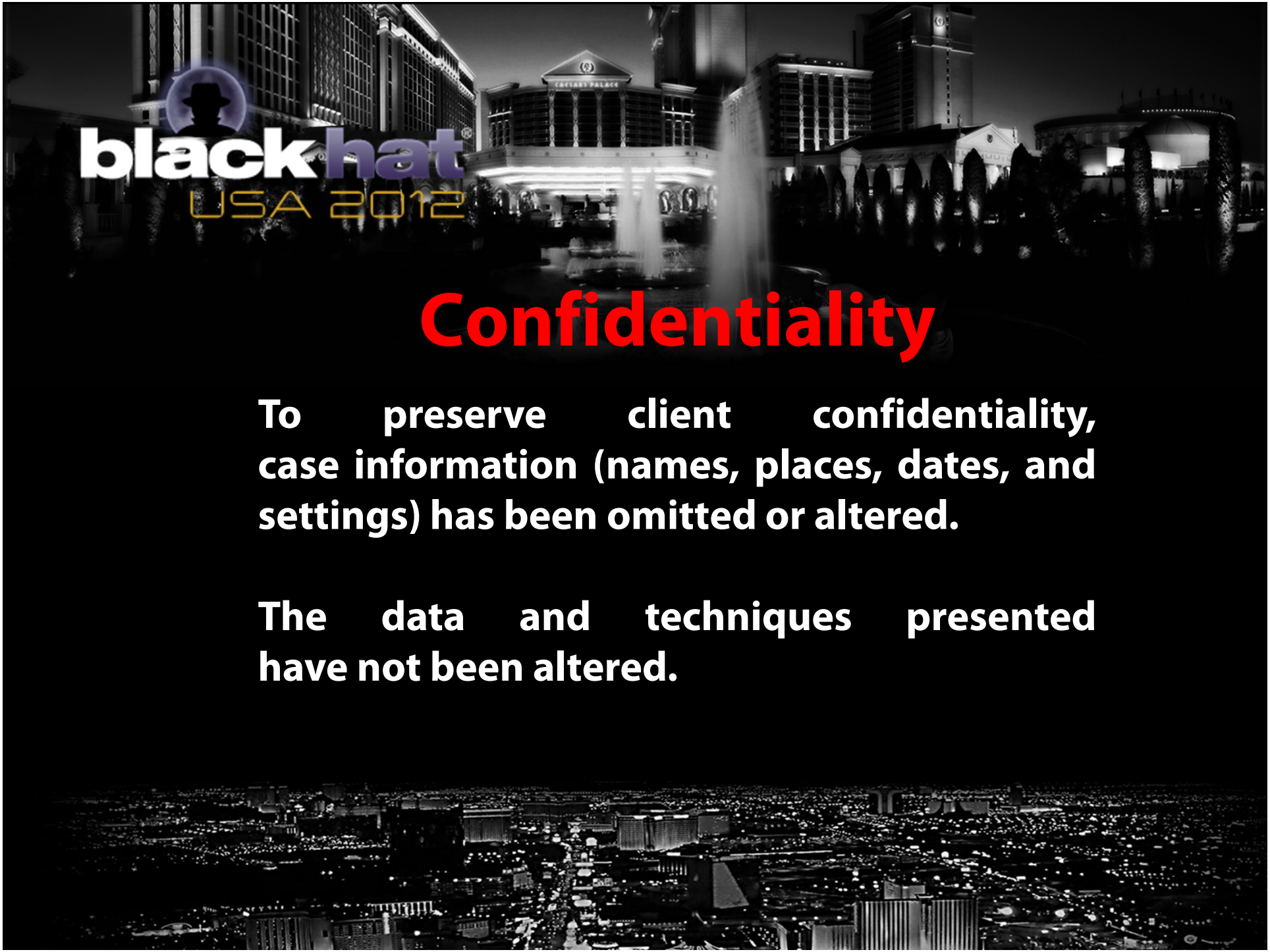

blackhat
USA 2012
Catching
Insider
Data Theft
with
Stochastic Forensics
Jonathan Grier



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 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

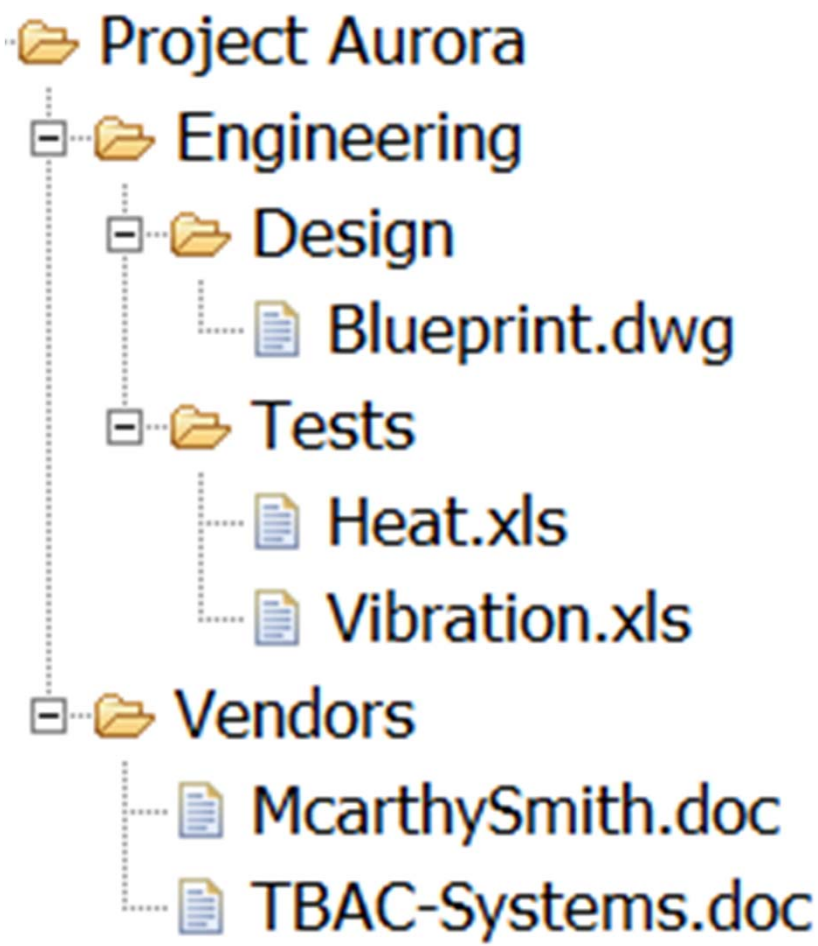
No Artifacts = No Forensics

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.

