

Employee Data Theft Case Study

Concerning Confidentiality

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

The data and techniques presented have not been altered.



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.

Harlan Carvey, *Windows Forensic Analysis*, 2009

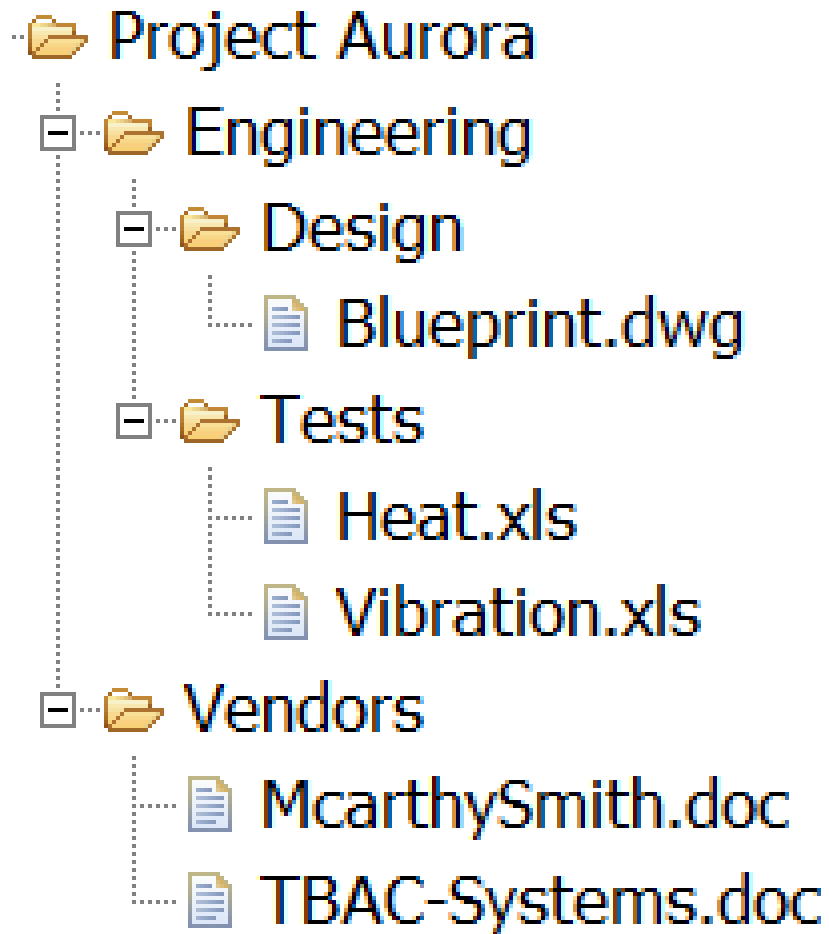
No Artifacts = No Forensics

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.

