Forensic Results from 236 Used Hard Drives

Simson L. Garfinkel, Ph.D.

11 August 2005

I purchased 10 used computers from a store in August 1998

Computer #1: 486-class machine with 32MB of RAM

- It boot!
- Law firm's file server...
- Still had client documents!

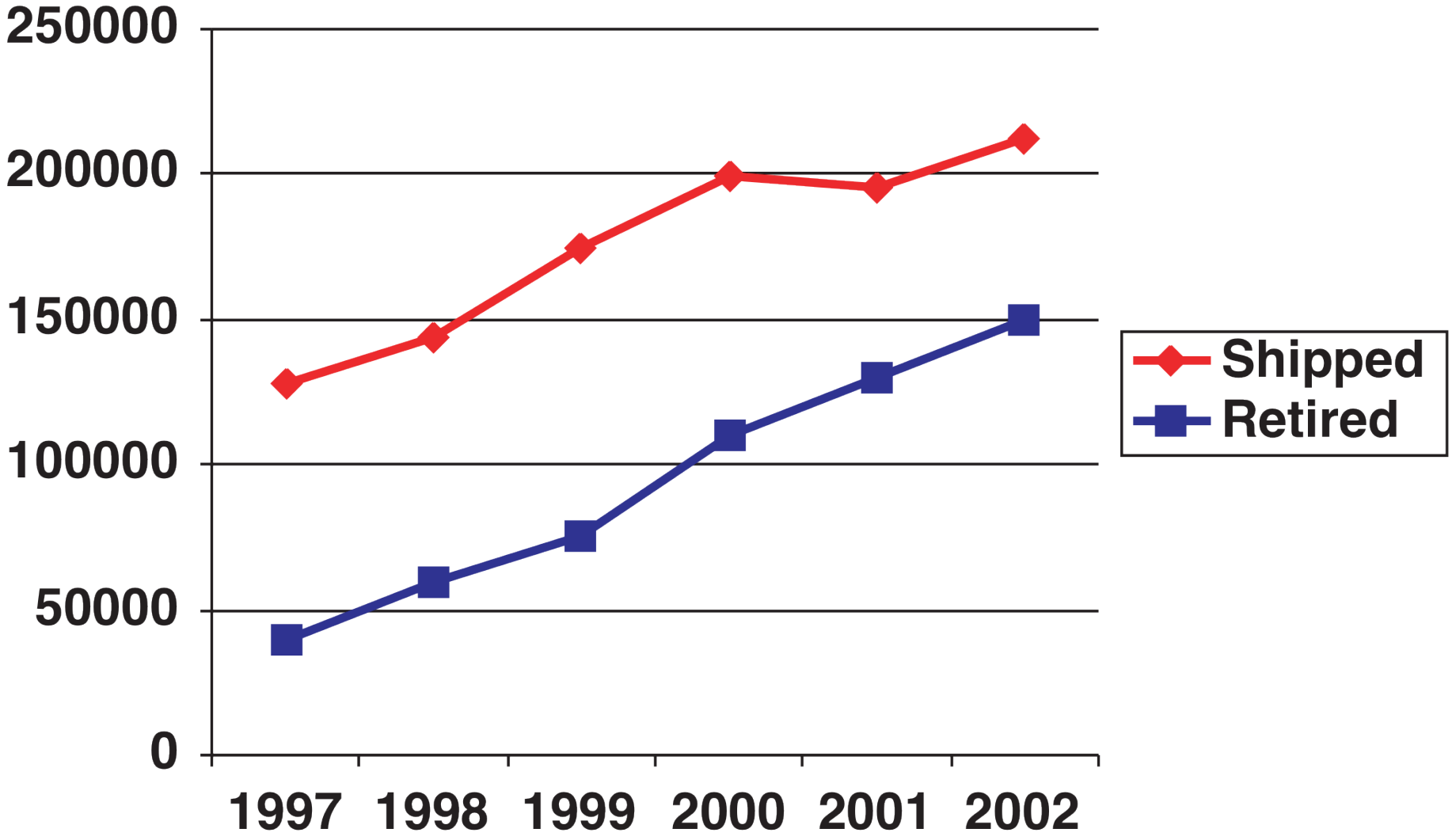
Other computers had equally interesting data.

- Database of community-outreach mental health services
- Quicken files (home finance)
- Draft manuscript from a novelist

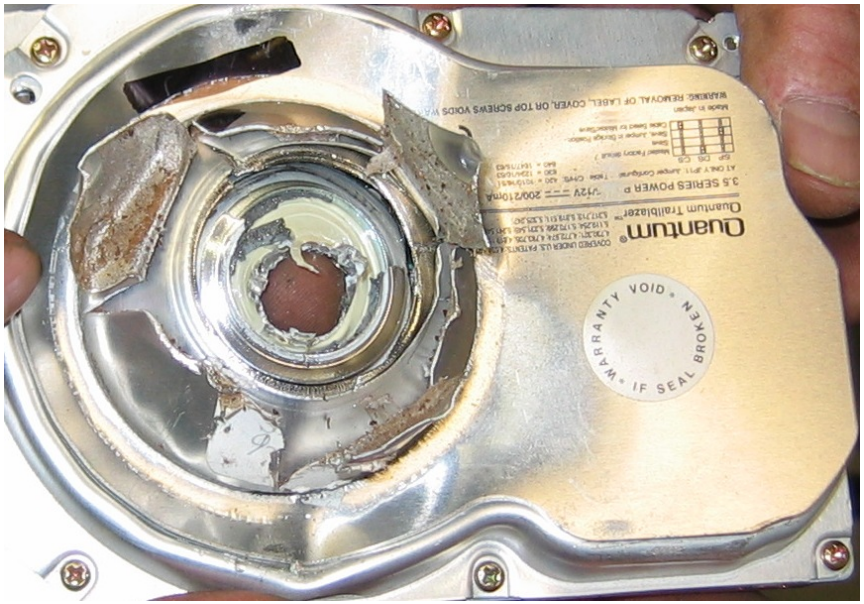
Hard Drives Pose Special Problem For Computer Security

- Do not forget data when power is removed.
- Can contain data that is not immediately visible.
- Today's computers can read hard drives that are 15 years old!
  - Electrically compatible (IDE/ATA)
  - Logically compatible (FAT16/32 file systems)
  - Very different from tape systems

149 million drives retired in 2002!



Physical destruction will remove the information...



...but many “retired” drives are not physically destroyed.

There is a significant secondary market for used disk drives.

- Re-used within organizations
- Given to charities
- Sold on eBay

All Categories Save this search

350 items found for **hard drives**

Sort by items: **ending first** | newly listed | lowest priced | highest priced

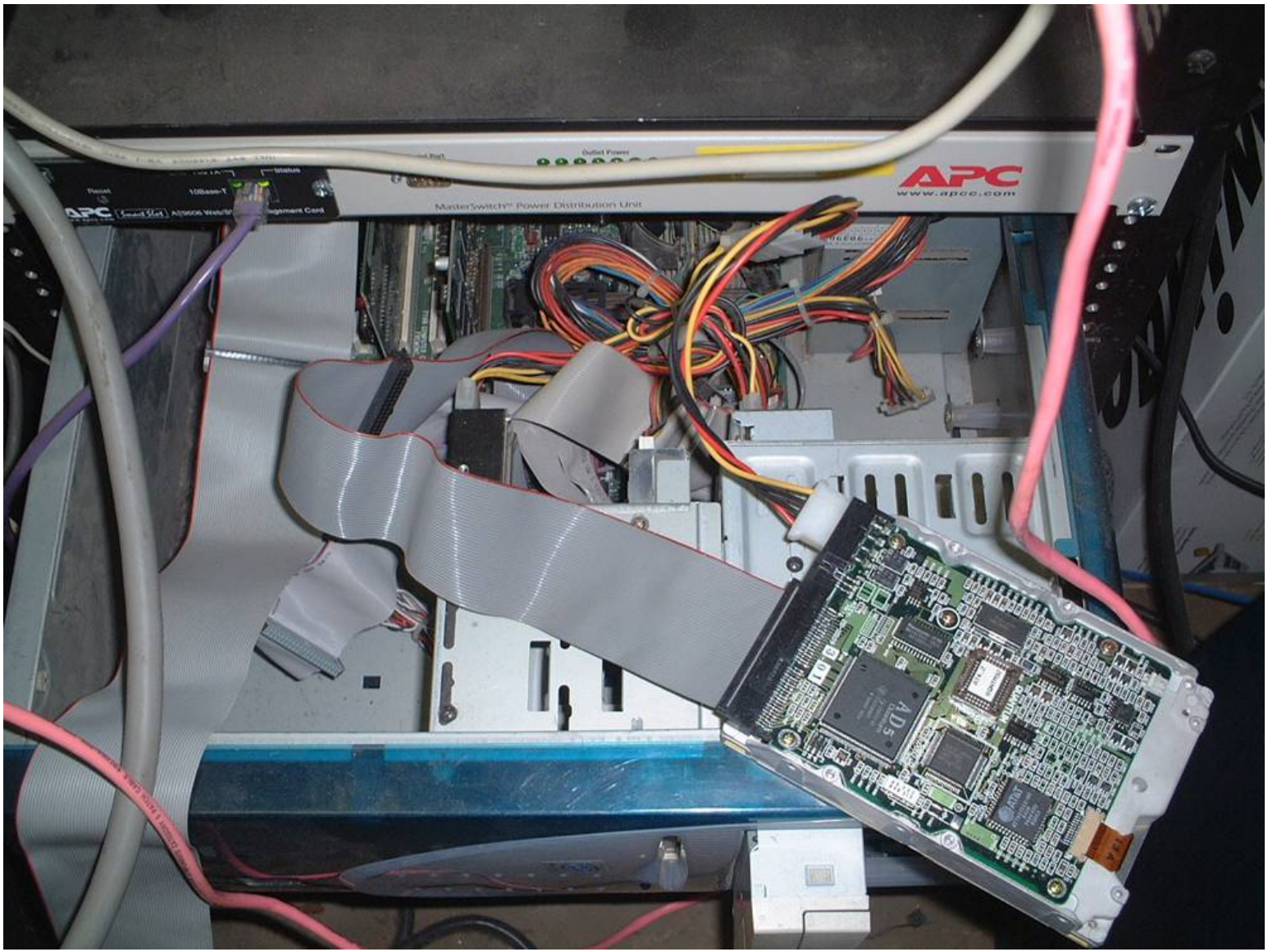
Picture Bids	Item Title	Price	Bids	Time Left
	Lot of hard and floppy drives	$5.50	2	14m
	Lot of hard and floppy drives	$5.50	2	22m
	Lot of hard and floppy drives	$5.50	2	25m
	Lot of 2 hard drives IDE	$8.00	12	29m
	3.2 gig Hard Drives	$180.00	-	59m
	(5) 1.2 hard drives & (15) 10/100 network	$15.00	1	1h 00m
	Lot of 3 Quantum 9.1 gig SCSI Hard Drives	$36.00	6	1h 25m
	IDE HARD DRIVES (3)	$6.50	6	1h 46m
	LOT OF 5 Hard Drives! 3.2 Gig Western Digital	$120.00 $124.95 Buy it Now	-	1h 50m
	QTY 3... IDE Hard Drives 2.5 Gig	$10.50	5	2h 02m
	5 WESTERN DIGITAL 2.5 GIG HARD DRIVES	$30.00	4	2h 03m
	QTY 3... IDE Hard Drives 1.0 Gig	$9.99	1	2h 04m
	Western Digital 850 meg IDE Hard Drives dutch	$6.00	1	2h 57m
	WINDOWS	$6.00	-	3h 18m

Retired drives given to schools, the poor, and sold.

Between January 1999 and April 2002, I acquired 236 hard drives on the secondary market.

Drives arrived by UPS

Copied data off using FreeBSD




```
dd if=/dev/ad0 of=file.img bs=65536 conv=noerror,sync
```

Stored images on a RAID

For every drive, I cataloged:

- Disk SN, date of manufacture, etc.
- All visible files.
- MD5 of every file.
- Every disk block.
- MD5 of the image.

