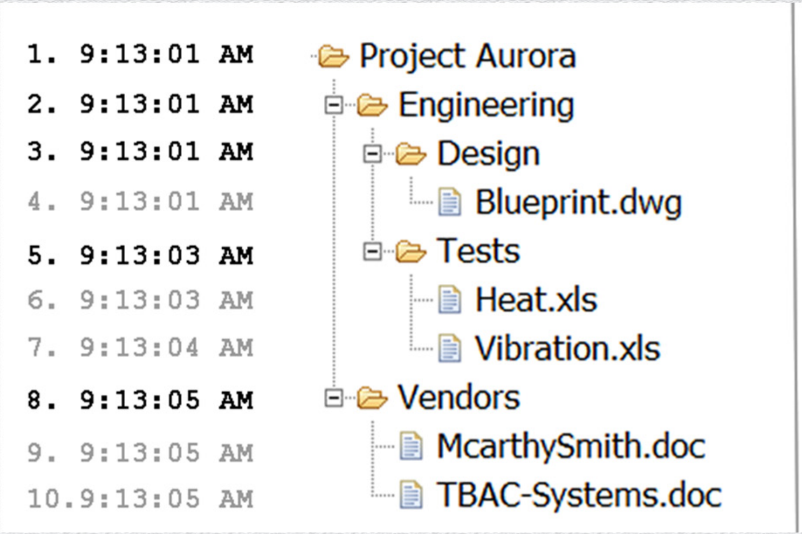
```
Project Aurora              1. 9:13:01 AM
  Engineering               2. 9:13:03 AM
    Design
      Blueprint.dwg         6. 9:21:47 AM
    Tests                   3. 9:13:04 AM
      Heat.xls
      Vibration.xls         4. 9:13:06 AM
  Vendors
    McarthySmith.doc        5. 9:17:25 AM
    TBAC-Systems.doc
```



- Project Aurora
  - Engineering
    - Design
      - Blueprint.dwg
    - Tests
      - Heat.xls
      - Vibration.xls
  - Vendors
    - McarthySmith.doc
    - TBAC-Systems.doc

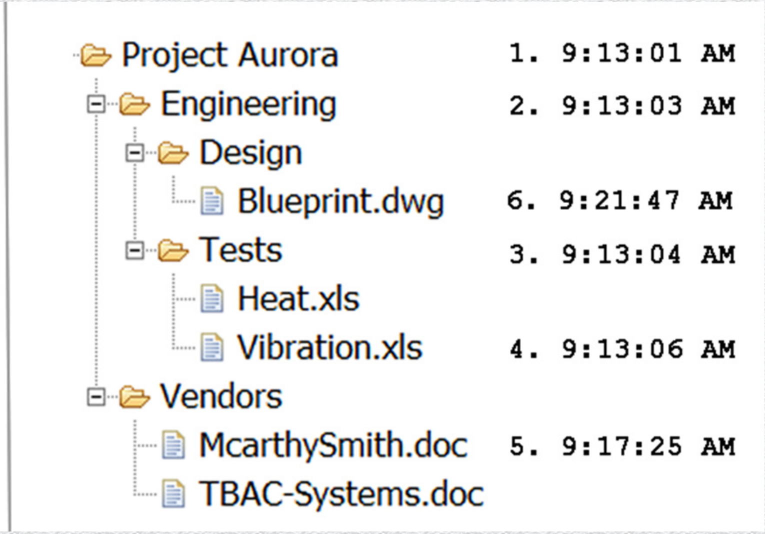
Access timestamps updates during:

Copying a folder



```
1. 9:13:01 AM     Project Aurora
2. 9:13:01 AM       Engineering
3. 9:13:01 AM         Design
4. 9:13:01 AM           Blueprint.dwg
5. 9:13:03 AM         Tests
6. 9:13:03 AM           Heat.xls
7. 9:13:04 AM           Vibration.xls
8. 9:13:05 AM       Vendors
9. 9:13:05 AM         McarthySmith.doc
10.9:13:05 AM         TBAC-Systems.doc
```

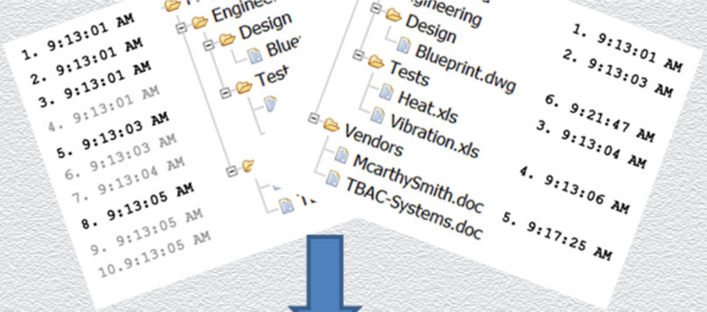
Routine access



```
Project Aurora                1. 9:13:01 AM
  Engineering                 2. 9:13:03 AM
    Design
      Blueprint.dwg           6. 9:21:47 AM
    Tests                     3. 9:13:04 AM
      Heat.xls
      Vibration.xls           4. 9:13:06 AM
  Vendors
    McarthySmith.doc          5. 9:17:25 AM
    TBAC-Systems.doc
```

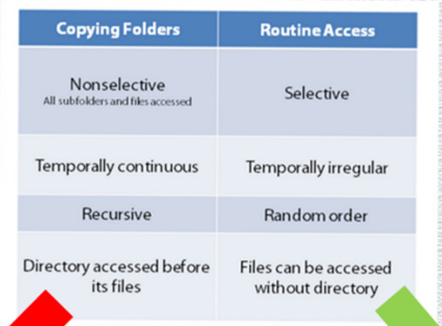
Emergent properties

Copying Folders	Routine Access
Nonselective	Selective
Temporally continuous	Temporally irregular
Recursive	Random order
Directory accessed before its files	File can be accessed without directory