Harlan Carvey, *Windows Forensic Analysis*, 2009

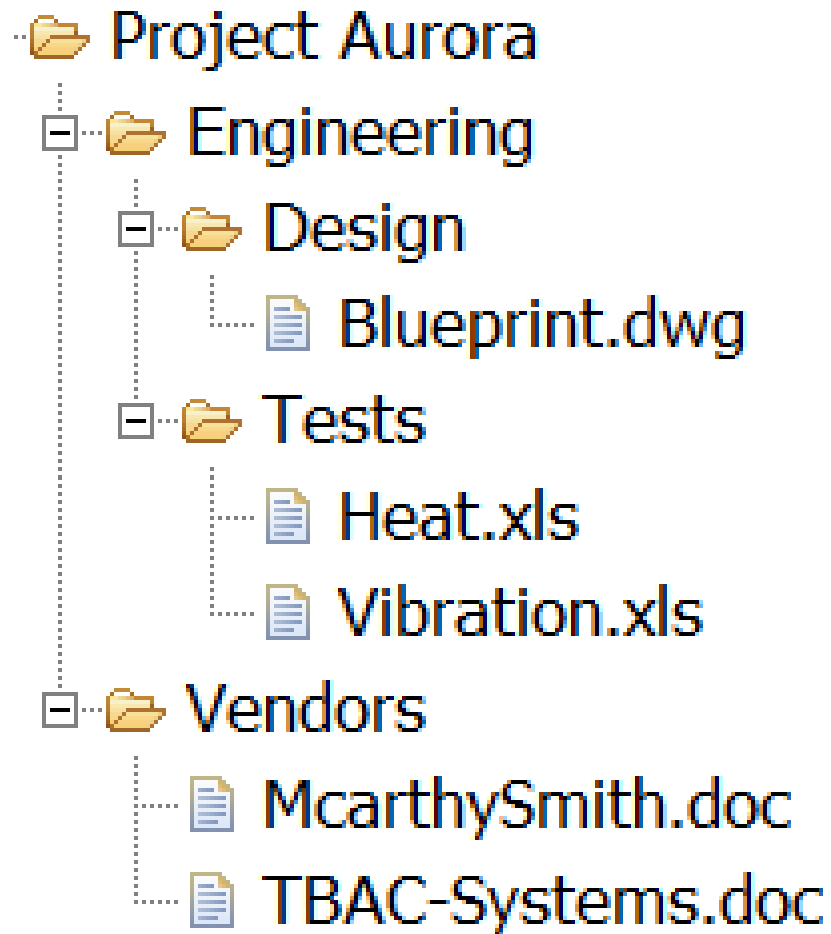
No Artifacts = No Forensics???



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		



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

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

Copying Folders

Routine Access

Nonselective

All subfolders and files accessed

Selective

Temporally continuous

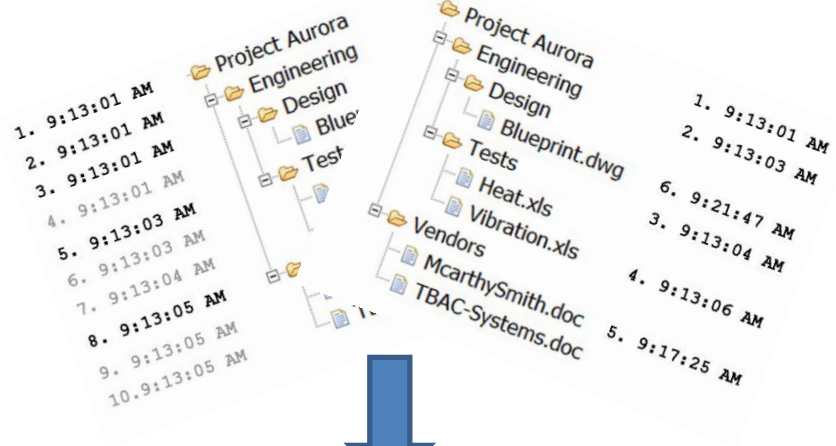
Temporally irregular

Recursive

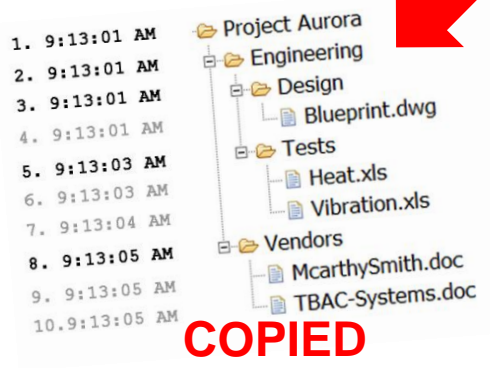
Random order

**Directory accessed
before its files**

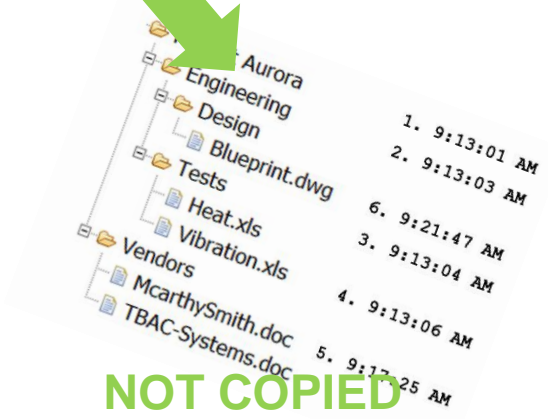
**Files can be accessed
without directory**