Example: Disk #70: IBM-DALA-3540/81B70E32

Purchased for $5 from a Mass retail store on eBay

Copied the data off: 541MB

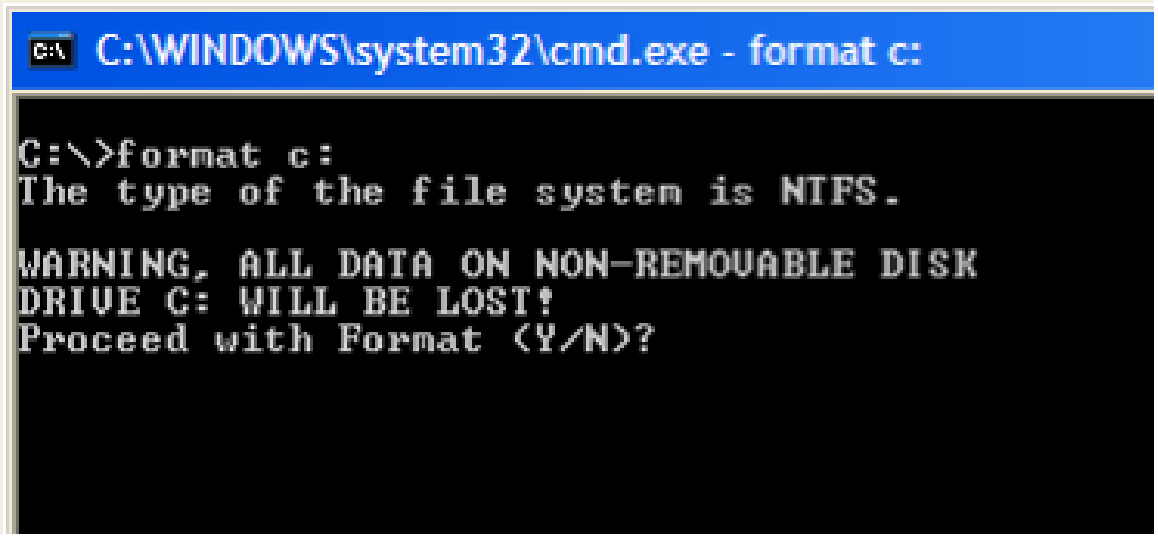
Initial analysis:

Total disk sectors:	1,057,392
Total non-zero sectors:	989,514
Total files:	3

The files:

```
drwxrwxrwx  0 root        0 Dec 31  1979 ./
-r-xr-xr-x  0 root   222390 May 11  1998 IO.SYS
-r-xr-xr-x  0 root        9 May 11  1998 MSDOS.SYS
-rwxrwxrwx  0 root    93880 May 11  1998 COMMAND.COM
```

Clearly, this disk was formatted...



But Windows FORMAT doesn't erase the disk...
FORMAT just writes a new root directory.

UNIX “strings” reveals the disk’s previous contents...

```
Insert diskette for drive
 and press any key when ready
Your program caused a divide overflow error.
If the problem persists, contact your program vendor.
Windows has disabled direct disk access to protect your lo
To override this protection, see the LOCK /? command for m
The system has been halted.  Press Ctrl+Alt+Del to restart
You started your computer with a version of MS-DOS incompa
version of Windows. Insert a Startup diskette matching thi

OEMString = "NCR 14 inch Analog Color Display Enchanced SV
        Graphics Mode: 640 x 480 at 72Hz vertical refresh.
        XResolution              = 640
        YResolution              = 480
        VerticalRefresh          = 72
```

70.img con't...

```
ling the Trial Edition
------------------------------
IBM AntiVirus Trial Edition is a full-function but time-li
evaluation version of the IBM AntiVirus Desktop Edition pr
may have received the Trial Edition on a promotional CD-RO
single-file installation program over a network.  The Tria
is available in seven national languages, and each languag
provided on a separate CC-ROM or as a separa
EAS.STCm
EET.STC
ELR.STCq
ELS.STC
```

70.img con’t...

MAB-DEDUCTIBLE

MAB-MOOP

MAB-MOOP-DED

METHIMAZOLE

INSULIN (HUMAN)

COUMARIN ANTICOAGULANTS

CARBAMATE DERIVATIVES

AMANTADINE

MANNITOL

MAPROTILINE

CARBAMAZEPINE

CHLORPHENESIN CARBAMATE

ETHINAMATE

FORMALDEHYDE

MAFENIDE ACETATE

[Garfinkel & Shelat 03] established the scale of the problem.

We found:

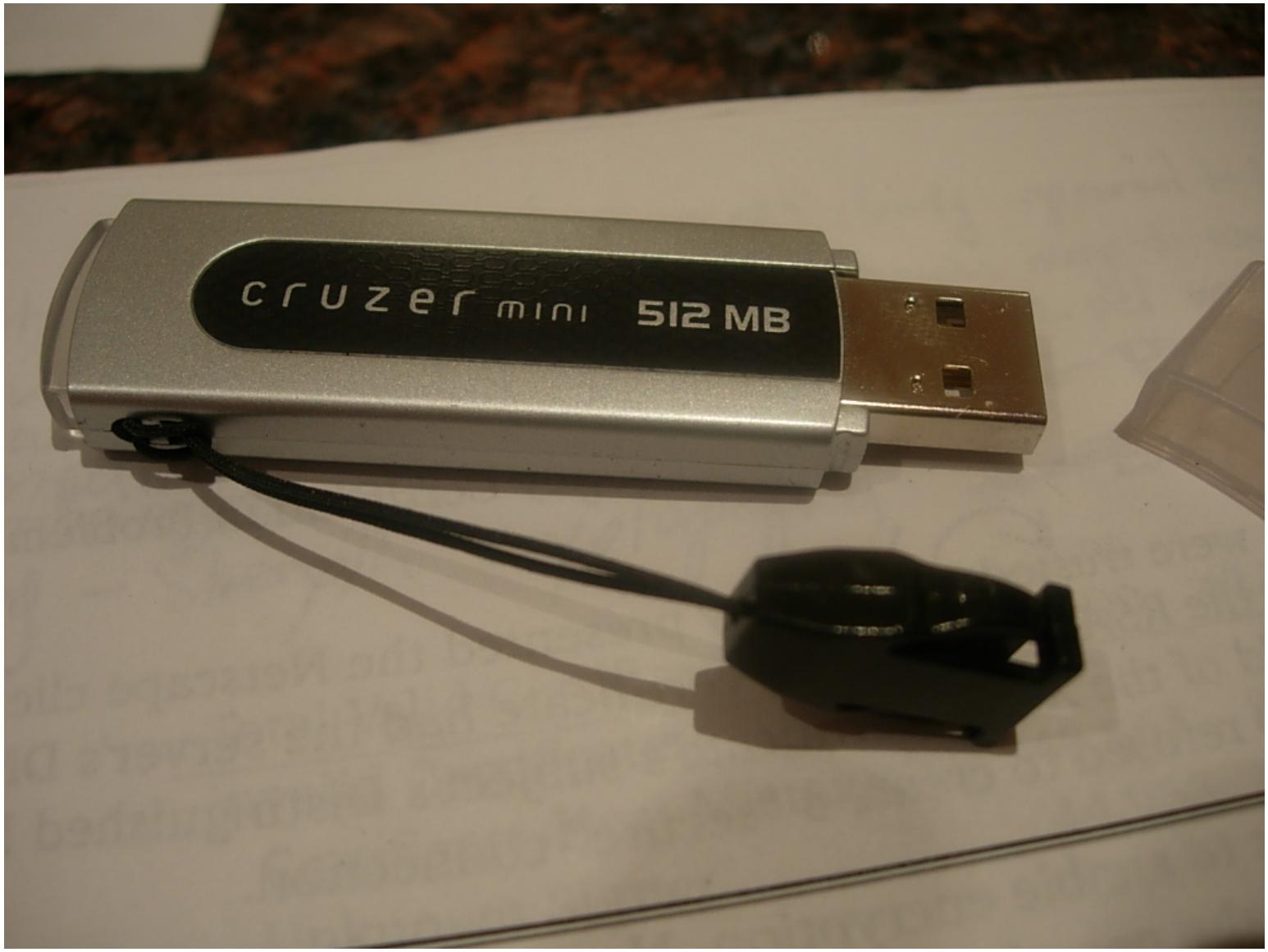
- Thousands of credit card numbers (many disks)
- Financial records
- Medical information
- Trade secrets
- Highly personal information

We did not determine why the data had been left there.

There are roughly a dozen documented cases of people purchasing old PCs and finding sensitive data.

- A woman in Pahrump, NV bought a used PC with pharmacy records [Markoff 97]
- Pennsylvania sold PCs with “thousands of files” on state employees [Villano 02]
- Paul McCartney’s bank records sold by his bank [Leyden 04]
- O&O Software GmbH – 100 drives (10% properly wiped) [O&O 04]

Information is even left behind on USB tokens...



A purchased token contained images from the previous owner.

With so many used systems, why so few stories of actual data disclosure?

- Hypothesis #1: Disclosure of “data passed” is exceedingly rare because most systems are properly sanitized.
- Hypothesis #2: Disclosures are so common that they are not newsworthy.
- Hypothesis #3: Systems aren’t properly sanitized, but few people notice the data.

How could people not notice the data?

C:\WINDOWS\system32\cmd.exe

```
C:\tmp>dir
 Volume in drive C has no label.
 Volume Serial Number is 1410-FC4A

 Directory of C:\tmp

10/15/2004  09:20 PM    <DIR>          .
10/15/2004  09:20 PM    <DIR>          ..
10/03/2004  11:34 AM        27,262,976 big_secret.txt
               1 File(s)     27,262,976 bytes
               2 Dir(s)   4,202,078,208 bytes free

C:\tmp>del big_secret.txt

C:\tmp>dir
 Volume in drive C has no label.
 Volume Serial Number is 1410-FC4A

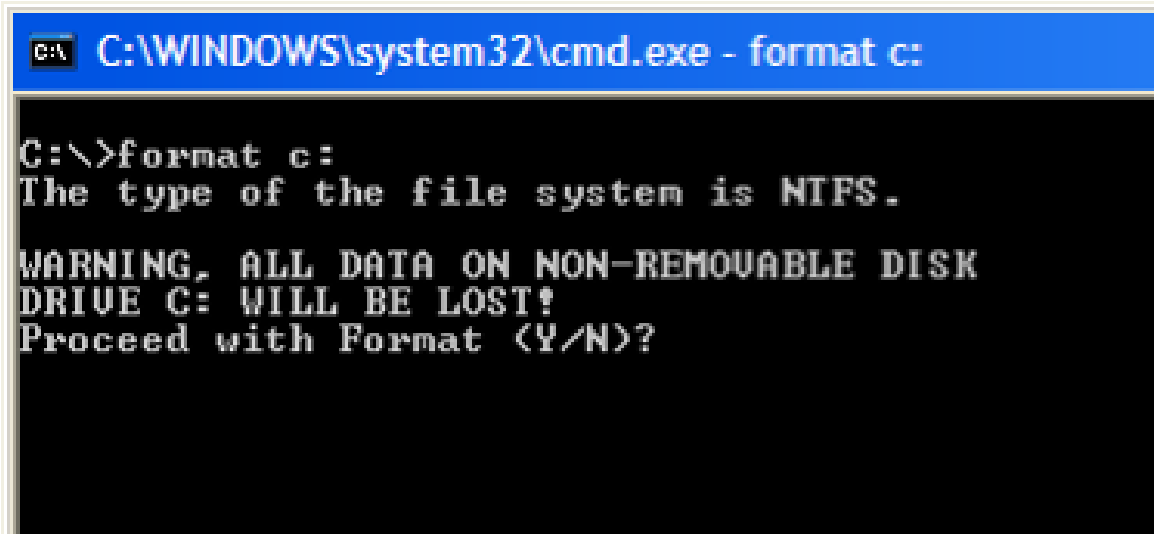
 Directory of C:\tmp

10/15/2004  09:22 PM    <DIR>          .
10/15/2004  09:22 PM    <DIR>          ..
               0 File(s)              0 bytes
               2 Dir(s)   4,229,296,128 bytes free

C:\tmp>_
```

DEL removes the file's name, but doesn't delete the files' data.