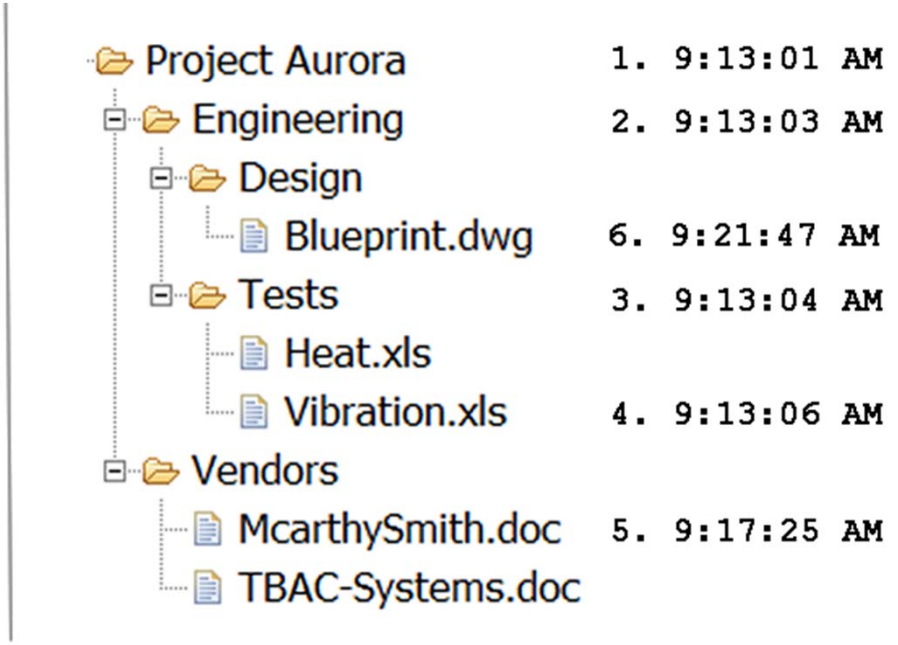
Harlan Carvey, *Windows Forensic Analysis*, 2009

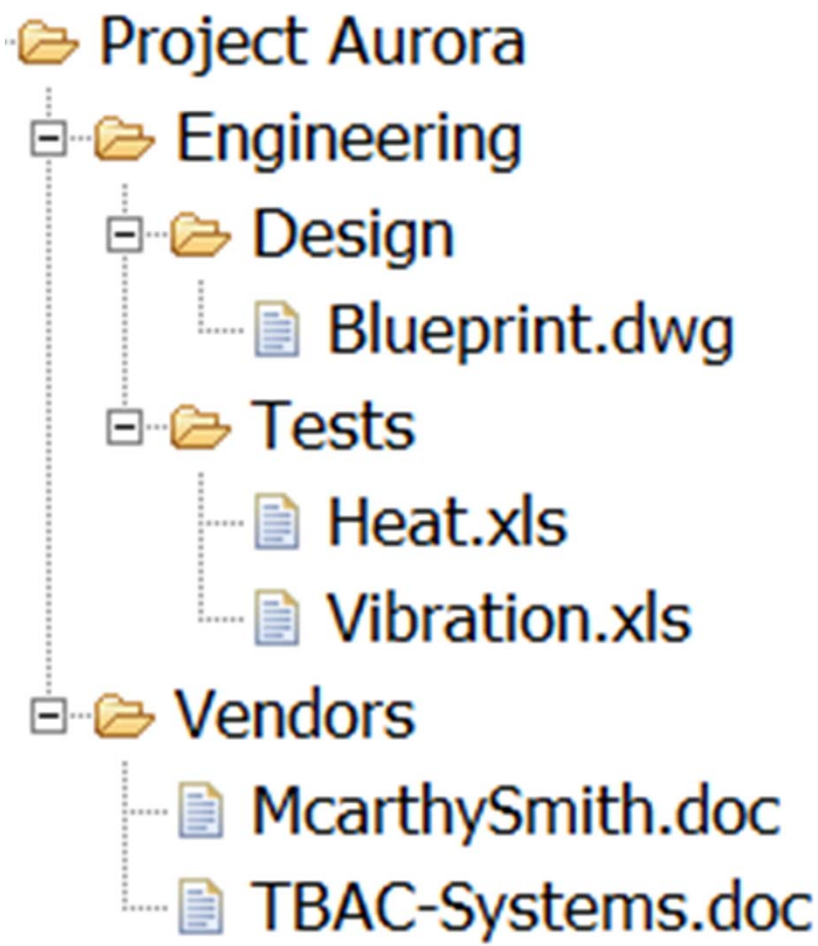
No Artifacts = No Forensics???


