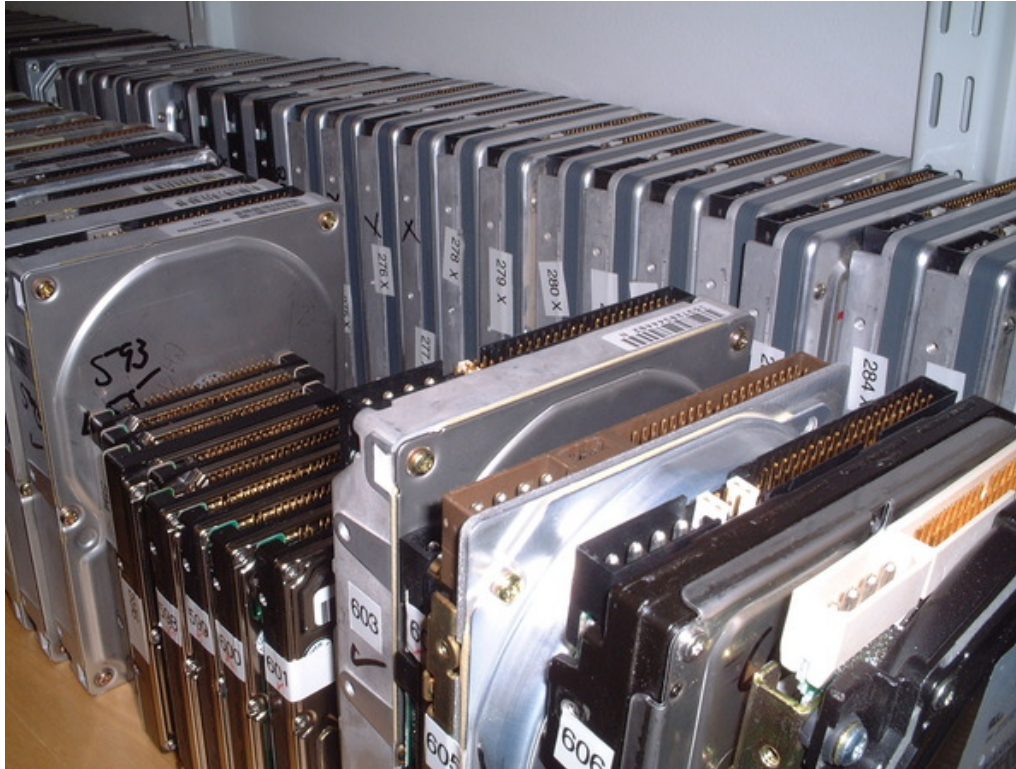


# Everything You Need to Know About the Destruction of Information on Computer Hard Drives



**Simson L. Garfinkel**

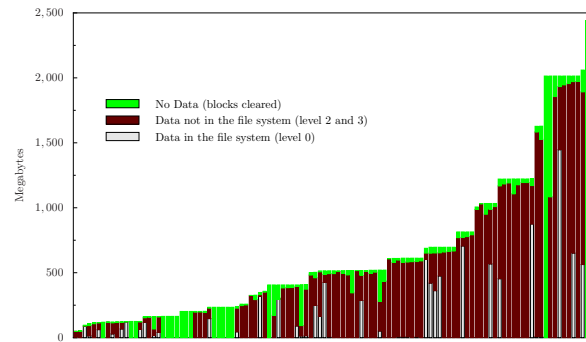
**May 16, 2006**

**1:30pm**

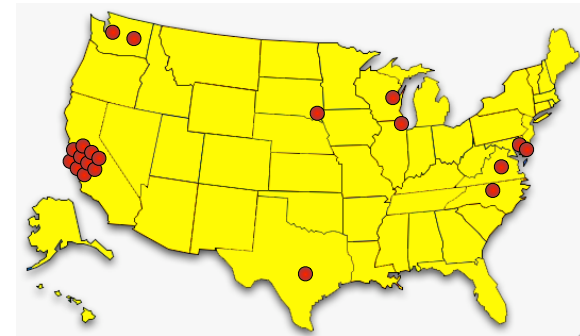
**Postdoctoral Fellow,  
Center for Research on Computation and Society  
Harvard University**

# The disk sanitization problem.

## 1. Scale of the problem



## 2. The Traceback Study



# Disk Sanitization

Recall some of the goals of computer security:

- Availability
- Confidentiality
- Data Integrity
- Control
- Audit

# Confidentiality means preventing unauthorized disclosure.

Data can be:

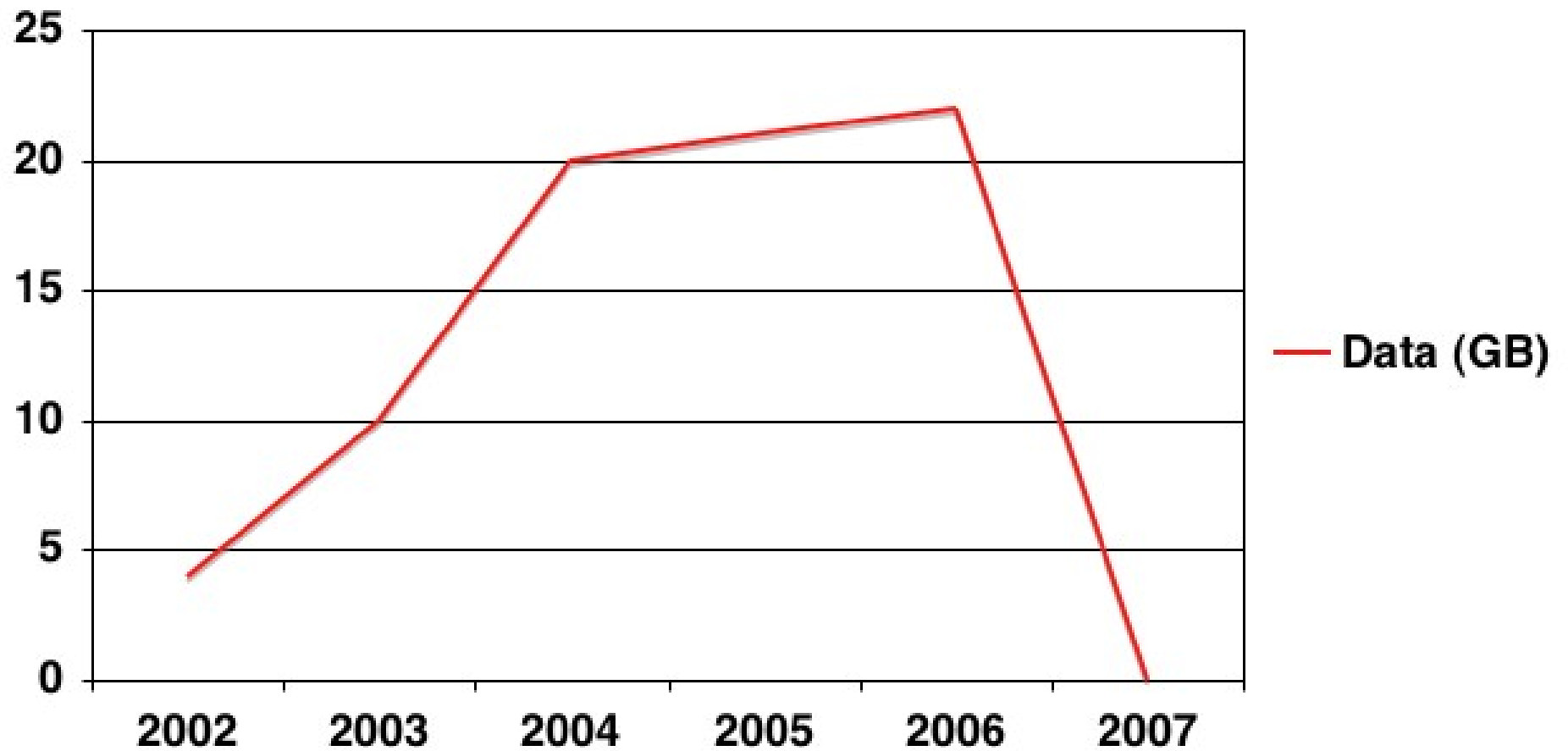
- In flight
- Stored



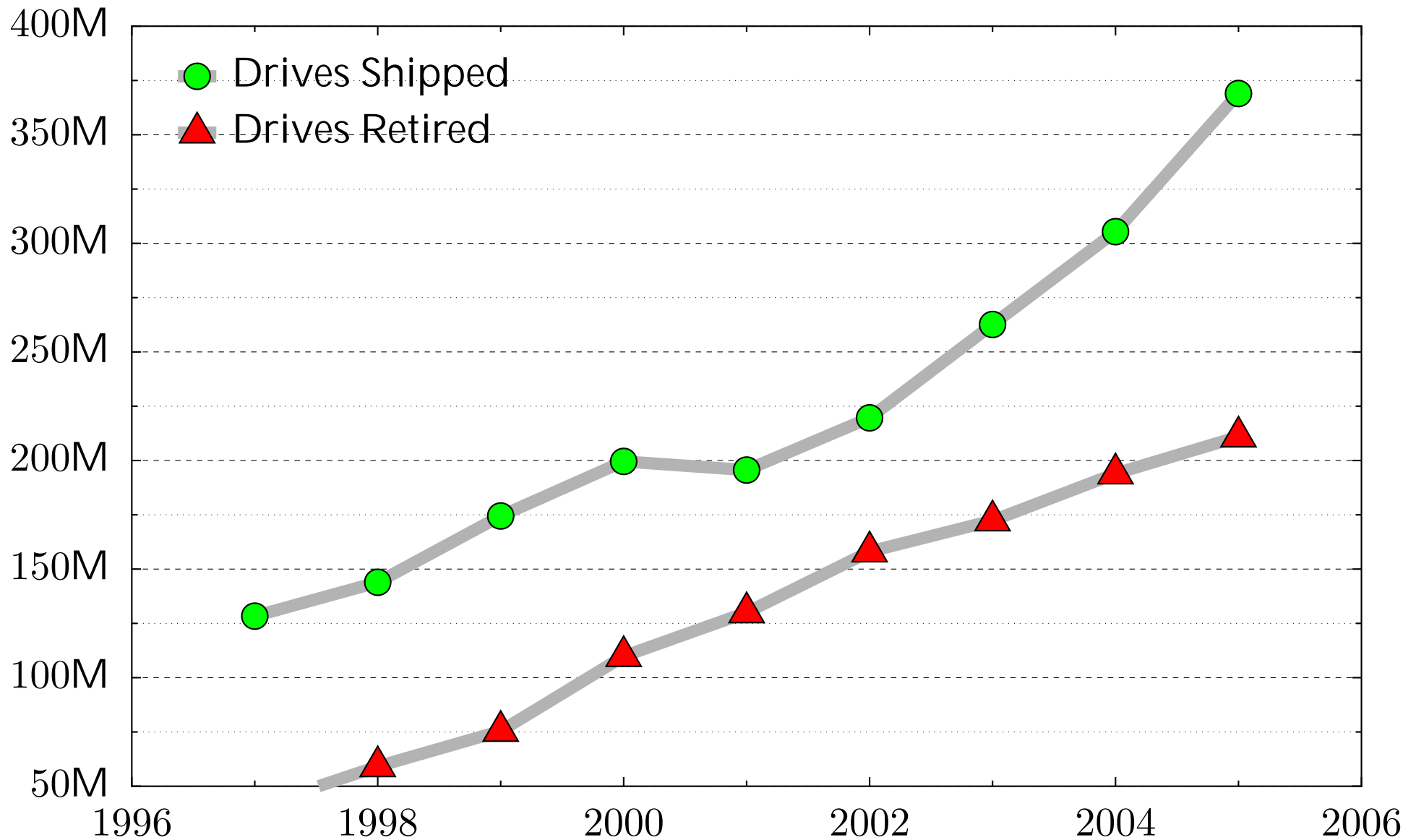
**Most data spends most of its time in storage.**



## Data over time: Conceptual



**210 million drives will be retired this year.**



**“Retire?”**



**Deckard (Harrison Ford) retiring a replicant.  
“Blade Runner” (1982)**

# Hard drives pose special problem for computer security

Do not forget data when power is removed.