FORMAT writes a new root directory and FAT.



FORMAT doesn’t doesn’t overwrite the disk sectors.

Weird Stuff, Sunnyvale California,January 1999

10GB drive: $19 “tested”

500 MB drive: $3 “as is”

Q: “How do you sanitize them?”

A: “We FDISK them!”

FDISK does not sanitize disks

10 GB drive: 20,044,160 sectors

Command	Sectors Written	%
FORMAT	21,541	0.11%
FDISK	2,563	0.01%

FORMAT erases the FAT,
complicating the recovery of fragmented files.

Note: We are not considering exotic recovery techniques.

We assume that writing a sector destroys its previous contents.

- Secret government agencies with advanced technology might be able to recover overwritten data.
- Nobody has ever publicly demonstrated this technology.

“The Protection of Information in Computer Systems.” Saltzer and Schroeder [1975]

h) Psychological acceptability: It is essential that the human interface be designed for ease of use, so that users routinely and automatically apply the protection mechanisms correctly. ... User interfaces that more closely match the mental models people have of information protection are needed.

To be effective, a solution to this problem needs to address the root cause.

Usability Problem:

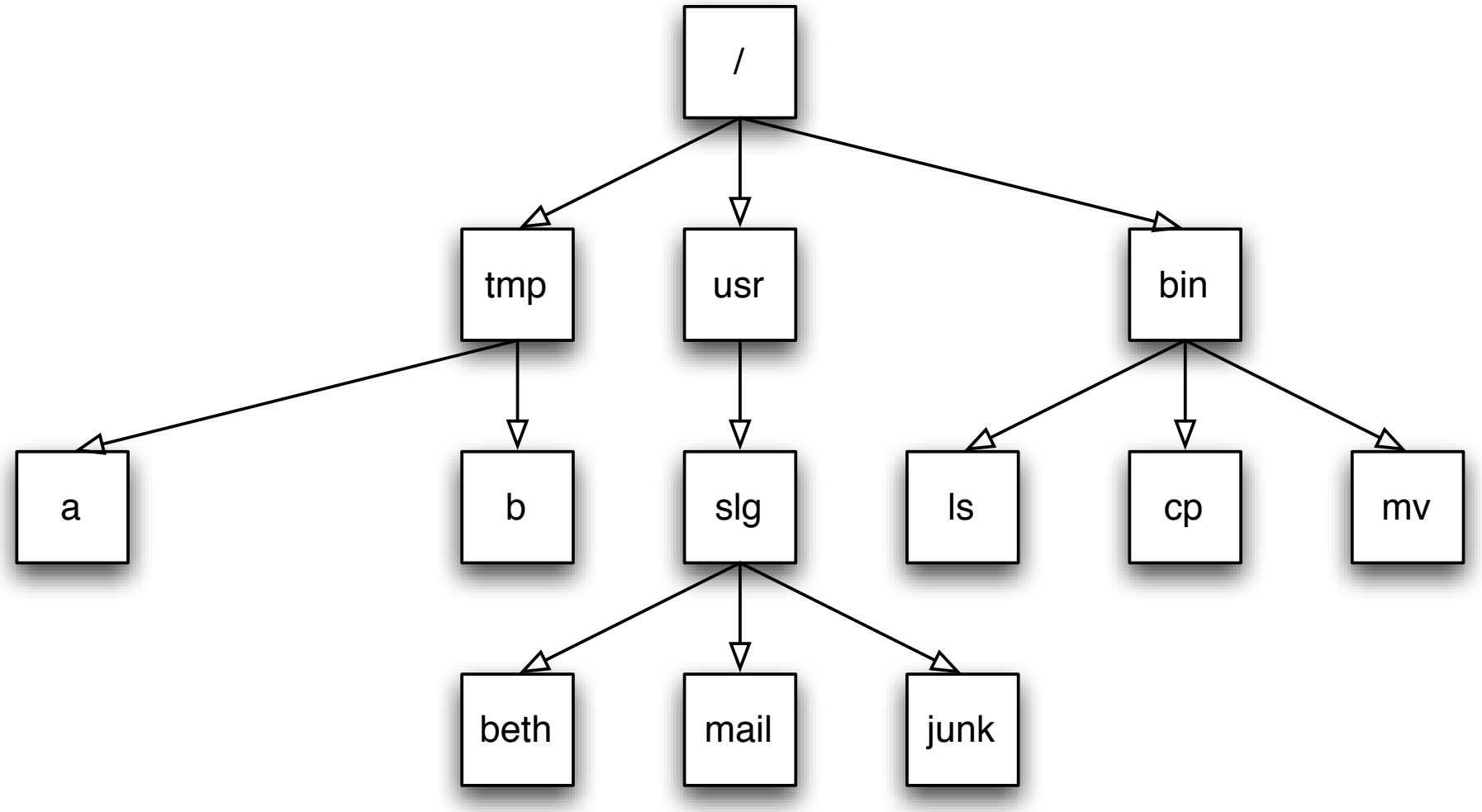
- Effective audit of information present on drives.
- Make DEL and FORMAT actually remove data. [Bauer & Priyantha 01]
- Provide alternative strategies for data recovery.

Education Problem:

- Add training to the interface. [Whitten 04]
- Regulatory requirements. [FTC 05, SEC 05]
- Legal liability.

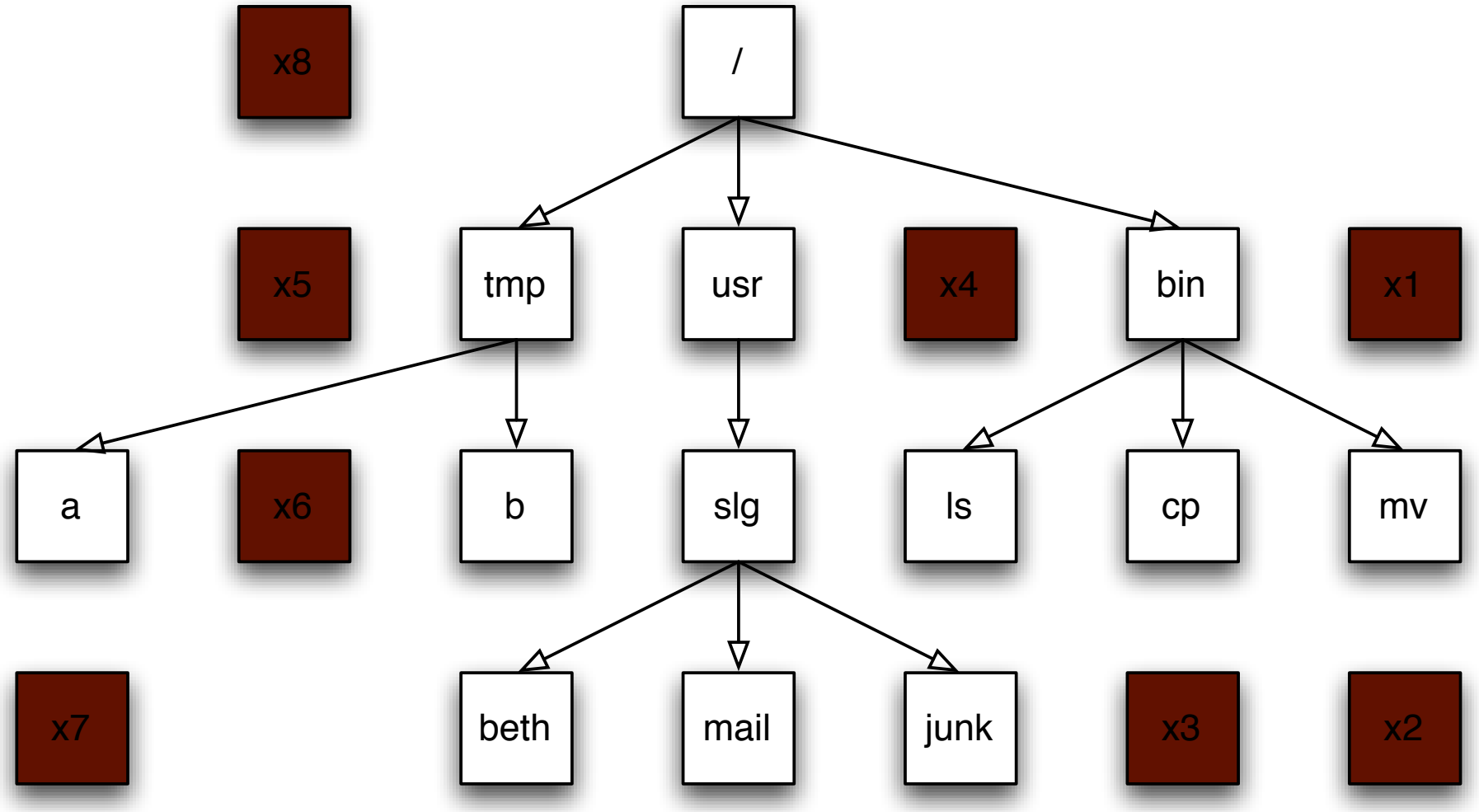
To determine the root cause, I looked *on the drives* and *contacted the data subjects*.

Data on a hard drive is arranged in sectors.



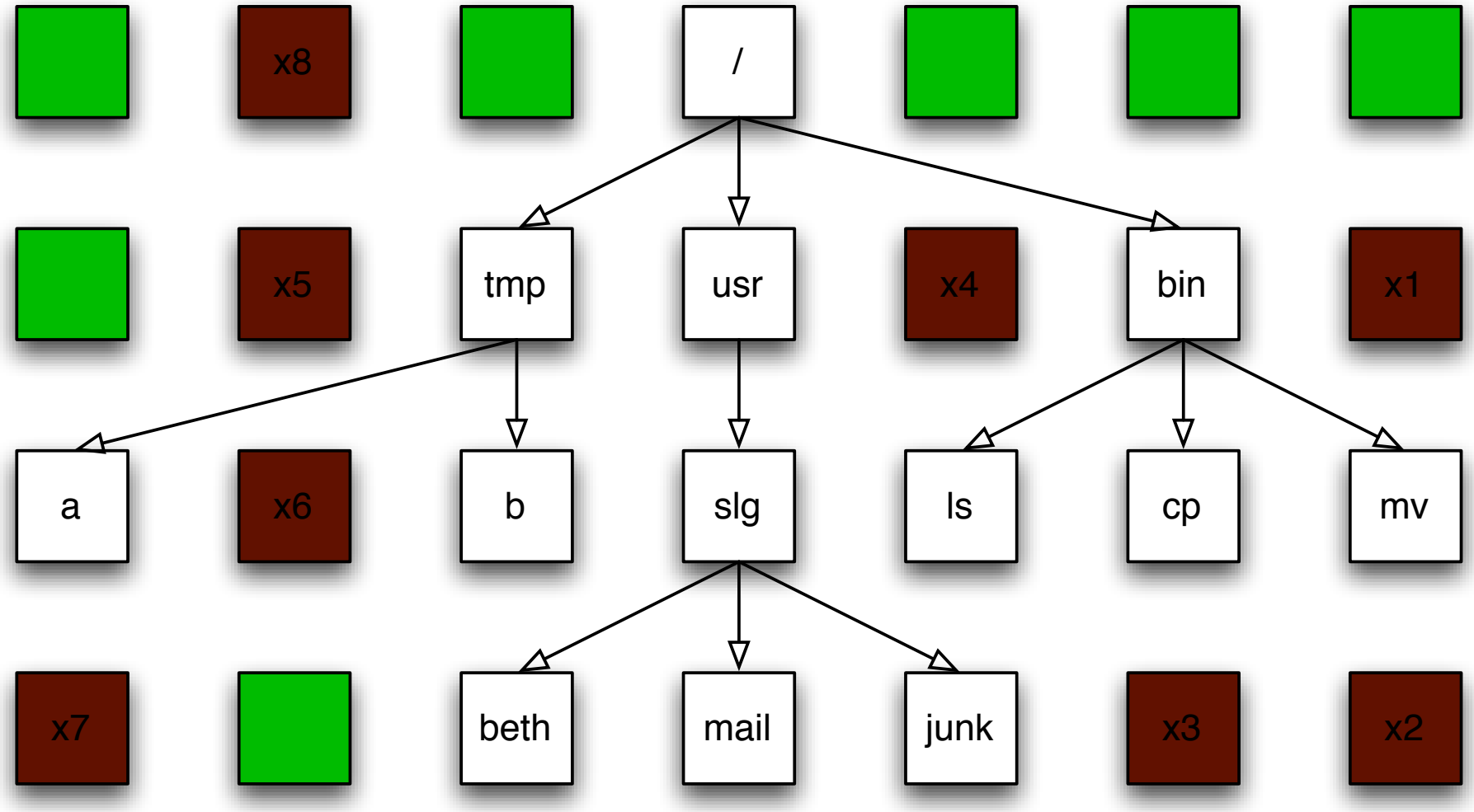
The white sectors indicate directories and files that are visible to the user.