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



COPIED



NOT COPIED

No Artifacts Yes Forensics

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

“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 method 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 method months later?

YES!

On a heavily used system?

Not really!

Won't most of the timestamps have been overwritten?

Two observations:

1. Timestamps 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

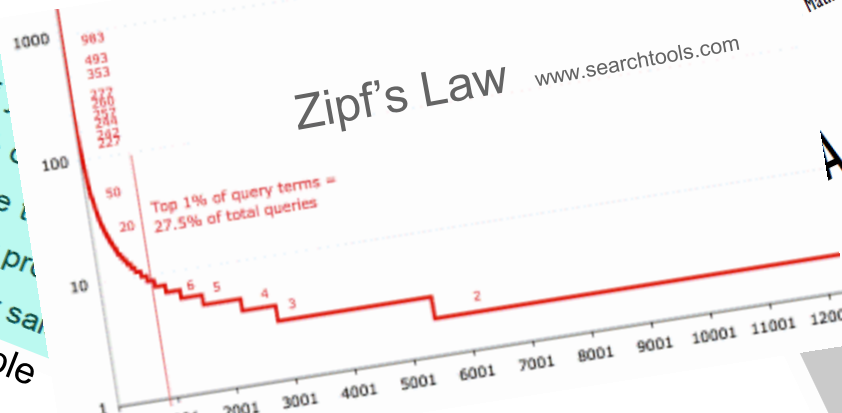
Table 1.1 *Percentage of files read or executed recently for a number of Internet servers*

	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 your customers
 - 80% of your complaints come from 20% of your customers
 - 80% of your profits come from 20% of the products
 - 80% of your sales come from 20% of your salespeople
 - 80% of your sales are made by 20% of your salespeople
- http://en.wikipedia.org/wiki/Pareto_principle



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

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

Mitzenmacher

At t_{copying} :

- All files have `access_timestamp = tcopying`

At t_{copying} :

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

Several weeks later:

- All files have $\text{access_timestamp} \geq t_{\text{copying}}$

At t_{copying} :