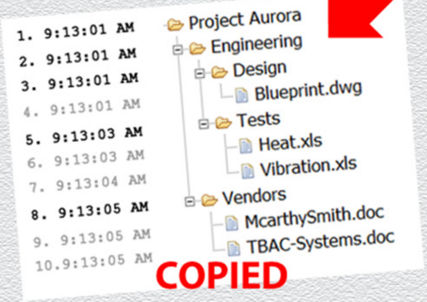


```
1. 9:13:01 AM
2. 9:13:01 AM
3. 9:13:01 AM
4. 9:13:01 AM
5. 9:13:03 AM
6. 9:13:03 AM
7. 9:13:04 AM
8. 9:13:05 AM
9. 9:13:05 AM
10.9:13:05 AM
```

```
gineering
  Design
    Blueprint.dwg      1. 9:13:01 AM
                       2. 9:13:03 AM
  Tests
    Heat.xls           6. 9:21:47 AM
    Vibration.xls      3. 9:13:04 AM
Vendors
  McarthySmith.doc     4. 9:13:06 AM
  TBAC-Systems.doc     5. 9:17:25 AM
```

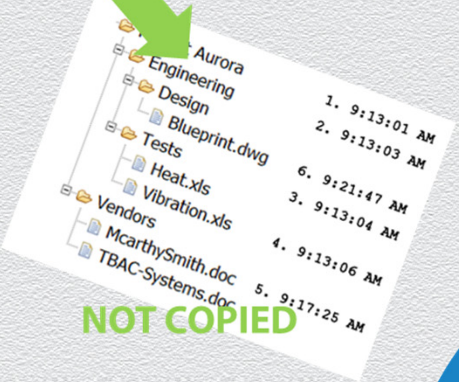


Copying Folders	Routine Access
Nonselective All subfolders and files accessed	Selective
Temporally continuous	Temporally irregular
Recursive	Random order
Directory accessed before its files	Files can be accessed without directory



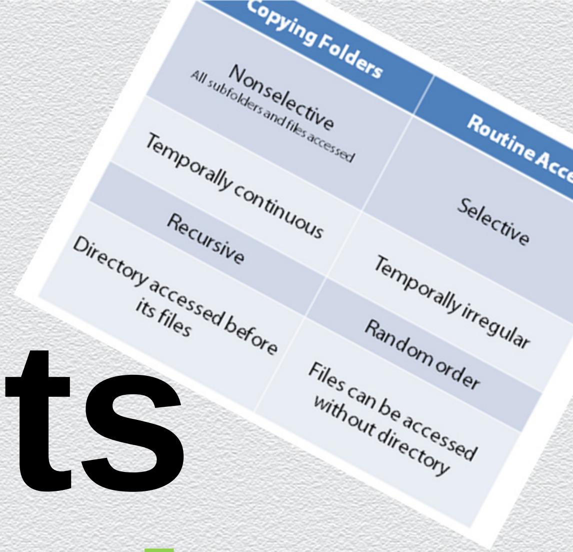
```
1. 9:13:01 AM     Project Aurora
2. 9:13:01 AM       Engineering
3. 9:13:01 AM         Design
4. 9:13:01 AM           Blueprint.dwg
5. 9:13:03 AM         Tests
6. 9:13:03 AM           Heat.xls
7. 9:13:04 AM           Vibration.xls
8. 9:13:05 AM       Vendors
9. 9:13:05 AM         McarthySmith.doc
10.9:13:05 AM         TBAC-Systems.doc
```

COPIED



```
  Aurora
    Engineering
      Design
        Blueprint.dwg      1. 9:13:01 AM
                           2. 9:13:03 AM
      Tests
        Heat.xls           6. 9:21:47 AM
        Vibration.xls      3. 9:13:04 AM
    Vendors
      McarthySmith.doc     4. 9:13:06 AM
      TBAC-Systems.doc     5. 9:17:25 AM
```

NOT COPIED



Copying Folders	Routine Acce
Nonselective All subfolders and files accessed	Selective
Temporally continuous	Temporally irregular
Recursive	Random order
Directory accessed before its files	Files can be accessed without directory

No Artifacts

Yes Forensics

"slap-your-head-and-say-'doh-wish-I'd-thought-of-that"

-- an anonymous colleague

Not so fast...

1. Timestamps are overwritten very quickly
2. There are other nonselective, recursive activities (besides copying)

Not so fast...

1. Timestamps are overwritten very quickly

Can we use this methods months later?

On a heavily used system?

Won’t most of the timestamps have been overwritten?

Not so fast...