Data on a hard drive is arranged in sectors.



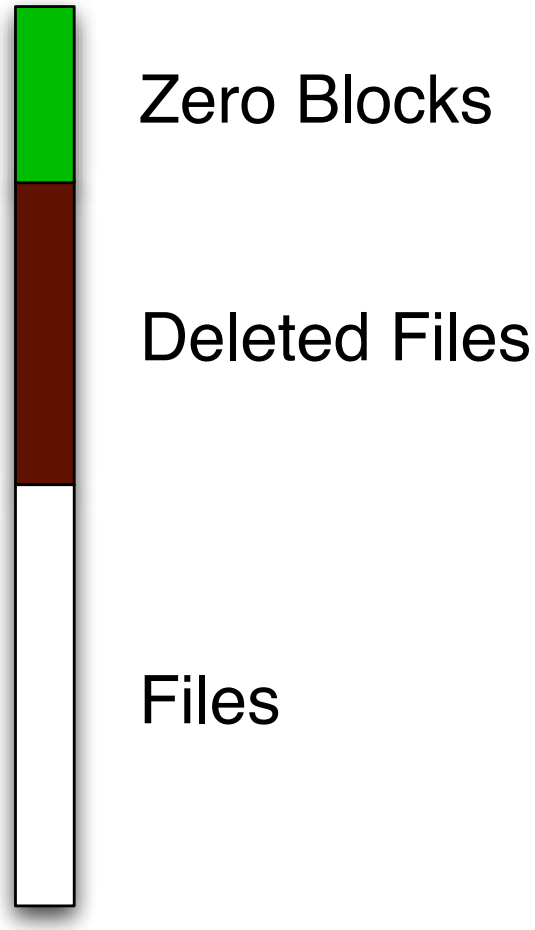
The brown sectors indicate files that were deleted.

Data on a hard drive is arranged in sectors.

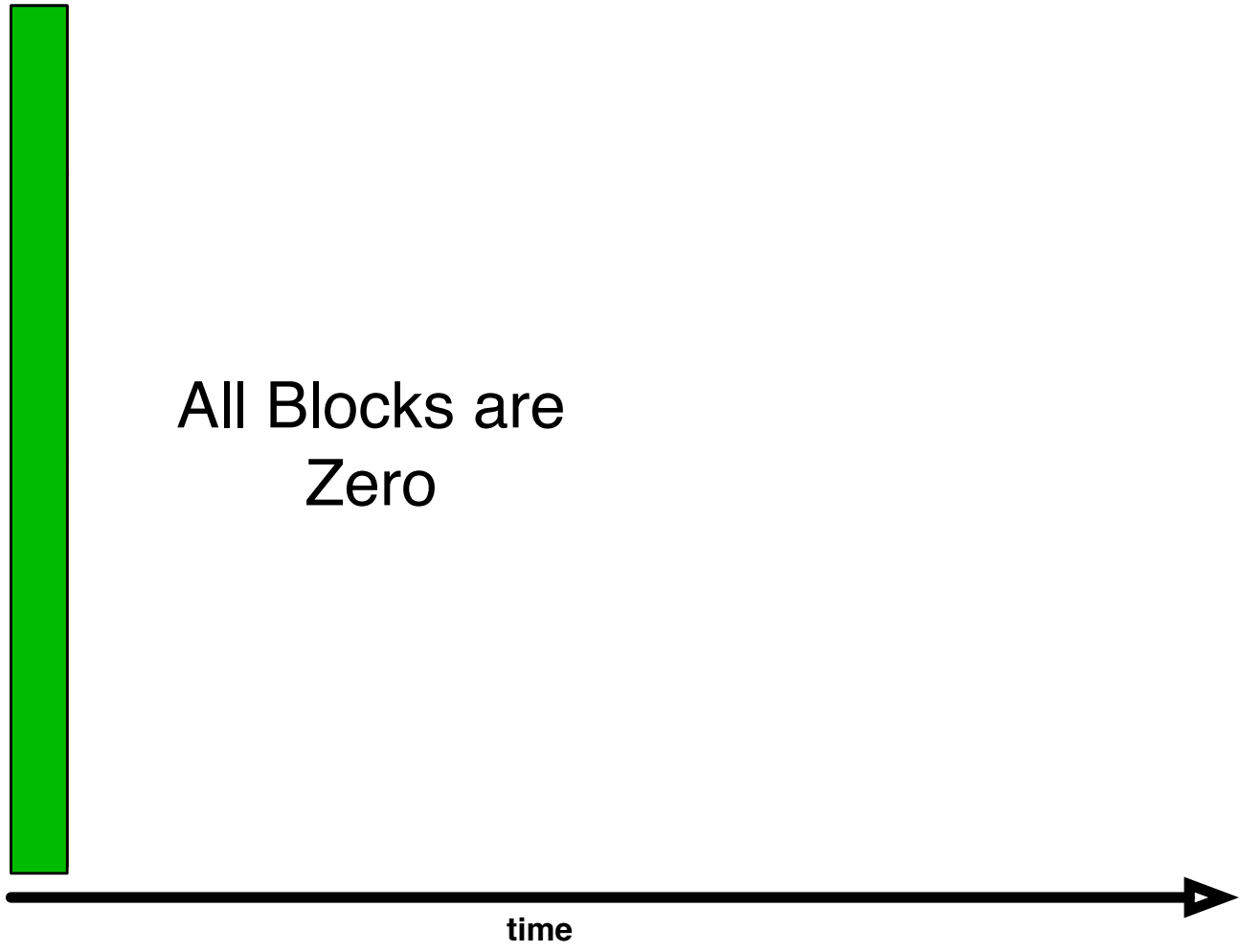


The green sectors indicate sectors that were never used (or that were wiped clean).

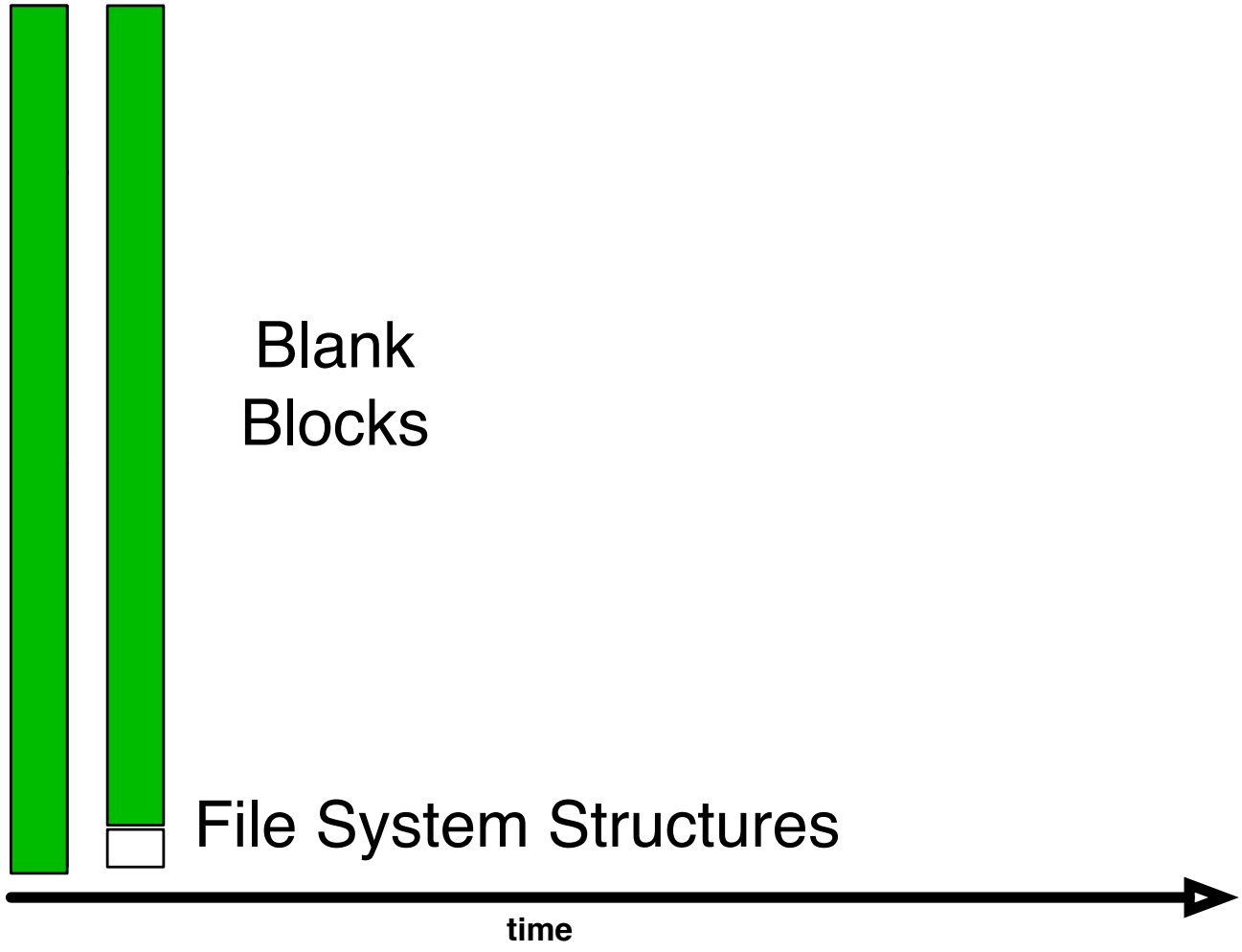
Stack the disk sectors:



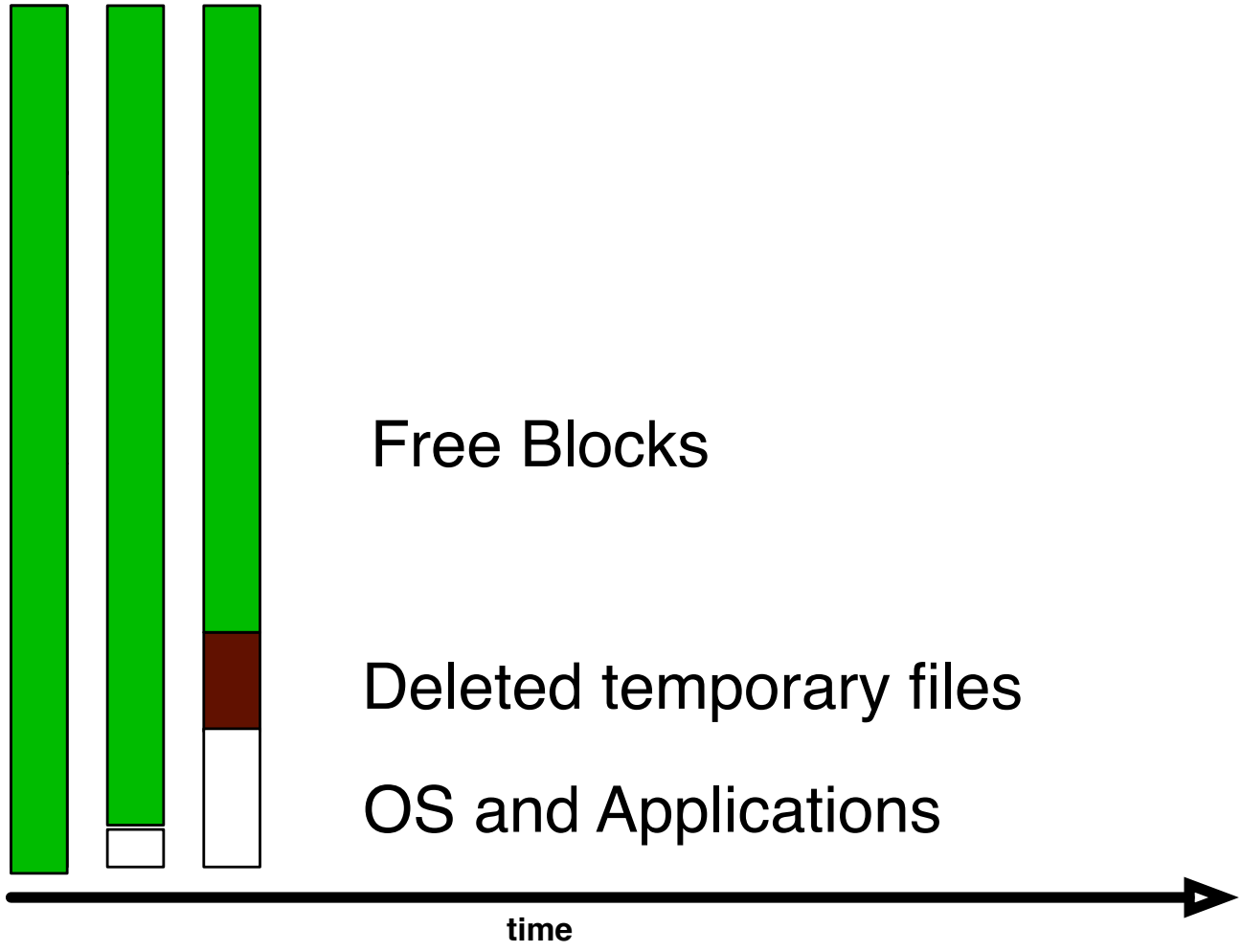
NO DATA: The disk is factory fresh.



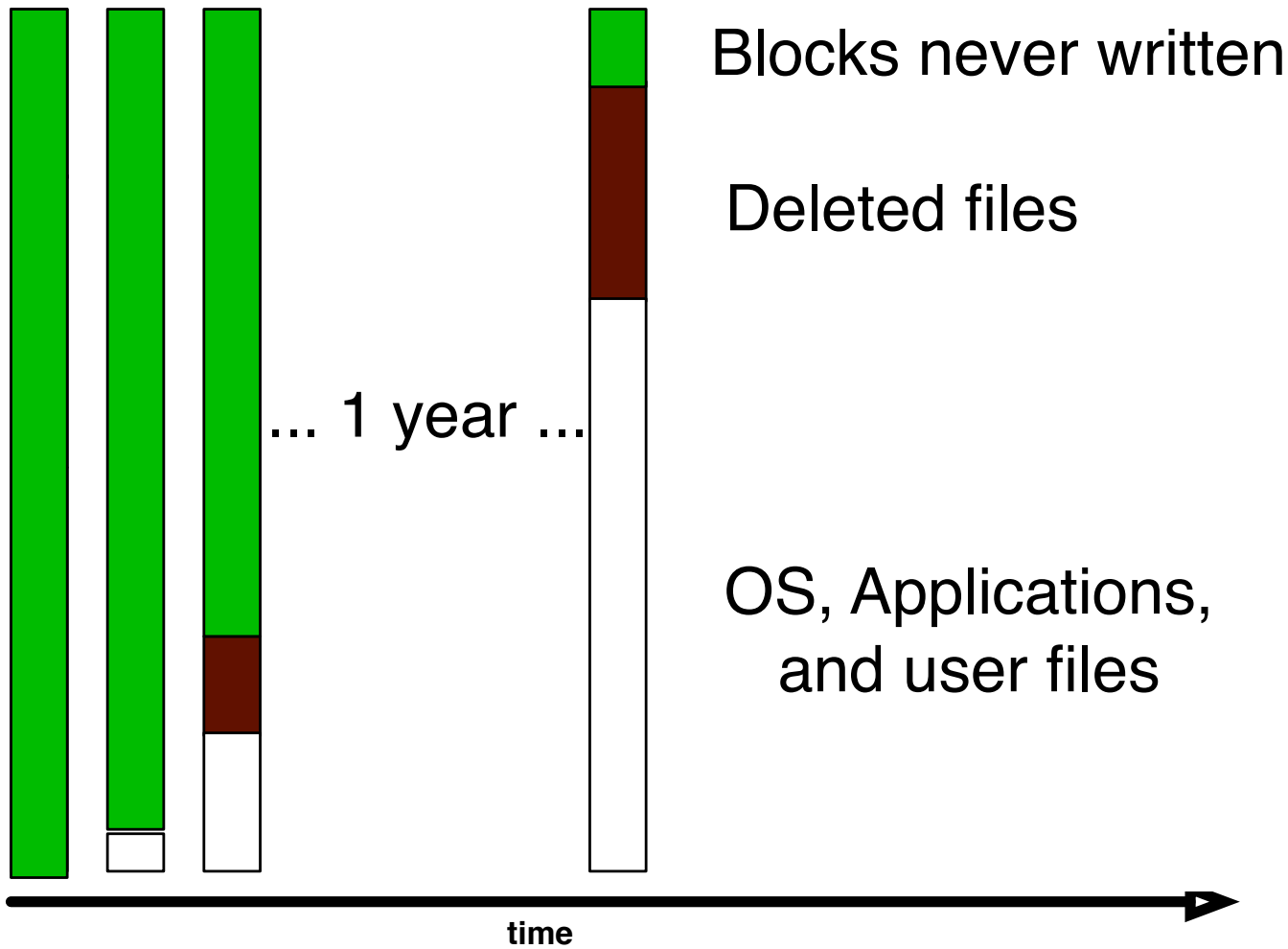
FORMATTED: The disk has an empty file system



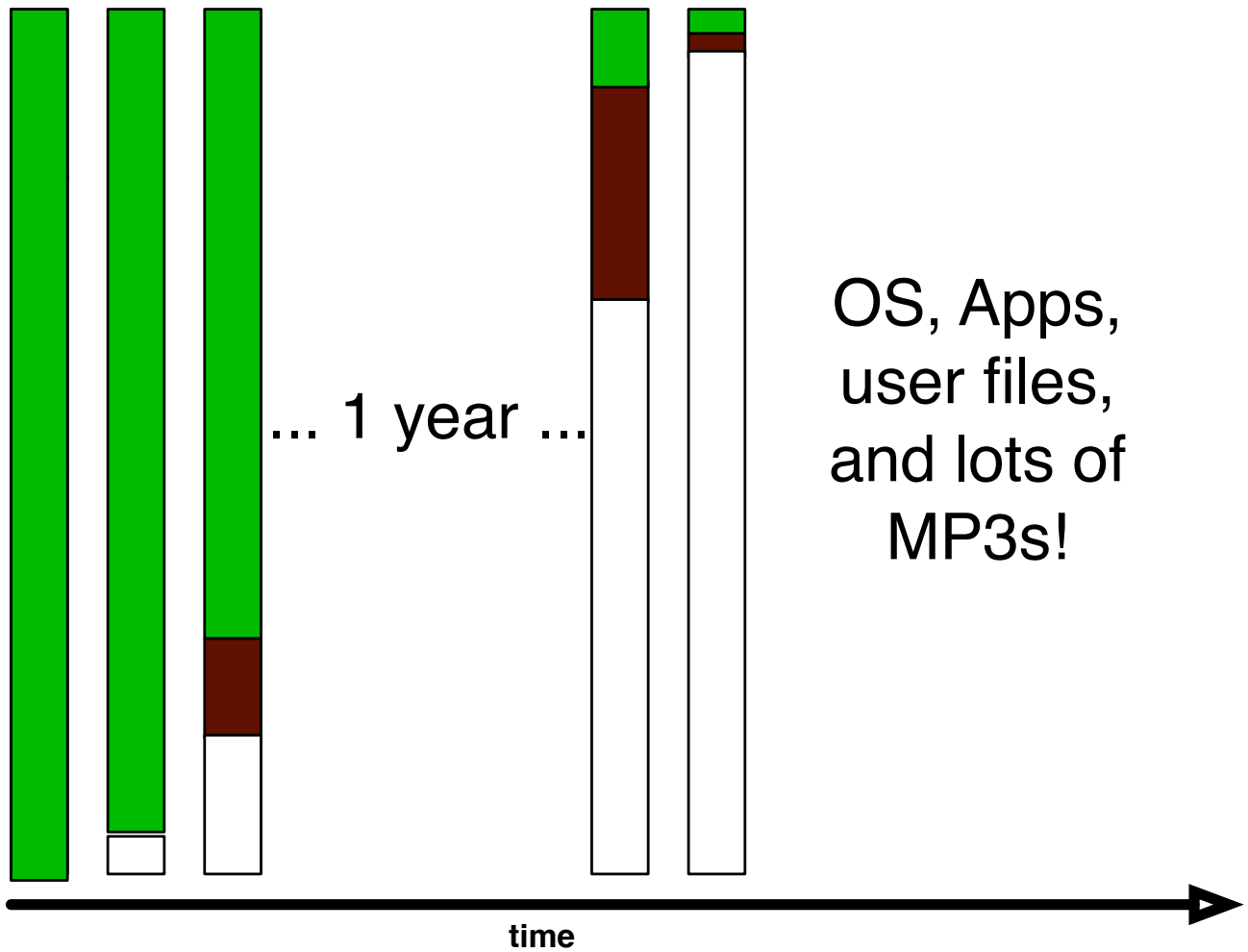
AFTER OS INSTALL: Temp. files have been deleted