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

Access timestamps updates during:

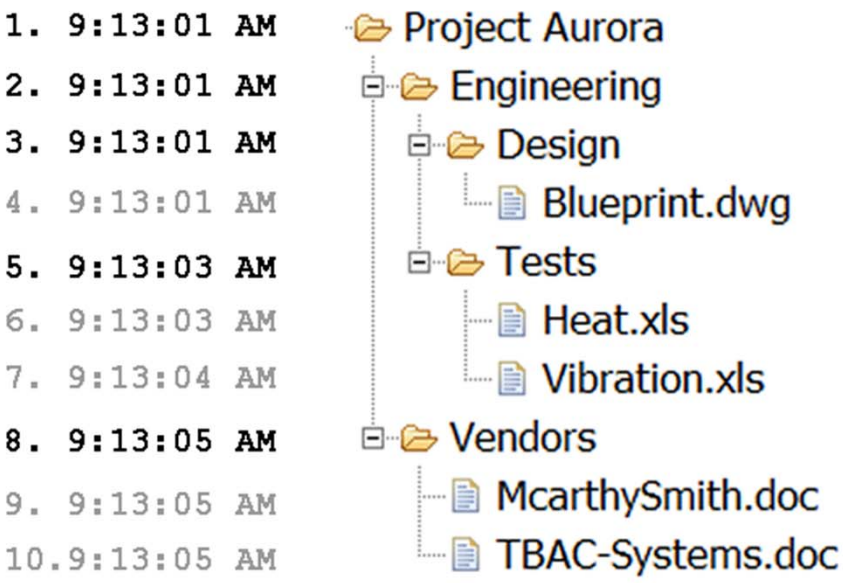
Routine access



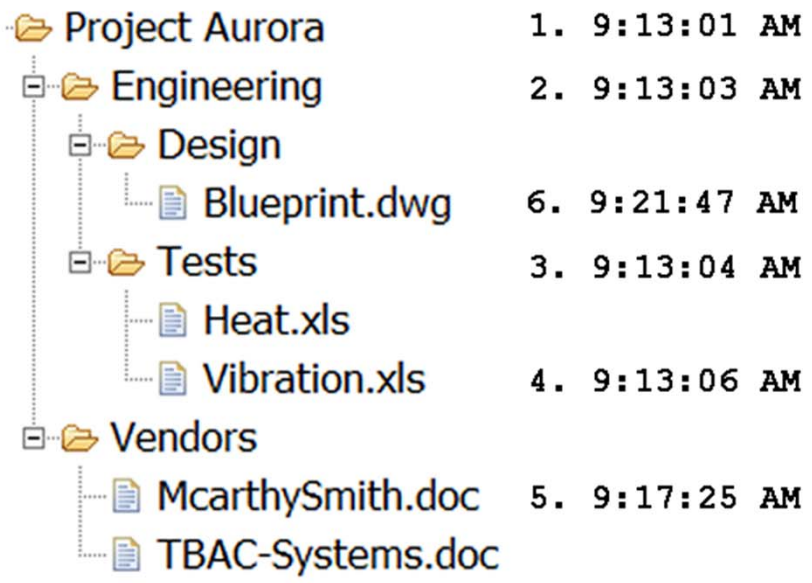

Project Aurora
Engineering
Design
Blueprint.dwg
Tests
Heat.xls
Vibration.xls
Vendors
McarthySmith.doc
TBAC-Systems.doc

Access timestamps updates during:

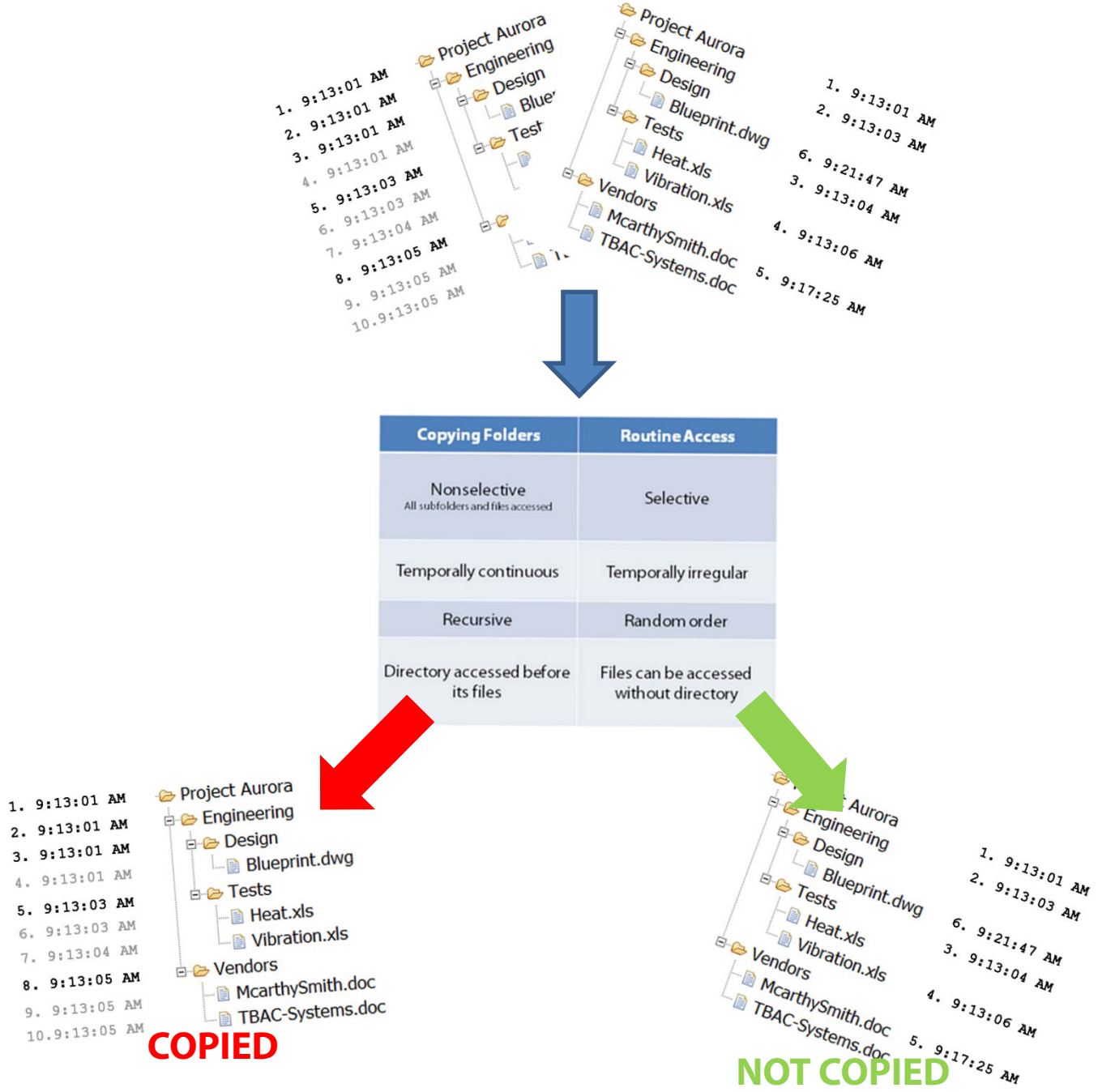
Copying a folder

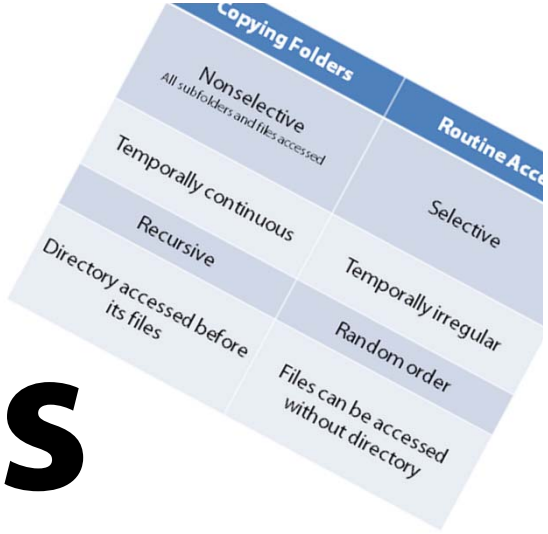


Routine access



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


Project Aurora
Engineering
Design
Blueprint.dwg
Tests
Heat.xls
Vibration.xls
Vendors
McarthySmith.doc
TBAC-Systems.doc
1. 9:13:01 AM
2. 9:13:03 AM
6. 9:21:47 AM
3. 9:13:04 AM
4. 9:13:06 AM
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
COPIED
NOT COPIED



No Artifacts
Yes Forensics

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

-- an anonymous reviewer

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*

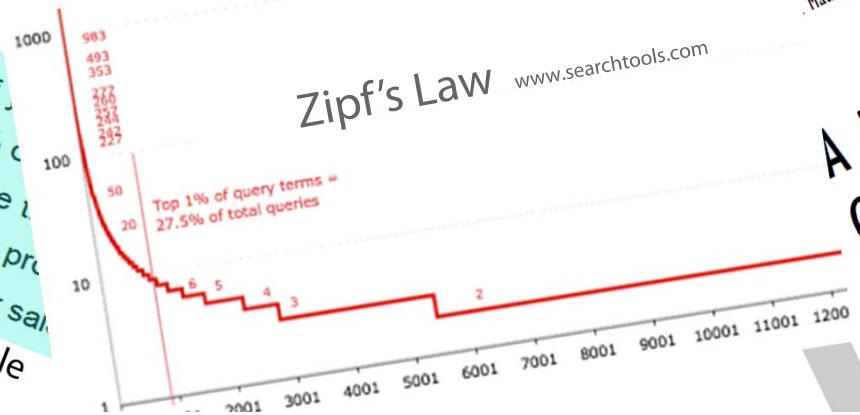
	www.things.org	www.fish.com	news.earthlink.net
Over one year:	76.6	75.9	10.9
Six months to one year:	7.6	18.6	7.2

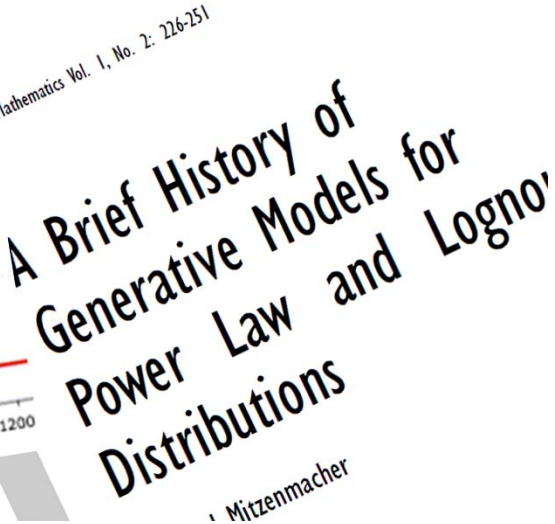
Farmer & Venema, *Forensic Discovery,* 2005

Pareto Principle

- 80% of your profits come from 20% of
- 80% of your complaints come from 20%
- 80% of your profits come from 20% of your pr
- 80% of your sales come from 20% of your sa
- 80% of your sales are made by 20% of your