1. Timestamps are overwritten very quickly

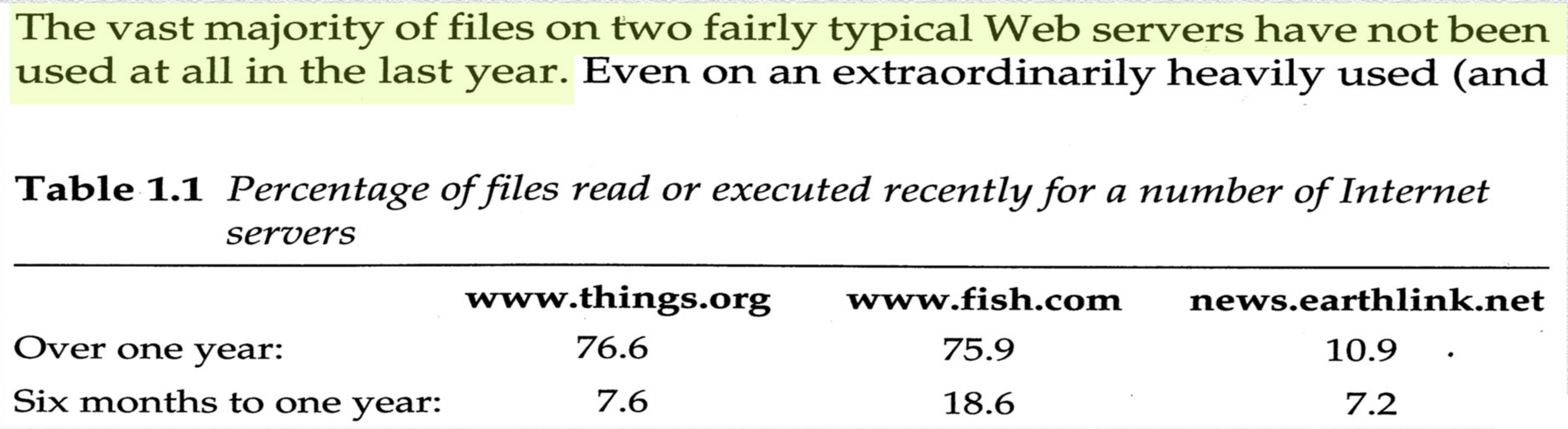
YES! **Can we use this methods months later?**

YES! **On a heavily used system?**

Not really! **Won’t most of the timestamps have been overwritten?**

Two observations

- 1. Timestamp values can *increase*, but never *decrease*.
- 2. A lot of files just collect dust. Most activity is on a minority of files.

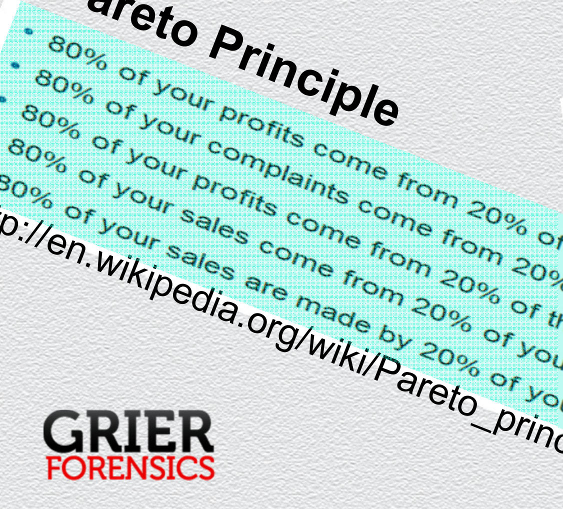


The vast majority of files on two fairly typical Web servers have not been used at all in the last year. Even on an extraordinarily heavily used (and

Table 1.1 *Percentage of files read or executed recently for a number of Internet servers*

	www.things.org	www.fish.com	news.earthlink.net
Over one year:	76.6	75.9	10.9
Six months to one year:	7.6	18.6	7.2

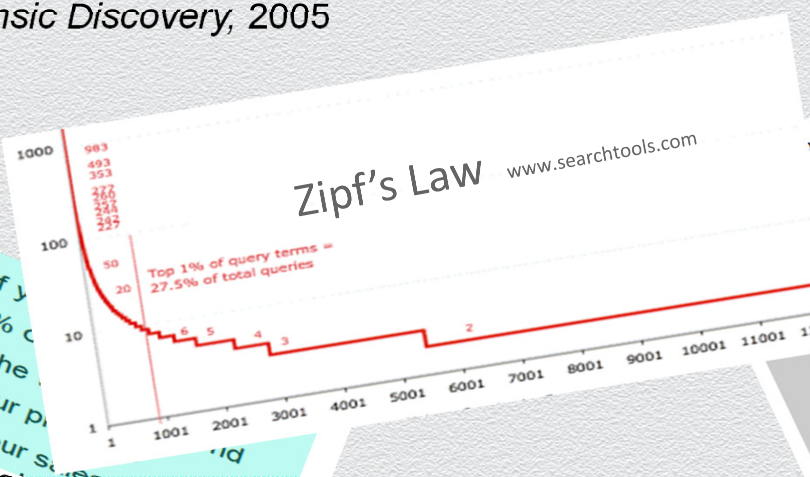
Farmer & Venema, *Forensic Discovery,* 2005



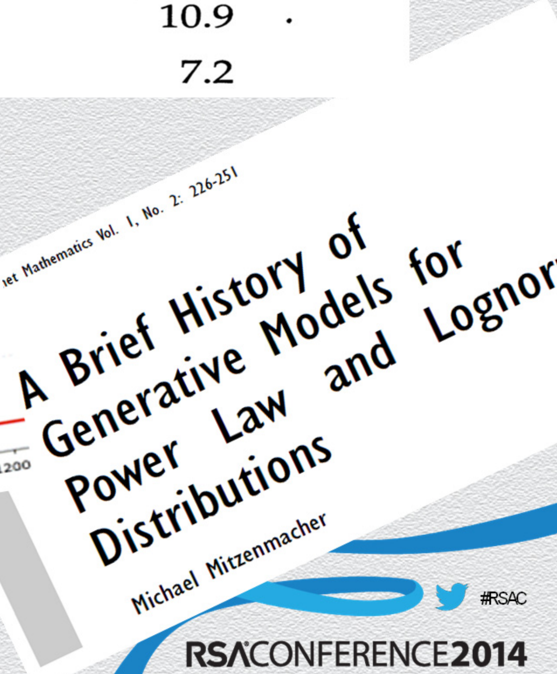
Pareto Principle

- 80% of your profits come from 20% of y
- 80% of your complaints come from 20% c
- 80% of your profits come from 20% of the
- 80% of your sales come from 20% of your p
- 80% of your sales are made by 20% of your s es staff[10]

p://en.wikipedia.org/wiki/Pareto_principle



Zipf's Law www.searchtools.com



et Mathematics Vol. 1, No. 2: 226-251

A Brief History of Generative Models for Power Law and Lognor Distributions

Michael Mitzenmacher

At $t_{copying}$:

- All files have access_timestamp = $t_{copying}$

At $t_{copying}$:

- All files have access_timestamp = $t_{copying}$

Several weeks later:

- All files have access_timestamp ≥ $t_{copying}$

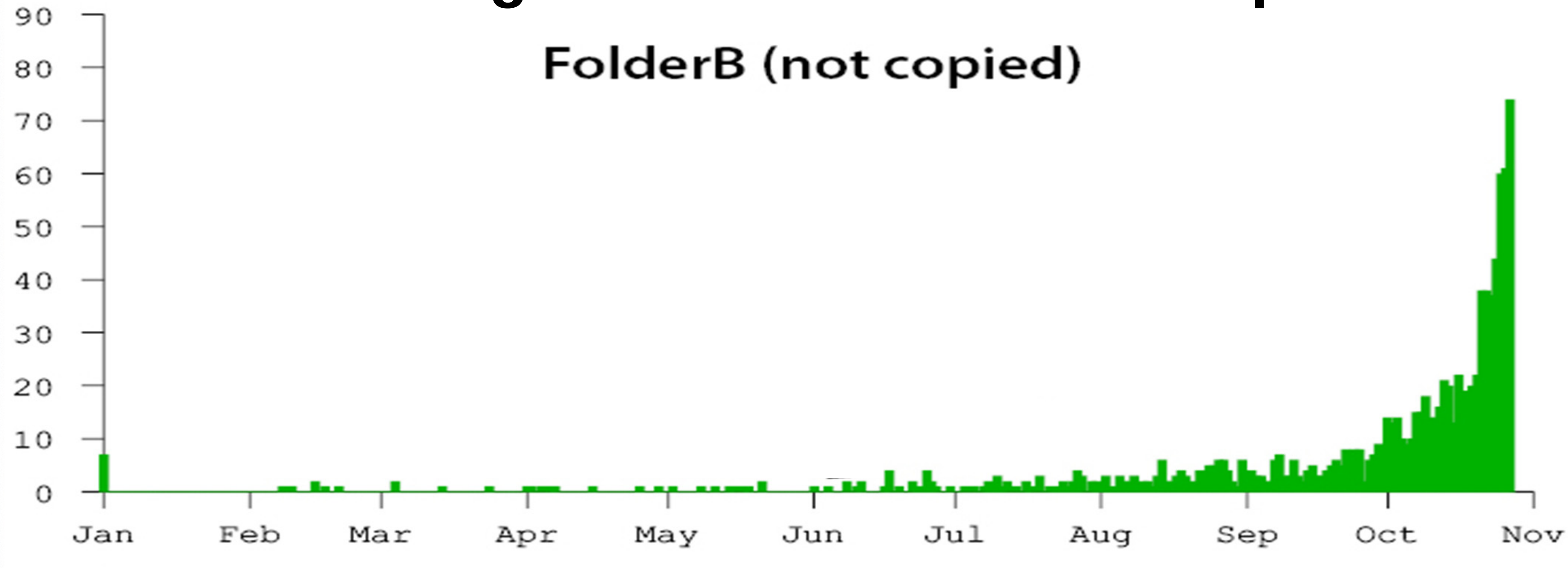
At $t_{copying}$:

- All files have access_timestamp = $t_{copying}$

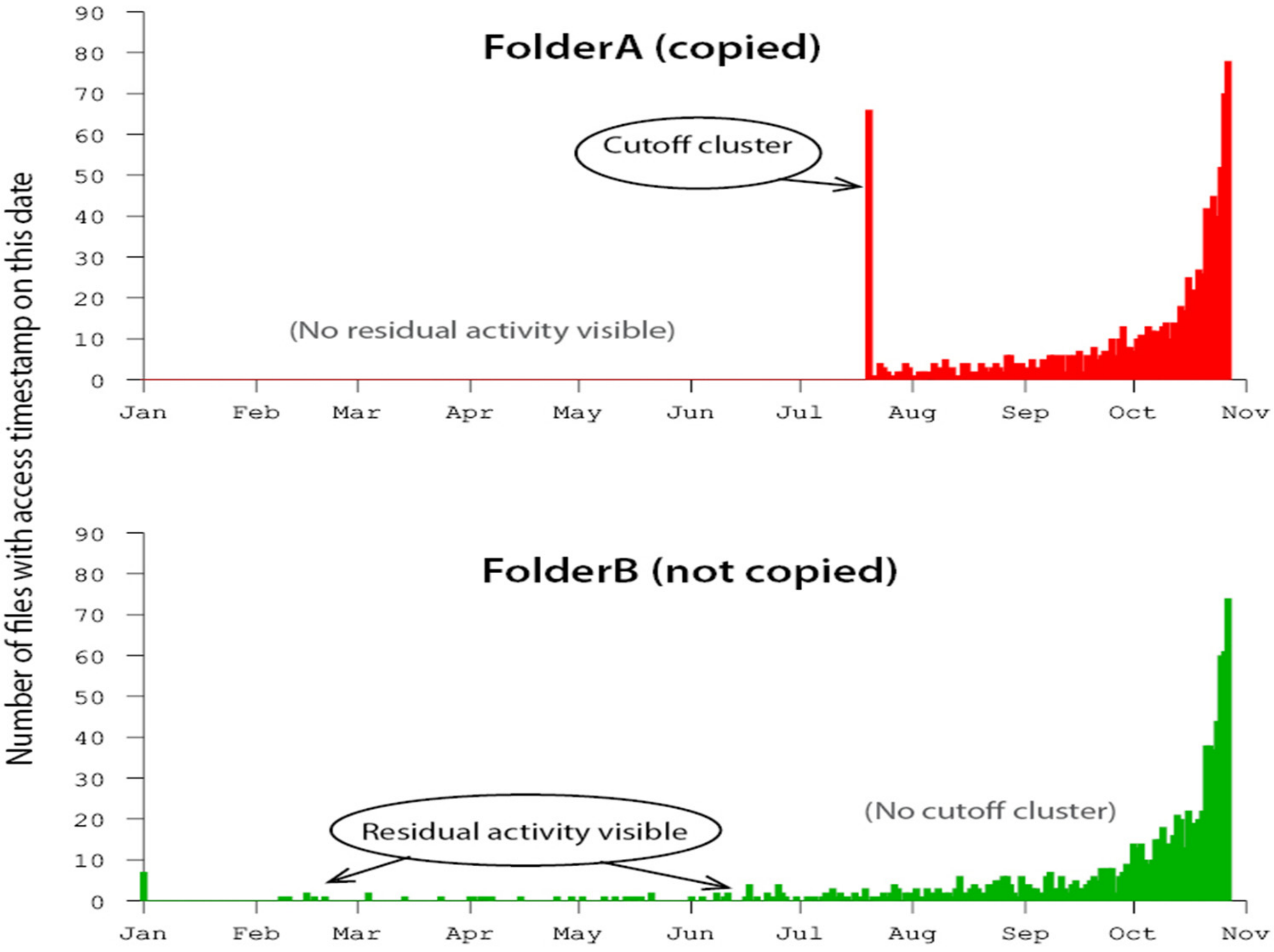
Several weeks later:

- All files have access_timestamp ≥ $t_{copying}$
- **Many** files still have access_timestamp **=** $t_{copying}$

Histogram of access timestamps



After 300 days of simulated activity



FolderA (copied)

Cutoff cluster

(No residual activity visible)

FolderB (not copied)

Residual activity visible

(No cutoff cluster)

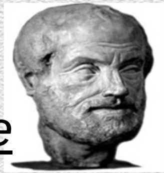
Copying creates a

cutoff cluster

cutoff – No file has timestamp < $t_{cluster}$

cluster – Many files have timestamp = $t_{cluster}$

Aren’t there other recursive access patterns besides copying?



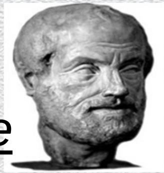
Affirming the consequent

A ⟶ B doesn’t prove B ⟶ A.

The *absence* of a cutoff cluster can disprove copying, but the *existence* can’t prove copying.

Perhaps they ran `grep`.

Indeed, there are!



Affirming the consequent

A ⟶ B doesn't prove B ⟶ A.

The *absence* of a cutoff cluster can disprove copying, but the *existence* can't prove copying.

Perhaps they ran `grep`.

vs.

Abductive reasoning

An unusual observation supports inferring a likely cause.

Who's trying to *prove* anything?

Investigate! One clue leads to another until the case unravels.

Indeed!
Check if `grep` is even installed.
Check why they were still in the building at 11 PM.

An actual investigation:

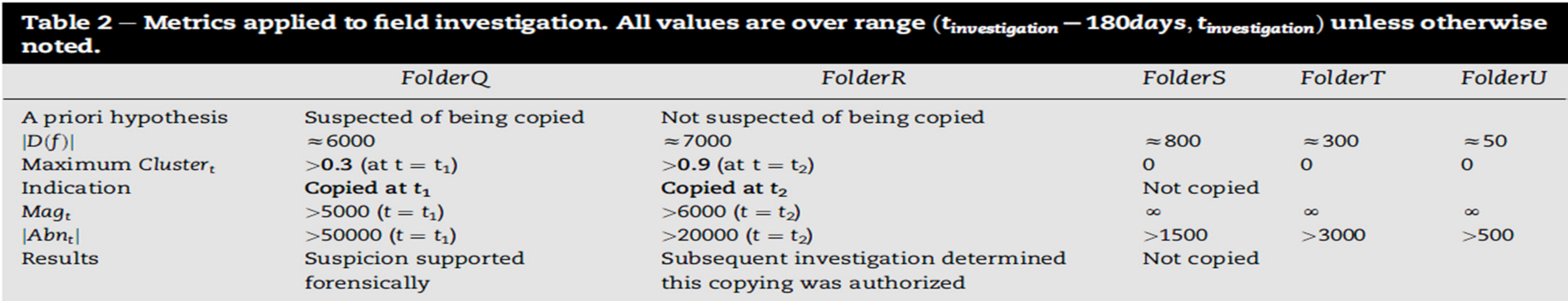
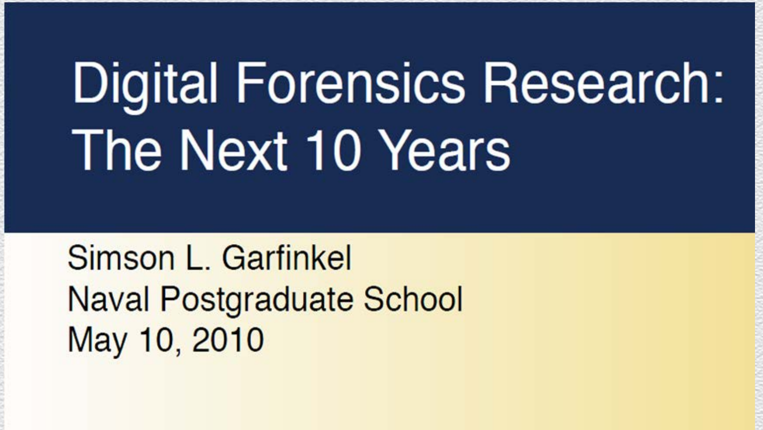


Table 2 — Metrics applied to field investigation. All values are over range ($t_{investigation} - 180 days, t_{investigation}$) unless otherwise noted.