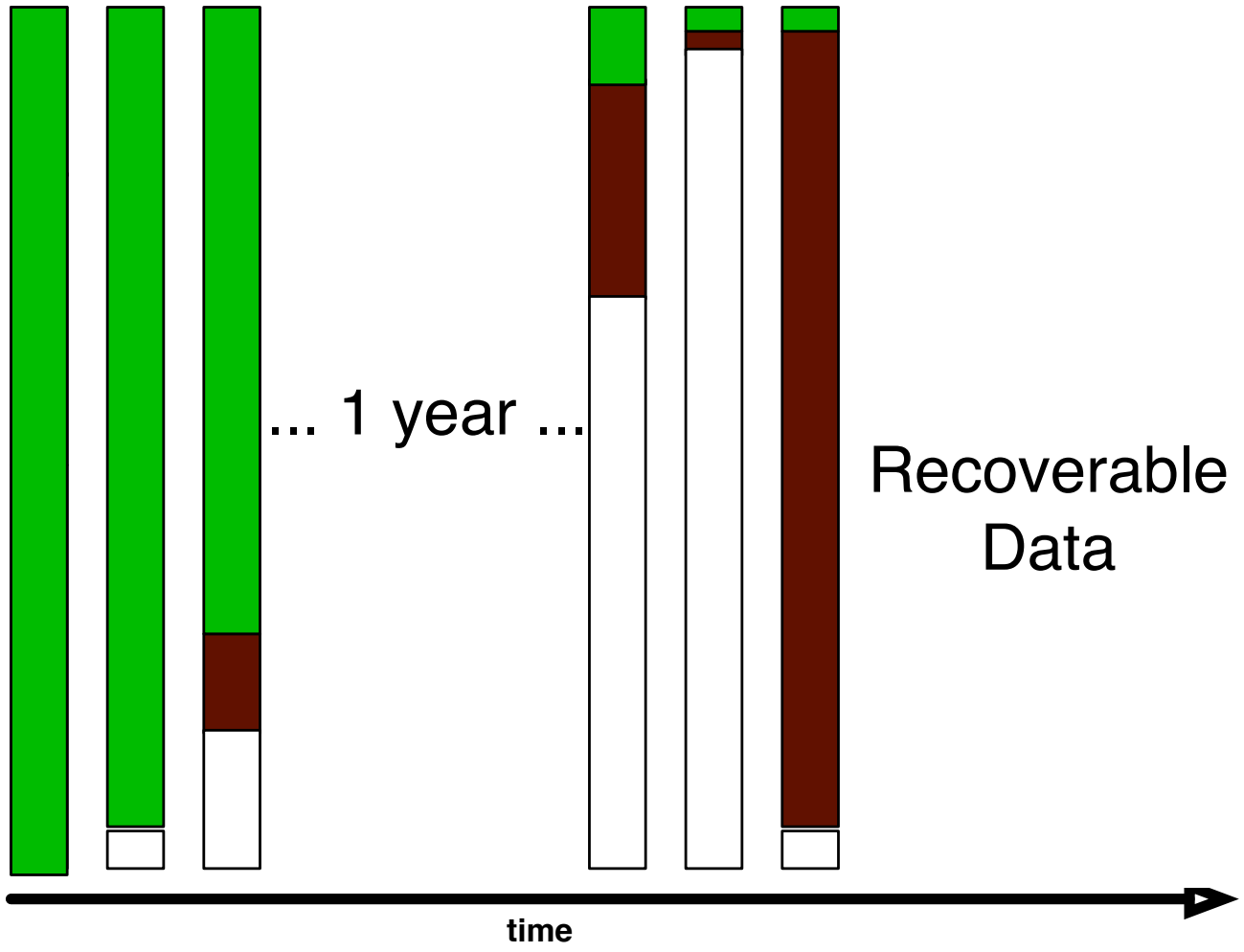
AFTER A YEAR OF SERVICE



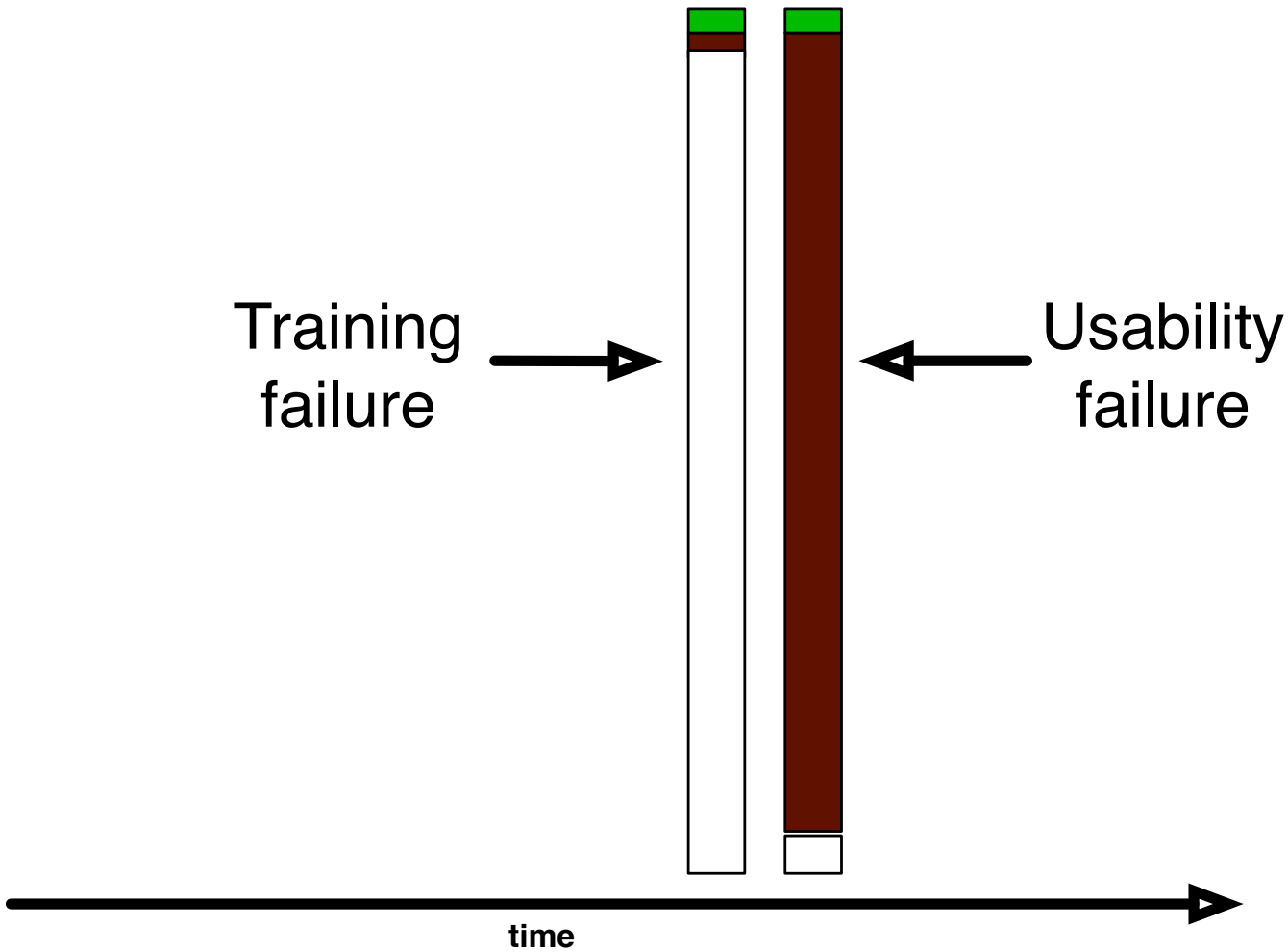
DISK NEARLY FULL!



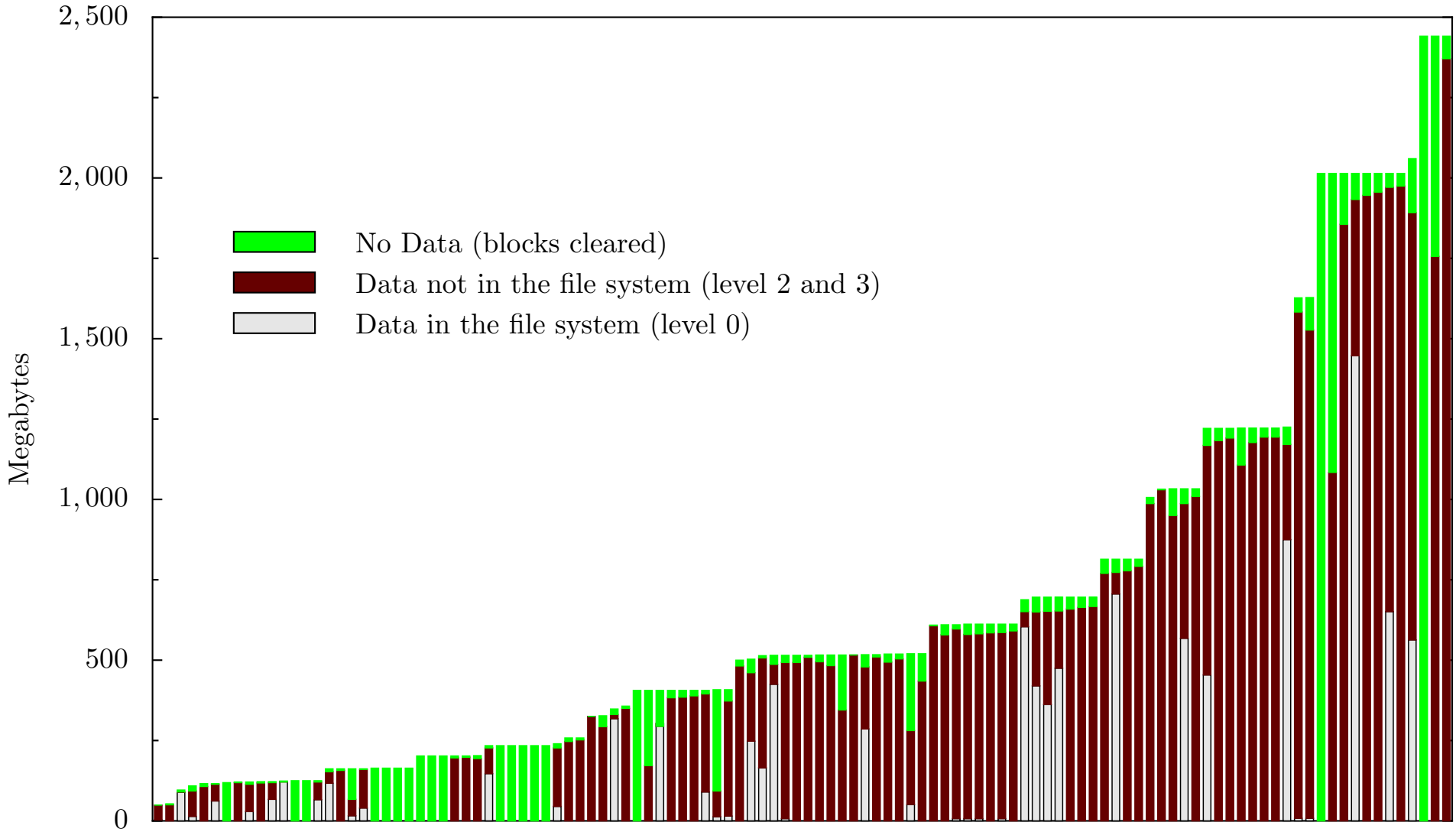
FORMAT C:\ (to sell the computer.)



We can use forensics to reconstruct motivations:



The 236 drives are dominated by failed sanitization attempts.



But training failures are also important.

Overall numbers

Drives Acquired:	236
Drives DOA:	60
Drives Images:	176
Drives Zeroed:	11
Drives “Clean Formatted:”	22
Total files:	168,459
Total data:	125G

“Clean Formatted”

Easily identified with SQL:

- ```
  img_blocks > 0
  and img_blocks != img_zblocks
  and img_blocks*.01 > img_zblocks
  ```

22 drives were “clean formatted.”