//en.wikipedia.org/wiki/Pareto_principle





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 ≥ $t_{copying}$

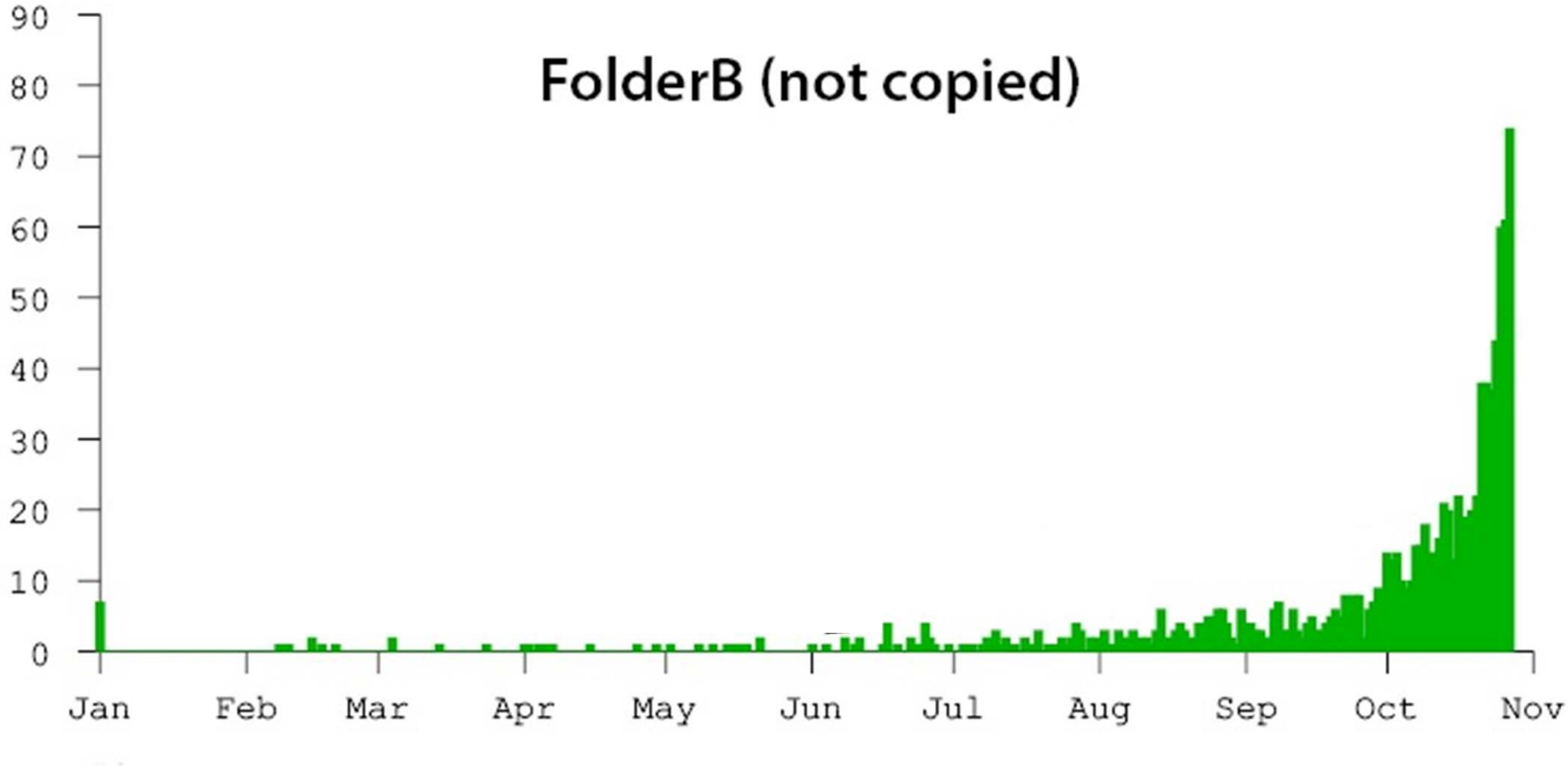
At $t_{copying}$:

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

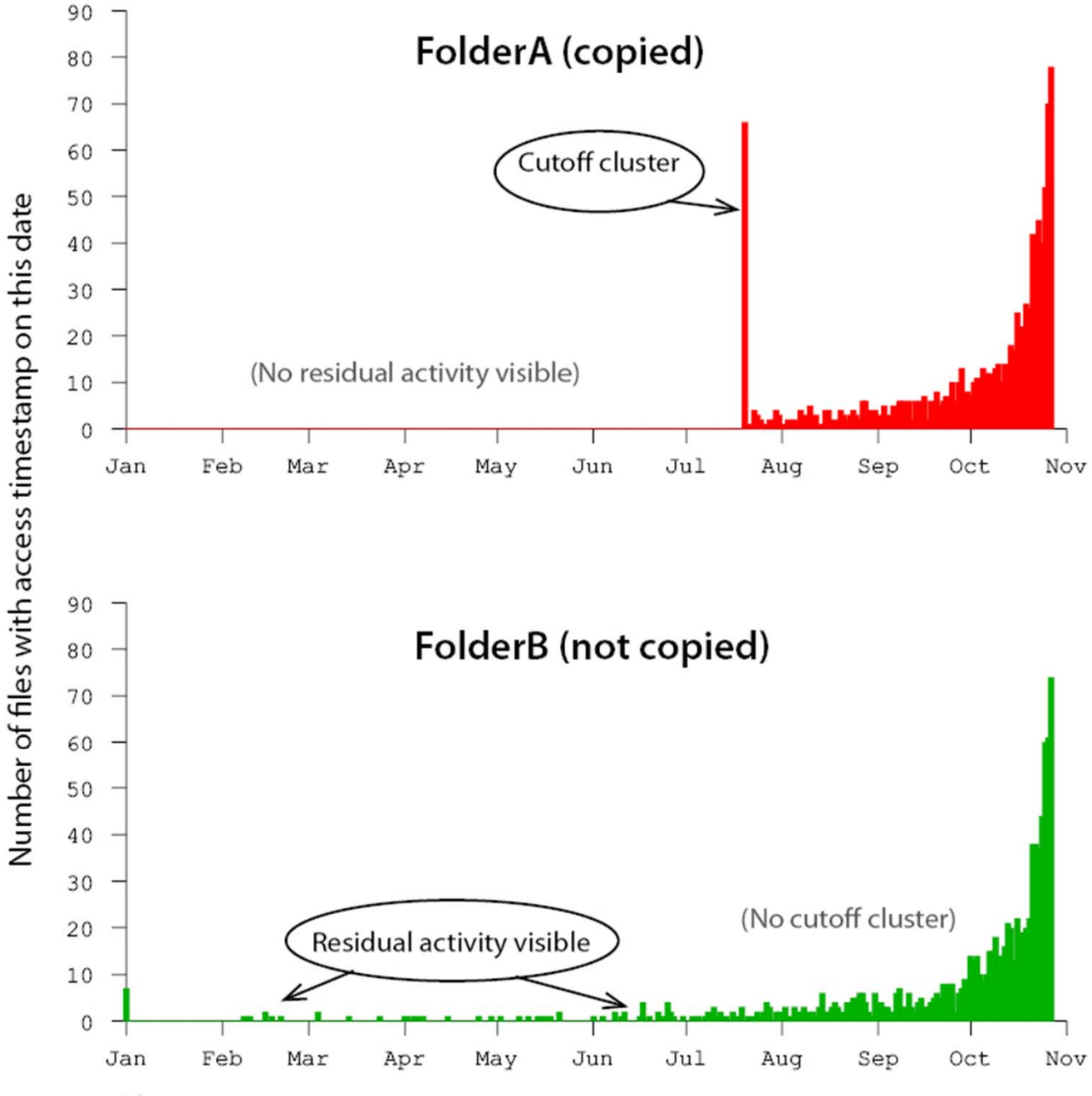
Several weeks later:

- 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

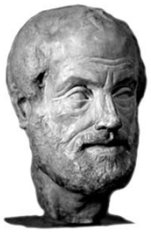

FolderA (copied)
Cutoff cluster
(No residual activity visible)
FolderB (not copied)
Residual activity visible
(No cutoff cluster)
Number of files with access timestamp on this date
0
10
20
30
40
50
60
70
80
90
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov

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

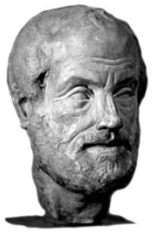
A ⟶ B doesn't prove B ⟶ A.

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.

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.

Who's trying to *prove* anything?

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

Perhaps they ran `grep`.

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.

Forensics

WHY ~~PROGRAMMING~~ IS A GOOD MEDIUM FOR investigating ~~EXPRESSING~~
POORLY UNDERSTOOD AND SLOPPILY-FORMULATED IDEAS.
-- Marvin Minsky, MIT, 1967

Forensics

WHY ~~PROGRAMMING~~ IS A GOOD MEDIUM FOR investigating ~~EXPRESSING~~
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

An actual investigation...

Part II:

Now for the real world…

NOISE

OpenSolaris **cp** command source code

```
        if (m < 0) {
                (void) fprintf(stderr, gettext(
                    "cp: cannot change owner and "
                    "group of %s:"), target);
                perror("");
        }
} else {
        /*
         * Copy the file.  If it happens to be a
         * symlink, copy the file referenced
         * by the symlink.
         */
        fi = open(source, O_RDONLY);
        if (fi < 0) {
                (void) fprintf(stderr,
                    gettext("%s: cannot open %s: "),
                    cmd, source);
```

Notice anything?

```
        if (m < 0) {
                (void) fprintf(stderr, gettext(
                    "cp: cannot change owner and "
                    "group of %s:"), target);
                perror("");
        }
} else {
        /*
         * Copy the file.  If it happens to be a
         * symlink, copy the file referenced
         * by the symlink.
         */
        fi = open(source, O_RDONLY);
        if (fi < 0) {
                (void) fprintf(stderr,
                    gettext("%s: cannot open %s: "),
```

Notice anything?

```
        if (m < 0) {
                (void) fprintf(stderr, gettext(
                    "cp: cannot change owner and "
                    "group of %s:"), target);
                perror("");
        }
} else {
        /*
         * Copy the file.  If it happens to be a
         * symlink, copy the file referenced
         * by the symlink.
         */
        fi = open(source, O_RDONLY);
        if (fi < 0) {
                (void) fprintf(stderr,
                    gettext("%s: cannot open %s: "),
                    cmd, source);
```

OpenSolaris **`cp`** command source code **`writefile()`** function

```
        /*
         * Mmap time!
         */
        if ((cp = mmap((caddr_t)NULL, mapsize, PROT_READ,
            MAP_SHARED, fi, (off_t)0)) == MAP_FAILED)
                mapsize = 0;    /* can't mmap today */
} else
        mapsize = 0;

if (mapsize != 0) {
        offset = 0;

        for (;;) {
                nbytes = write(fo, cp, mapsize);
                /*
                 * if we write less than the mmaped size it's due to a
                 * media error on the input file or out of space on
                 * the output file.  So, try again, and look for errno.
```

CopyFile function

msdn

Copies an existing file to a new file.

The **CopyFileEx** function provides two additional capabilities. **CopyFileEx** can call a specified callback function each time a portion of the copy operation is completed, and **CopyFileEx** can be canceled during the copy operation.

To perform this operation as a transacted operation, use the **CopyFileTransacted** function.