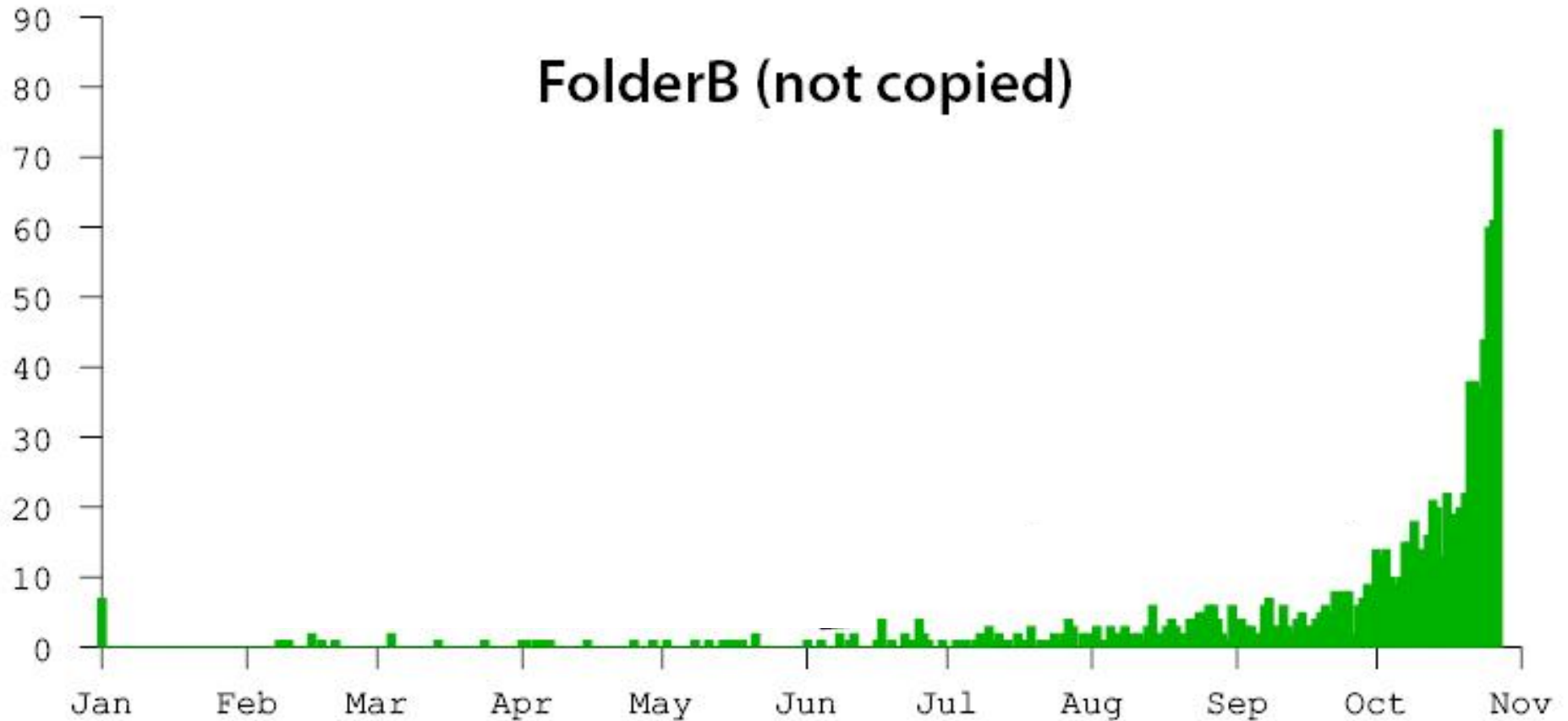
- All files have $\text{access_timestamp} = t_{\text{copying}}$

Several weeks later:

- All files have $\text{access_timestamp} \geq t_{\text{copying}}$
- **Many** files still have $\text{access_timestamp} = t_{\text{copying}}$

Histogram of access timestamps

FolderB (not copied)