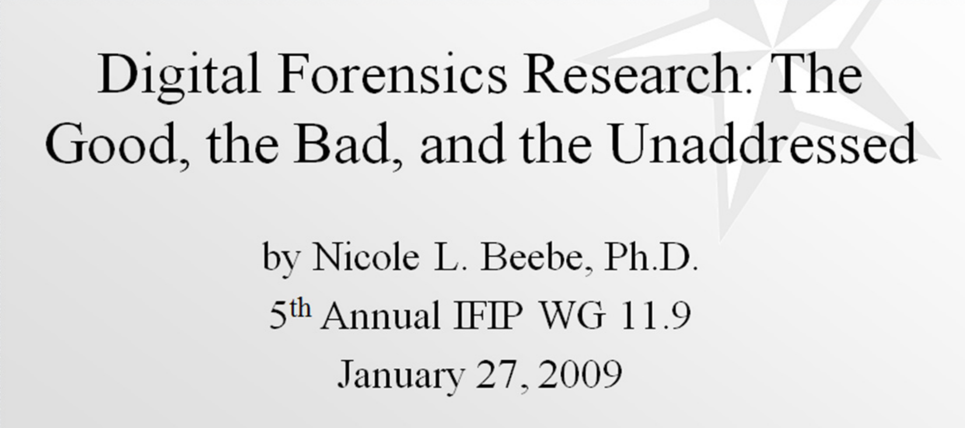
	FolderQ	*FolderR*	*FolderS*	*FolderT*	*FolderU*
A priori hypothesis	Suspected of being copied	Not suspected of being copied			
$\|D(f)\|$	$\approx$6000	$\approx$7000	$\approx$800	$\approx$300	$\approx$50
Maximum $Cluster_t$	$>$**0.3** (at t = t_1)	$>$**0.9** (at t = t_2)	0	0	0
Indication	**Copied at t_1**	**Copied at t_2**	Not copied		
Mag_t	$>$5000 ($t = t_1$)	$>$6000 ($t = t_2$)	∞	∞	∞
$\|Abn_t\|$	$>$50000 ($t = t_1$)	$>$20000 ($t = t_2$)	$>$1500	$>$3000	$>$500
Results	Suspicion supported forensically	Subsequent investigation determined this copying was authorized	Not copied		

Jonathan Grier, *Detecting Data Theft Using Stochastic Forensics*, J. Digital Investigation 2011



Digital Forensics Research: The Next 10 Years

Simson L. Garfinkel
Naval Postgraduate School
May 10, 2010



Digital Forensics Research: The Good, the Bad, and the Unaddressed

by Nicole L. Beebe, Ph.D.
5th Annual IFIP WG 11.9
January 27, 2009

Leading forensic researchers have called to move from:

> “What data can we find?”

To:

> “What did this person do?”

Classical Forensics:

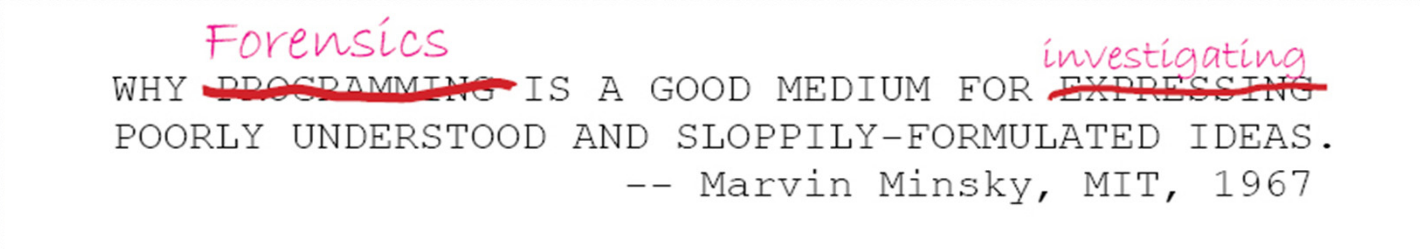
Look at the Surviving Data → Reconstruct Previous Data → This previous data is our deliverable.

Classical Forensics:

Look at the Surviving Data → Reconstruct Previous Data → This previous data is our deliverable.

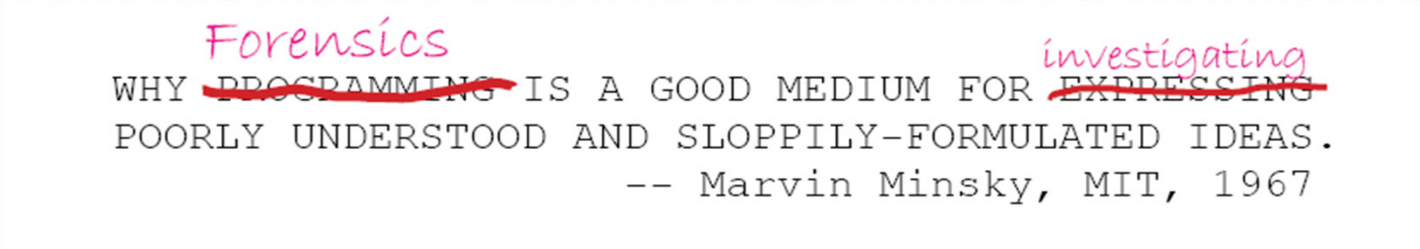
Stochastic Forensics:

What do I want to know about? → What behavior is associated? → How does that behavior affect the system? → Measure those effects. Draw a (quantifiable) inference.