Syntax

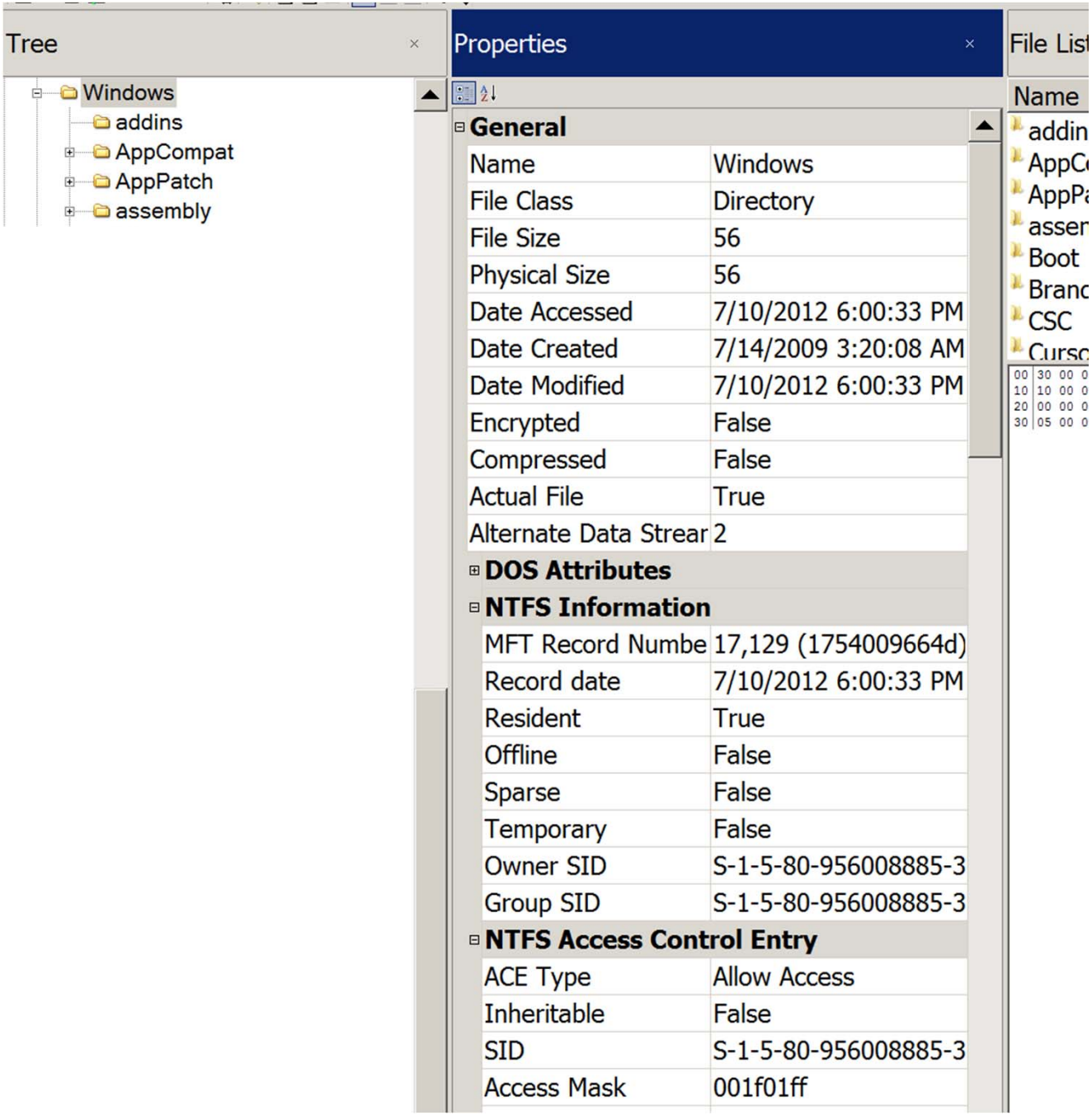
C++

```
BOOL WINAPI CopyFile(
  _In_  LPCTSTR lpExistingFileName,
  _In_  LPCTSTR lpNewFileName,
  _In_  BOOL bFailIfExists
);
```