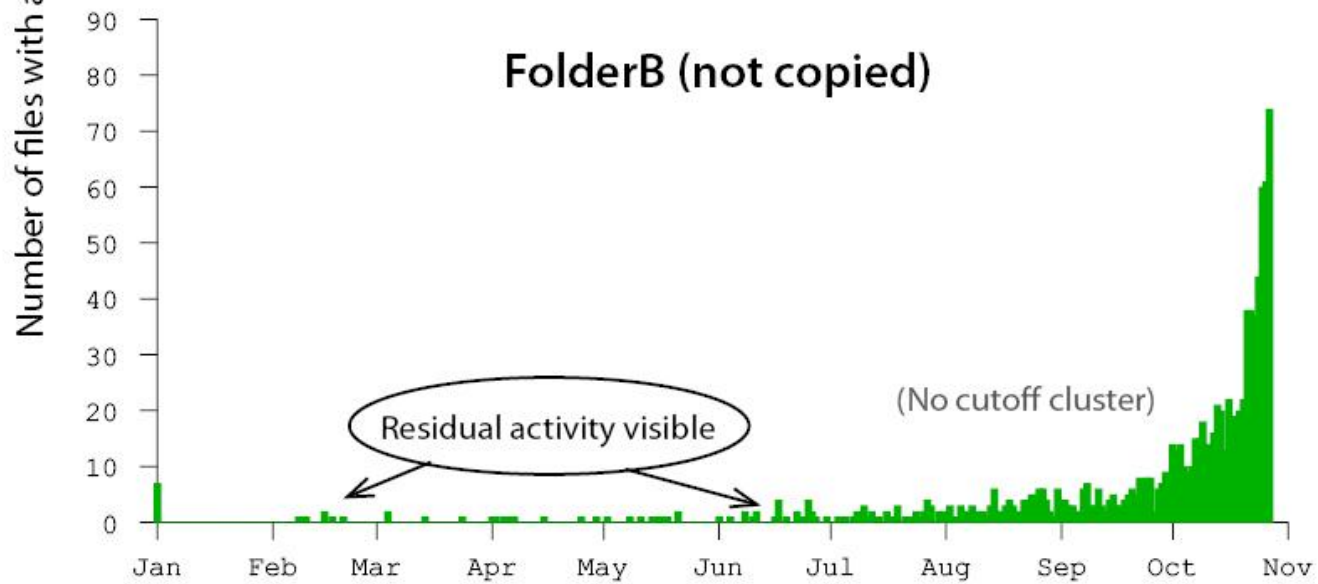
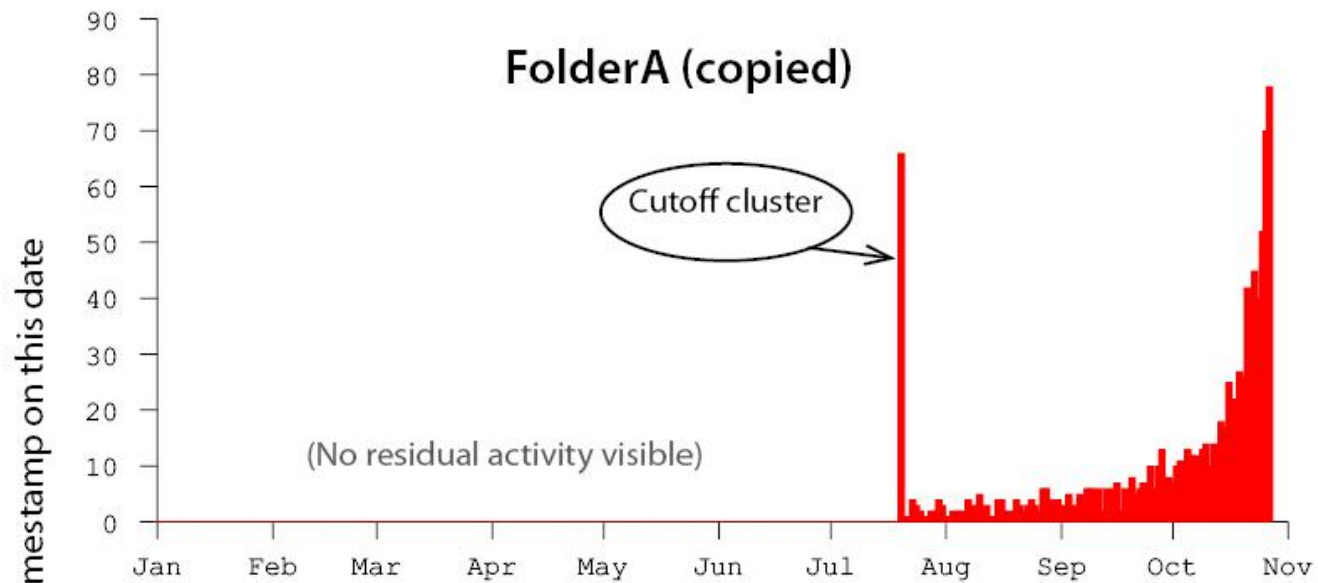
After 300 days of simulated activity

Data from investigation:

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

	FolderQ	FolderR	FolderS	FolderT	FolderU
A priori hypothesis	Suspected of being copied	Not suspected of being copied			
$ D(f) $	≈ 6000	≈ 7000	≈ 800	≈ 300	≈ 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



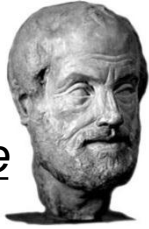
Copying creates a

cutoff cluster

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

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

Aren't there other recursive access patterns besides copying?