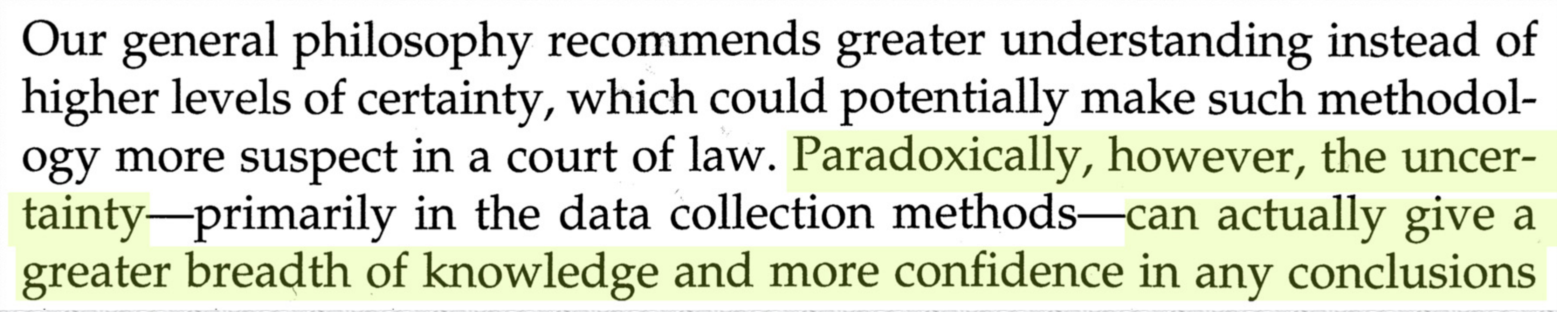
WHY ~~PROGRAMMING~~ Forensics IS A GOOD MEDIUM FOR ~~EXPRESSING~~ investigating POORLY UNDERSTOOD AND SLOPPILY-FORMULATED IDEAS.

-- Marvin Minsky, MIT, 1967



WHY ~~PROGRAMMING~~ Forensics IS A GOOD MEDIUM FOR ~~EXPRESSING~~ investigating POORLY UNDERSTOOD AND SLOPPILY-FORMULATED IDEAS.

-- Marvin Minsky, MIT, 1967



Our general philosophy recommends greater understanding instead of higher levels of certainty, which could potentially make such methodology more suspect in a court of law. Paradoxically, however, the uncertainty—primarily in the data collection methods—can actually give a greater breadth of knowledge and more confidence in any conclusions

Farmer & Venema, *Forensic Discovery,* 2005

Research Questions

1. Delving deeper

   Scientific testing

   Probability value

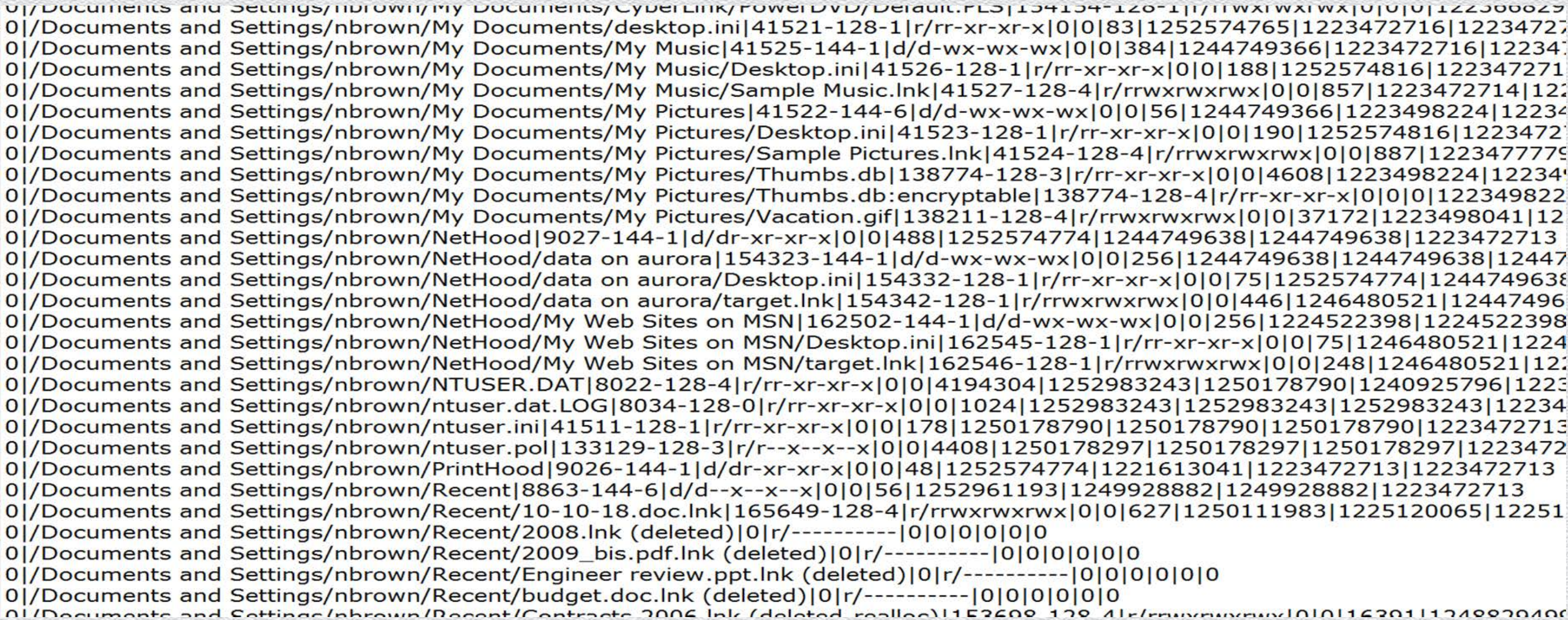
   Automation

2. What other questions can stochastic forensics address?

   Let's find sloppy questions and answer them less precisely!