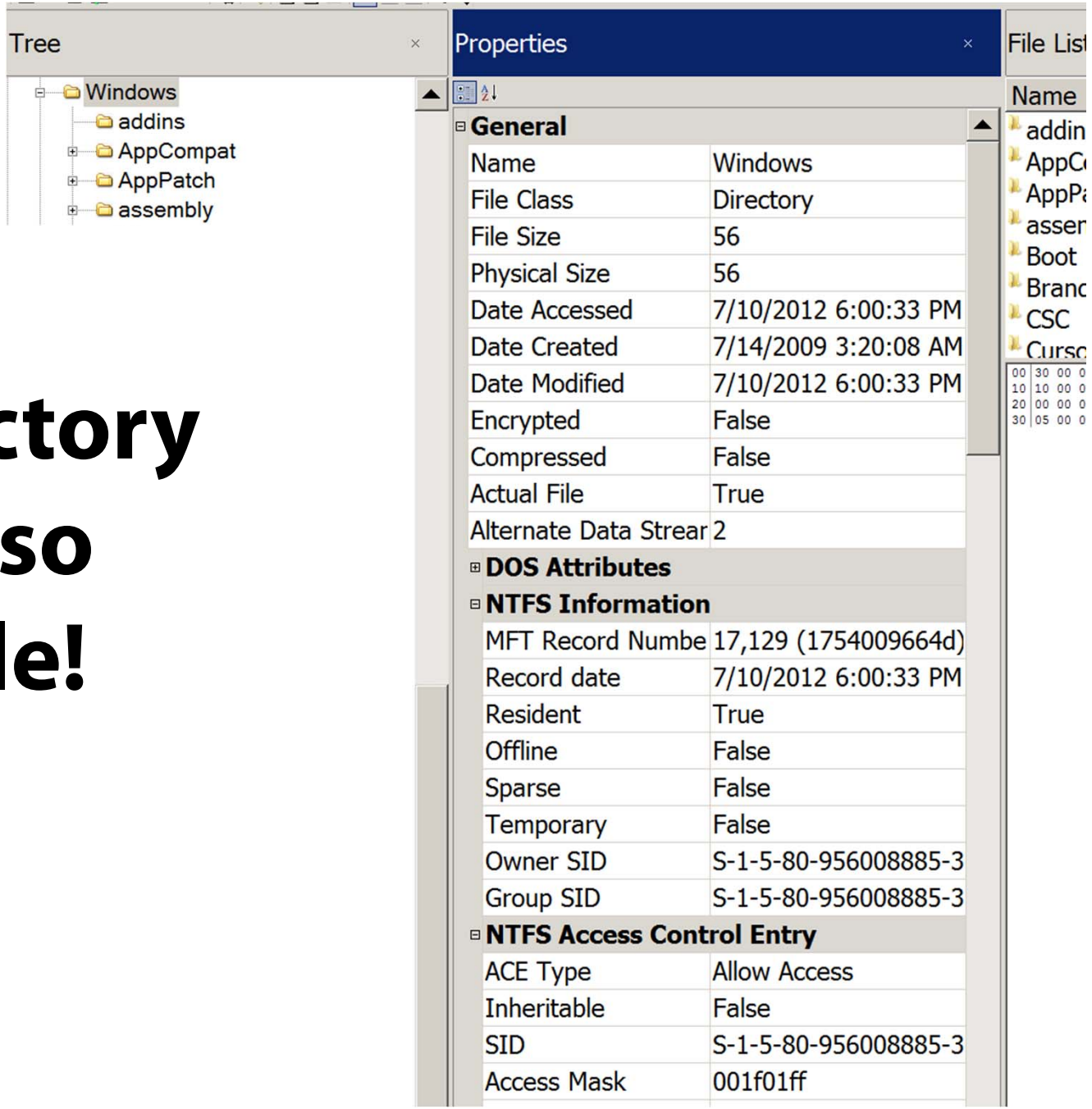
Is all lost

(on Windows at least)

?


Tree
Windows
addins
AppCompat
AppPatch
assembly
Properties
General
Name Windows
File Class Directory
File Size 56
Physical Size 56
Date Accessed 7/10/2012 6:00:33 PM
Date Created 7/14/2009 3:20:08 AM
Date Modified 7/10/2012 6:00:33 PM
Encrypted False
Compressed False
Actual File True
Alternate Data Strear 2
DOS Attributes
NTFS Information
MFT Record Numbe 17,129 (1754009664d)
Record date 7/10/2012 6:00:33 PM
Resident True
Offline False
Sparse False
Temporary False
Owner SID S-1-5-80-956008885-3
Group SID S-1-5-80-956008885-3
NTFS Access Control Entry
ACE Type Allow Access
Inheritable False
SID S-1-5-80-956008885-3
Access Mask 001f01ff
File List
Name
addin
AppC
AppPa
assen
Boot
Branc
CSC
Cursc

a Directory is also a File!



Filter...


NOISE

ACCURACY?

Who needs

ACCURACY?

Part III:

Applying Stochastic Forensics

Eyeball?

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|12234
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|12
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|12234
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|12
0|/Documents and Settings/nbrown/NTUSER.DAT|8022-128-4|r/rr-xr-xr-x|0|0|4194304|1252983243|1250178790|1240925796|1223
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|1223472713
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

Filter
&
Plot

Filter

1. By folder

Filter

1. By folder
2. Directories versus Files

Filter

1. By folder
2. Directories versus Files
3. Permissions

Filter

1. By folder
2. Directories versus Files
3. Permissions
4. Other

Plot

Our visual cognition is amazingly robust

Ploticus: http://ploticus.sourceforge.net

Interpret
&
Advance

No Cluster?

Strong evidence of *no* copying

Found Cluster?

1. Check control folders
2. Search for causes
3. Fingerprint it

Found Cluster?

A cluster defines a tight *window of opportunity*.

Use it to *propel the investigation forward.*

Part IV:

Forensic Hacking

hack *v.*

Exploring the inner workings of something by using it in a way its creators never imagined.

Classical Forensics:

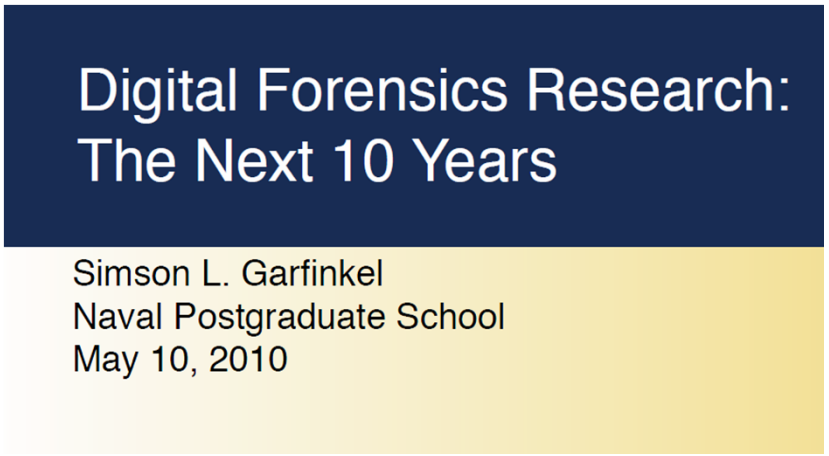
Look at the Surviving Data → Reconstruct Previous Data → This previous data is our deliverable.

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.



Leading researchers have called to move from:

"What data can we find?"

To:

"What did this person do?"

Forensics

WHY ~~PROGRAMMING~~ IS A GOOD MEDIUM FOR investigating ~~EXPRESSING~~
POORLY UNDERSTOOD AND SLOPPILY-FORMULATED IDEAS.
-- Marvin Minsky, MIT, 1967

Forensics

WHY ~~PROGRAMMING~~ IS A GOOD MEDIUM FOR investigating ~~EXPRESSING~~
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 Agenda
(i.e. a request for help)

1. Scientific testing

   Automate, build corpus, confidence levels, validate

2. Fingerprinting

   We *can* distinguish copying from `grep`!

3. Probability value

4. What other questions can stochastic forensics address?

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

Questions?
Comments?
Want More Info?

Please speak to me,
here at Black Hat
or jdgrier at grierforensics com.