Affirming the consequent

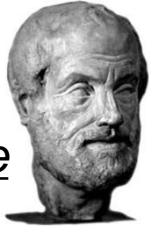
$A \rightarrow B$ doesn't prove $B \rightarrow A$.

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



$A \rightarrow B$ doesn't prove $B \rightarrow A$.

VS.



Abductive reasoning

An unusual observation supports inferring a likely cause.

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

Perhaps they ran `grep`.

Who's trying to *prove* anything?

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

Indeed!

Check if `grep` is installed, if they've ever run it before, or after, on any folder.

Check why they were still in the building at 11 PM.

***Implications
for the field of
forensics...***

Classical Forensics:

Look at the
Surviving Data

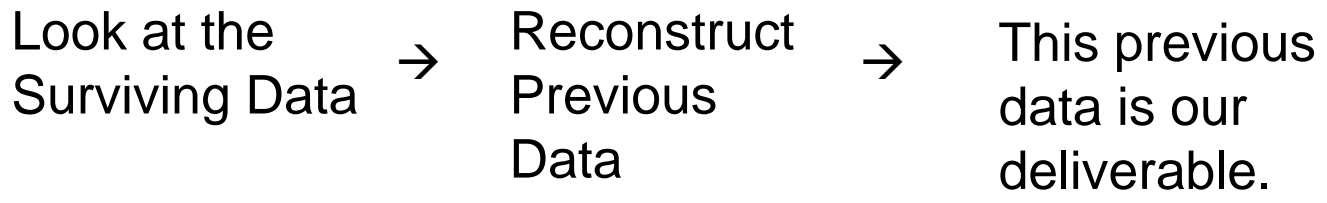


Reconstruct
Previous
Data

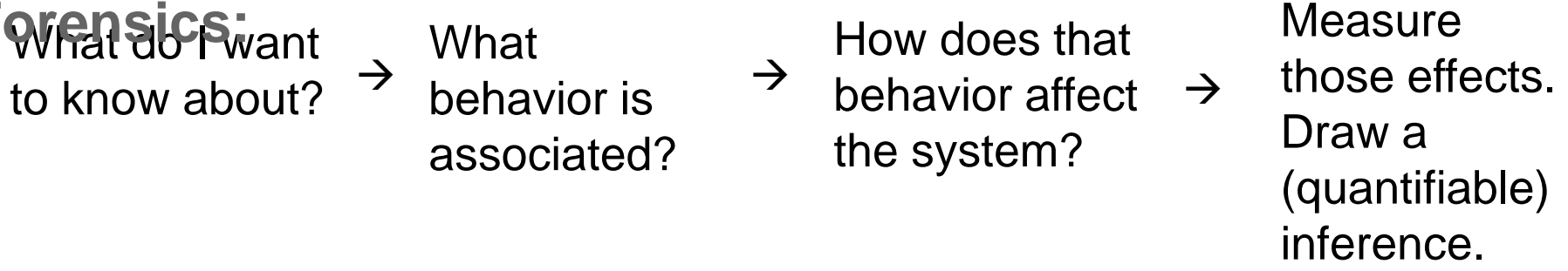


This previous
data is our
deliverable.