- 1 from Driveguys (but other 2 had lots of data)
- 18 from pcjunkyard (out of 25; 1 had parish data)
- 1 from Mr. M. who sold his 2GB drive on eBay.
- 1 from a VA reseller (1 DOA; 3 dirty formats)
- 1 from an unknown source (1 DOA; 1 dirty format)
  ```

MD5 factoring allows the rapid identification of files by their fingerprints.

This allows quick determination of:

- Unique files
- Operating system files
- Most common files

Coming soon: factoring sectors.

- A 60GB file has 3.6GB of MD5 codes...
- May require a specialized database.

We found relatively few unique files that had not been deleted.

Microsoft Word files:	783
Microsoft Excel files:	184
Microsoft PowerPoint files:	30
Outlook PST files:	11
audio files:	977

Why were there so few unique files?

But what *really* happened?

To answer this question, I needed to contact the original drive owners.

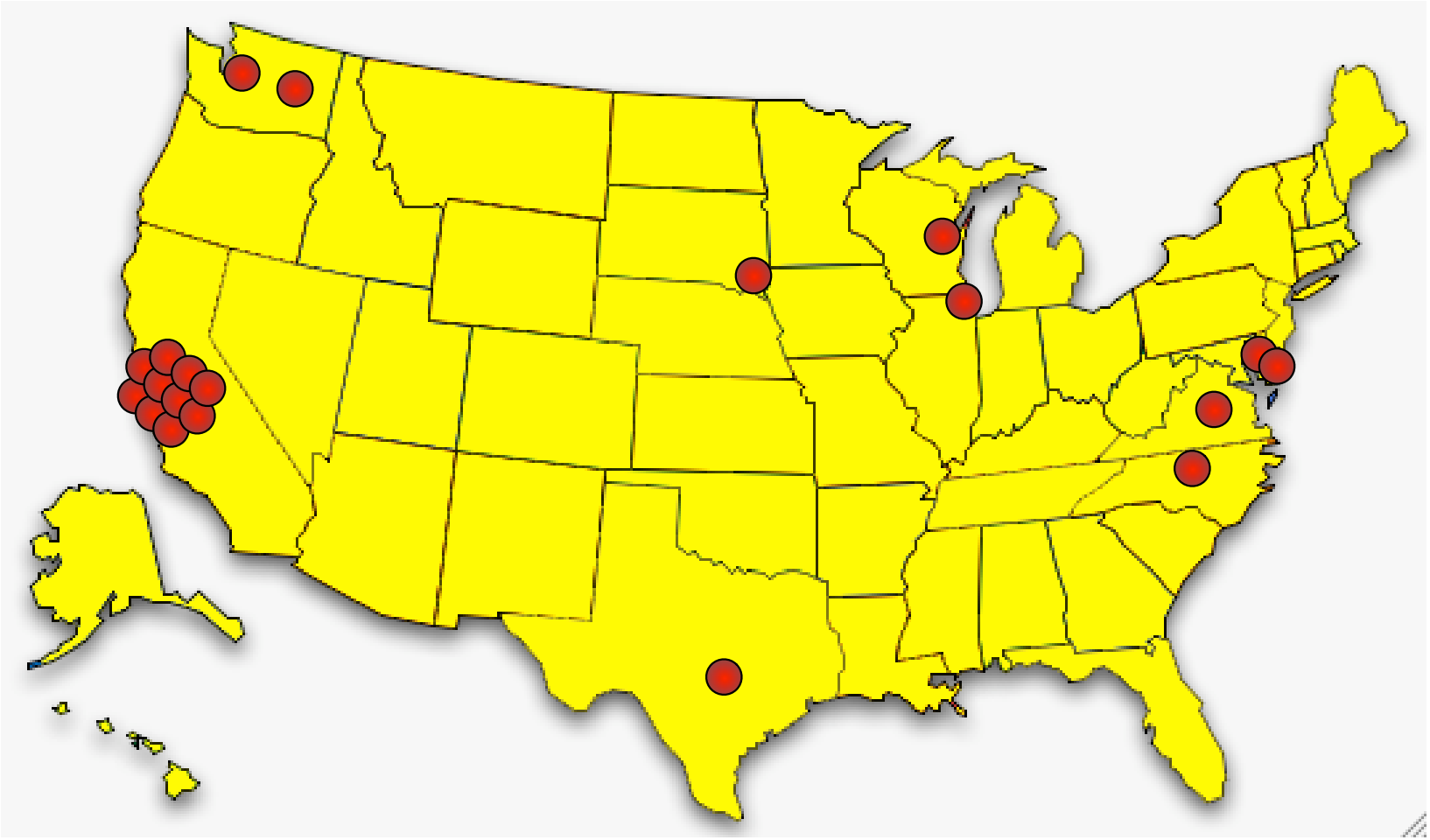
The *Remembrance of Data Passed Traceback Study*.

1. Find data on hard drive
2. Determine the owner
3. Get contact information for organization
4. Find the right person *inside* the organization
5. Set up interviews
6. Follow guidelines for human subjects work