Part II: Applying Stochastic Forensics

Eyeball?



```
0|/Documents and Settings/nbrown/My Documents/CyberLink/PowerDVD/Default.PLS|154154-128-1|r/rrwxrwxrwx|0|0|0|12236682
0|/Documents and Settings/nbrown/My Documents/desktop.ini|41521-128-1|r/rr-xr-xr-x|0|0|83|1252574765|1223472716|1223472
0|/Documents and Settings/nbrown/My Documents/My Music|41525-144-1|d/d-wx-wx-wx|0|0|384|1244749366|1223472716|122347
0|/Documents and Settings/nbrown/My Documents/My Music/Desktop.ini|41526-128-1|r/rr-xr-xr-x|0|0|188|1252574816|122347271
0|/Documents and Settings/nbrown/My Documents/My Music/Sample Music.lnk|41527-128-4|r/rrwxrwxrwx|0|0|857|1223472714|122
0|/Documents and Settings/nbrown/My Documents/My Pictures|41522-144-6|d/d-wx-wx-wx|0|0|56|1244749366|1223498224|12234
0|/Documents and Settings/nbrown/My Documents/My Pictures/Desktop.ini|41523-128-1|r/rr-xr-xr-x|0|0|190|1252574816|1223472
0|/Documents and Settings/nbrown/My Documents/My Pictures/Sample Pictures.lnk|41524-128-4|r/rrwxrwxrwx|0|0|887|122347777
0|/Documents and Settings/nbrown/My Documents/My Pictures/Thumbs.db|138774-128-3|r/rr-xr-xr-x|0|0|4608|1223498224|122349
0|/Documents and Settings/nbrown/My Documents/My Pictures/Thumbs.db:encryptable|138774-128-4|r/rr-xr-xr-x|0|0|0|122349822
0|/Documents and Settings/nbrown/My Documents/My Pictures/Vacation.gif|138211-128-4|r/rrwxrwxrwx|0|0|37172|1223498041|12
0|/Documents and Settings/nbrown/NetHood|9027-144-1|d/dr-xr-xr-x|0|0|488|1252574774|1244749638|1244749638|1223472713
0|/Documents and Settings/nbrown/NetHood/data on aurora|154323-144-1|d/d-wx-wx-wx|0|0|256|1244749638|1244749638|12447
0|/Documents and Settings/nbrown/NetHood/data on aurora/Desktop.ini|154332-128-1|r/rr-xr-xr-x|0|0|75|1252574774|124474963
0|/Documents and Settings/nbrown/NetHood/data on aurora/target.lnk|154342-128-1|r/rrwxrwxrwx|0|0|446|1246480521|12447496
0|/Documents and Settings/nbrown/NetHood/My Web Sites on MSN|162502-144-1|d/d-wx-wx-wx|0|0|256|1224522398|1224522398
0|/Documents and Settings/nbrown/NetHood/My Web Sites on MSN/Desktop.ini|162545-128-1|r/rr-xr-xr-x|0|0|75|1246480521|1224
0|/Documents and Settings/nbrown/NetHood/My Web Sites on MSN/target.lnk|162546-128-1|r/rrwxrwxrwx|0|0|248|1246480521|122
0|/Documents and Settings/nbrown/NTUSER.DAT|8022-128-4|r/rr-xr-xr-x|0|0|4194304|1252983243|1250178790|1240925796|122
0|/Documents and Settings/nbrown/ntuser.dat.LOG|8034-128-0|r/rr-xr-xr-x|0|0|1024|1252983243|1252983243|1252983243|12234
0|/Documents and Settings/nbrown/ntuser.ini|41511-128-1|r/rr-xr-xr-x|0|0|178|1250178790|1250178790|1250178790|122347271
0|/Documents and Settings/nbrown/ntuser.pol|133129-128-3|r/r--x--x--x|0|0|4408|1250178297|1250178297|1250178297|1223472
0|/Documents and Settings/nbrown/PrintHood|9026-144-1|d/dr-xr-xr-x|0|0|48|1252574774|1221613041|1223472713|1223472713
0|/Documents and Settings/nbrown/Recent|8863-144-6|d/d--x--x--x|0|0|56|1252961193|1249928882|1249928882|1223472713
0|/Documents and Settings/nbrown/Recent/10-10-18.doc.lnk|165649-128-4|r/rrwxrwxrwx|0|0|627|1250111983|1225120065|12251
0|/Documents and Settings/nbrown/Recent/2008.lnk (deleted)|0|r/----------|0|0|0|0|0|0
0|/Documents and Settings/nbrown/Recent/2009_bis.pdf.lnk (deleted)|0|r/----------|0|0|0|0|0|0
0|/Documents and Settings/nbrown/Recent/Engineer review.ppt.lnk (deleted)|0|r/----------|0|0|0|0|0|0
0|/Documents and Settings/nbrown/Recent/budget.doc.lnk (deleted)|0|r/----------|0|0|0|0|0|0
0|/Documents and Settings/nbrown/Recent/Contracts 2006.lnk (deleted-realloc)|153698-128-4|r/rrwxrwxrwx|0|0|16391|124882949
```

Filter & Plot

Filter

- 1. By folder

Filter

-
- 2. Directories versus Files

Filter

- 3. Permissions

Filter

- 4. Other

Filter

- 1. By folder
- 2. Directories versus files
- 3. Permissions
- 4. Other

Plot

Our visual cognition is amazingly robust

Interpret
&
Advance

No Cluster?

Strong evidence of ***no*** copying

Found Cluster?

- 1. Check control folders
- 2. Search for causes

Found Cluster?

- A cluster defines a tight *window of opportunity*
- Use it to *propel the investigation forward*

For more information

- *Detecting Data Theft Using Stochastic Forensics*

  http://www.grierforensics.com/datatheft/Detecting_Data_Theft_Using_Stochastic_Foresnics.pdf
- Digital Forensics Magazine, May 2012
- Ask me! Jonathan Grier, jdgrier@grierforensics.com