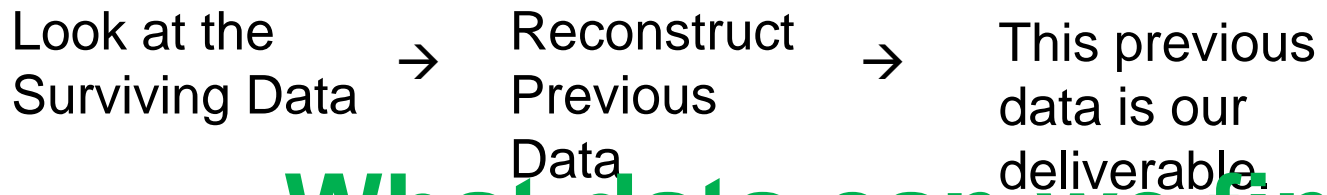
Classical Forensics:



Stochastic Forensics:

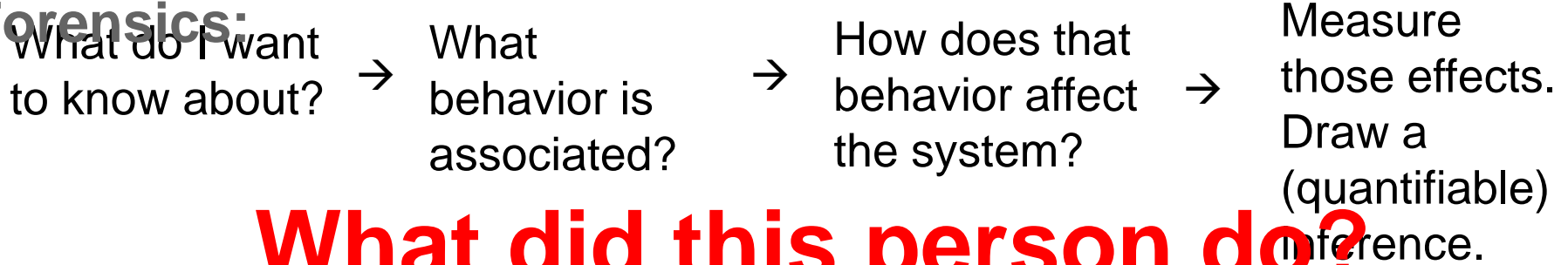


Classical Forensics:



What data can we find?

Stochastic Forensics:



What did this person do?



Lesson Learned:

Forensics doesn't really matter...

Col. John Boyd
Military Strategist
Author, *Patterns of Conflict*

For more information:

- Read my paper

Detecting Data Theft

Using Stochastic Forensics

http://www.grierforensics.com/datatheft/Detecting_Data_Theft_Using_Stochastic_Forensics.pdf

- These slides will be available at

http://www.grierforensics.com/datatheft/Employee_Data_Theft_Case_Study_ACSAC.pdf

- Ask me!

See next slide for my contact info

I'm very interested in hearing
your
feedback, ideas, and questions.

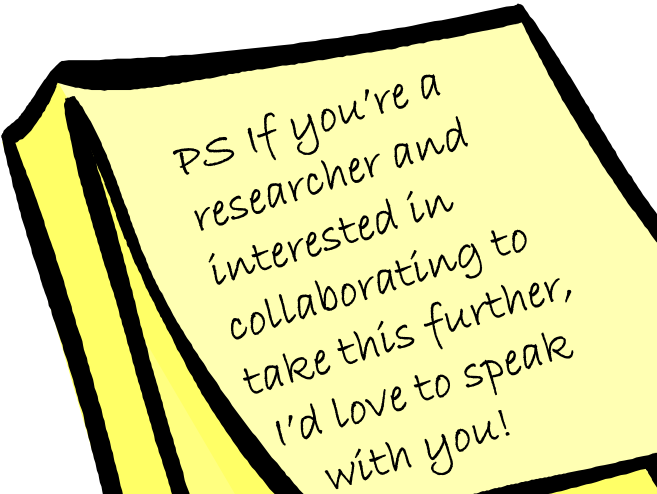
Please share them with me
here at ACSAC.

Or, if we miss each other:

Jonathan Grier

443.501.4044 x1

jdgrier at grierforensics.com



PS If you're a
researcher and
interested in
collaborating to
take this further,
I'd love to speak
with you!