```
06/19/1999 /:dir216/Four H Resume.doc
03/31/1999 /:dir216/U.M. Markets & Society.doc
08/27/1999 /:dir270/Resume-Deb.doc
03/31/1999 /:dir270/Deb-Marymount Letter.doc
03/31/1999 /:dir270/Links App. Ltr..doc
08/27/1999 /:dir270/Resume=Marymount U..doc
03/31/1999 /:dir270/NCR App. Ltr..doc
03/31/1999 /:dir270/Admissions counselor, NCR.doc
08/27/1999 /:dir270/Resume, Deb.doc
03/31/1999 /:dir270/UMUC App. Ltr..doc
03/31/1999 /:dir270/Ed. Coordinator Ltr..doc
03/31/1999 /:dir270/American College ...doc
04/01/1999 /:dir270/Am. U. Admin. Dir..doc
04/05/1999 /:dir270/IR Unknown Lab.doc
04/06/1999 /:dir270/Admit Slip for Modernism.doc
04/07/1999 /:dir270/Your Honor.doc
```

This was a lot harder than I thought it would be.

Ultimately, I contacted 20 organizations between April 2003 and April 2005.



The leading cause of compromised privacy was betrayed trust.

Trust Failure: 5 cases

- ✔ Home computer; woman's son took to "PC Recycle"
- ✔ Community college; no procedures in place
- ✔ Church in South Dakota; administrator "kind of crazy"
- ✔ Auto dealership; consultant sold drives he "upgraded"
- ✔ Home computer, financial records; same consultant

This specific failure wasn't considered in [GS 03]; it was the most common failure.

Poor training or supervision was the second leading cause.

Trust Failure: 5 cases

Lack of Training: 3 cases

- ✔ California electronic manufacturer
- ✔ Supermarket credit-card processing terminal
- ✔ ATM machine from a Chicago bank

Alignment between the interface and the underlying representation would overcome this problem.

In two cases, the data custodians simply didn't care.

Trust Failure: 5 cases
Lack of Training: 3 cases

Lack of Concern: 2 cases

- ✔ Bankrupt Internet software developer
- ✔ Layoffs at a computer magazine

Regulation on resellers might have prevented these cases.

In seven cases, no cause could be determined.

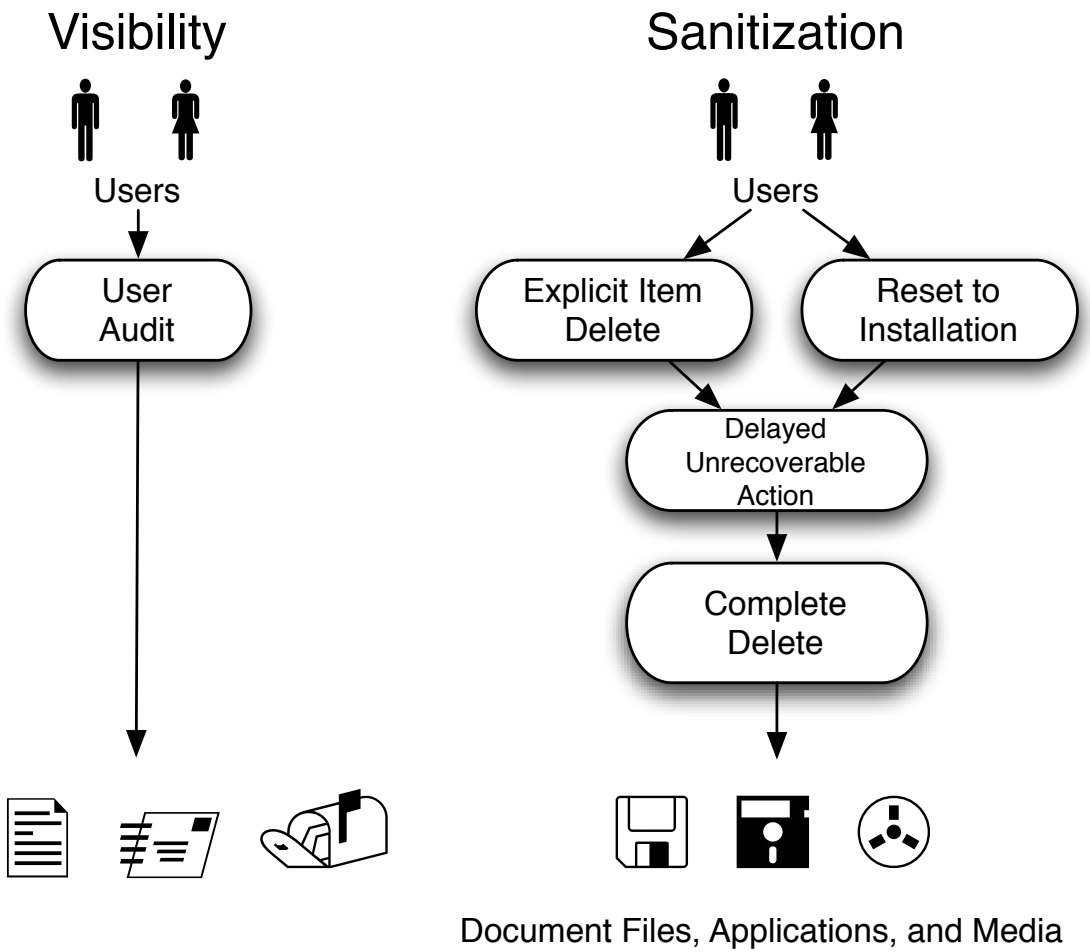
Trust Failure: 5 cases
Lack of Training: 3 cases
Lack of Concern: 2 cases

Unknown Reason: 7 cases

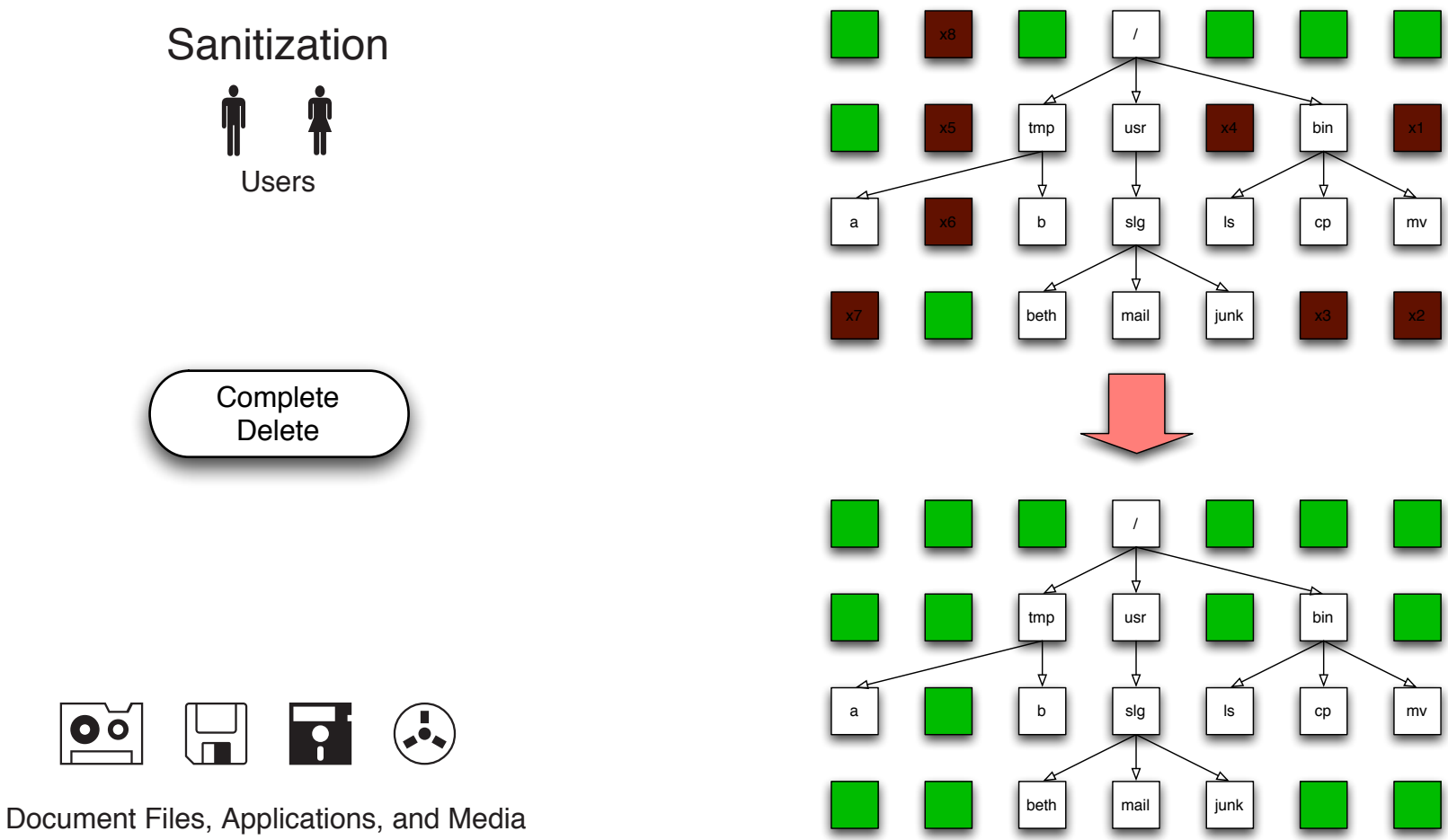
- ✘ Bankrupt biotech startup
- ✘ Another major electronics manufacturer
- ✘ Primary school principal's office
- ✘ Mail order pharmacy
- ✘ Major telecommunications provider
- ✘ Minnesota food company
- ✘ State Corporation Commission

Regulation might have helped here, too.

I have identified five distinct patterns for addressing the sanitization problem.

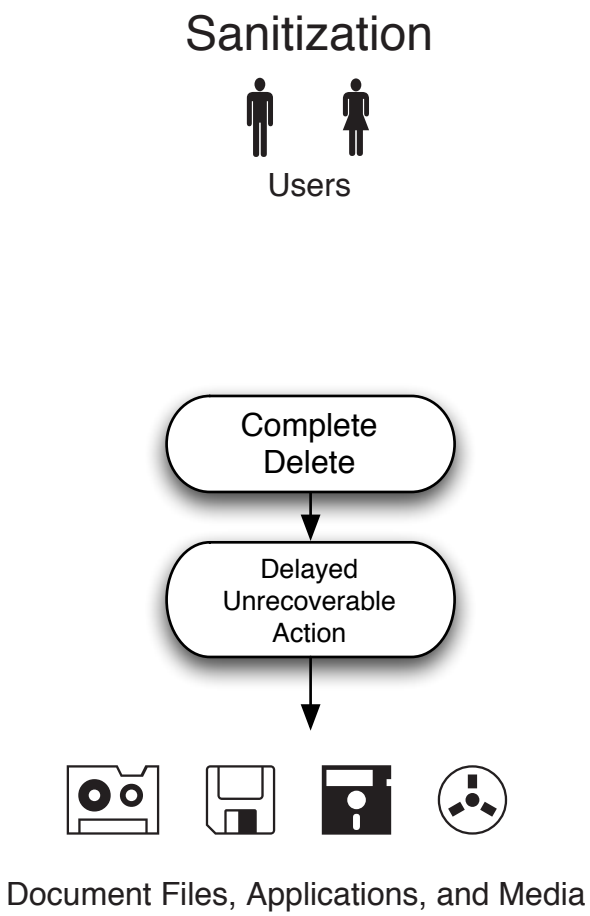


Complete Delete: assure that deleting the *visible* representation deletes the *hidden* data as well.



Naming this pattern lets us discuss its absence in modern operating systems.

Delayed Unrecoverable Action: give the users a chance to change their minds.



[Norman 83] and [Cooper 99] both suggest this functionality, but they do not name or integrate it.

Legislative reactions to this research: "Fair and Accurate Credit Transactions Act of 2003" (US)

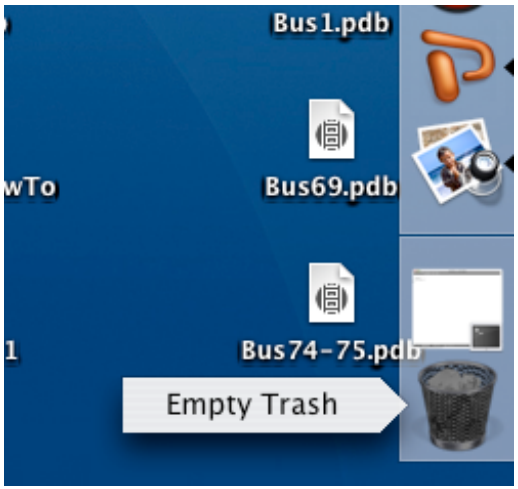
- Introduced in July 2003. Signed December 2003.
- Regulations adopted in 2004, effective June 2005.
- Amends the FCRA to standardize consumer reports.
- Requires destruction of paper or electronic "consumer records."

Testimony: `http://tinyurl.com/cd2my`

Technical reactions to this research: “Secure Empty Trash” in MacOS 10.3.

Unfortunately, “Secure Empty Trash” is incomplete.

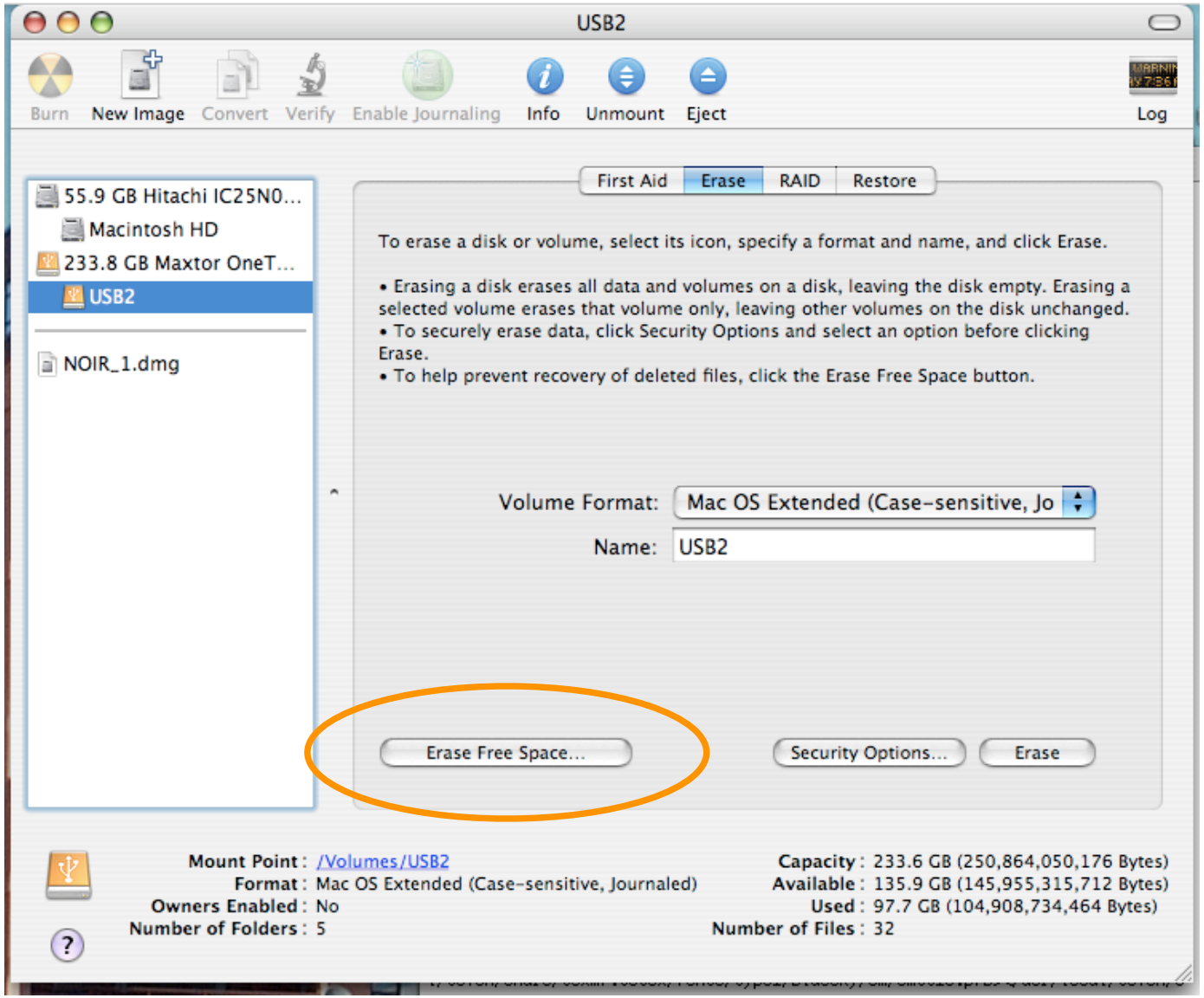
- Implemented in Finder (inconsistently)
- Locks trash can
- Can’t change your mind



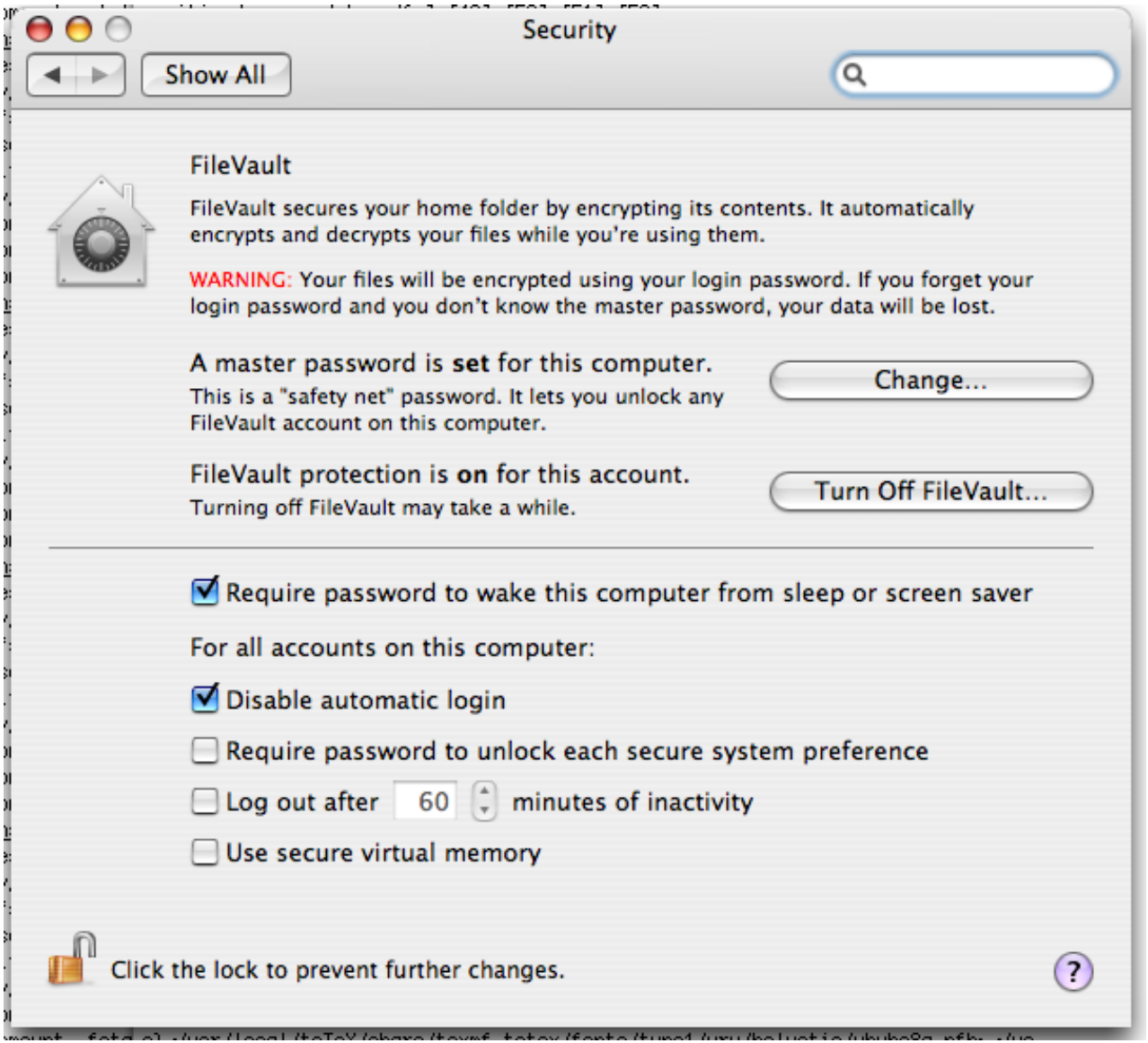
What's the difference between...

...an action that can't be undone and one that can't be recovered?

MacOS 10.4 “Erase Free Space” makes a big file.



MacOS “File Vault” gives users an encrypted file system.



The Clean Delete Agenda:

- Make FORMAT actually erase the disk.
- Make “Empty Trash” actually overwrite data.
- Integrate this functionality with web browsers, word processors, etc.

This will complicate some police investigations.

Questions?