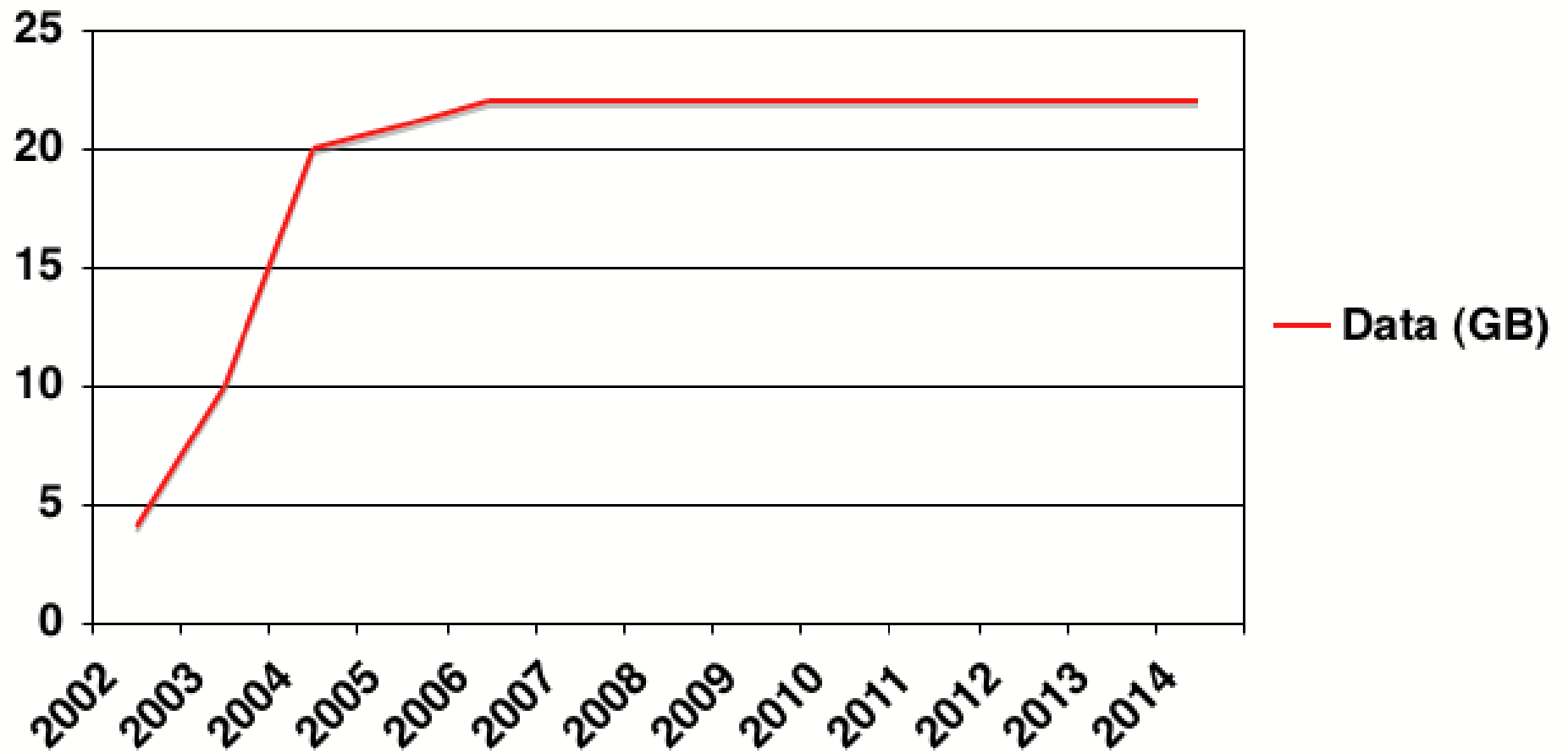
Contain data that is not immediately visible.

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



## Data over time: Actual



# There is a significant secondary market for used disk drives.



Retired drives are:

- Re-used within organizations
- Given to charities
- Sold at auction

**All Categories** [Save this search](#)  
350 items found for hard drives  
Sort by items: [ending first](#) | [newly listed](#) | [lowest priced](#) | [highest priced](#)

Picture Size	Item Title	Price	Bids	Time Left
	<a href="#">Lot of hard and floppy drives</a>	\$5.50	2	14m
	<a href="#">Lot of hard and floppy drives</a>	\$5.50	2	22m
	<a href="#">Lot of hard and floppy drives</a>	\$5.50	2	25m
	<a href="#">Lot of 2 hard drives IDE</a>	\$8.00	12	29m
	<a href="#">3.2 gig Hard Drives</a>	\$180.00	-	59m
	<a href="#">(5) 1.2 hard drives &amp; (15) 10/100 network</a>	\$15.00	1	1h 00m
	<a href="#">Lot of 3 Quantum 9.1 gig SCSI Hard Drives</a>	\$16.00	6	1h 25m
	<a href="#">IDE HARD DRIVES (3)</a>	\$6.50	6	1h 46m
	<a href="#">LOT OF 5 Hard Drives! 3.2 Gig Western Digital</a>	\$120.00 \$124.95 <del>78% off</del>	-	1h 50m
	<a href="#">QTY 3... IDE Hard Drives 2.5 Gg</a>	\$10.50	5	2h 02m
	<a href="#">5 WESTERN DIGITAL 2.5 GIG HARD DRIVES</a>	\$30.00	4	2h 03m
	<a href="#">QTY 3... IDE Hard Drives 1.0 Gg</a>	\$9.99	1	2h 04m
	<a href="#">Western Digital 850 meg IDE Hard Drives dutch</a>	\$6.00	1	2h 57m
	<a href="#">WINDOWS</a>	\$6.00	-	3h 18m

About 1000 used drives/day sold on eBay.

**Today there are three primary techniques for assuring data confidentiality.**

1. Physical security.
2. Logical access controls. (operating system)
3. Cryptography (disk & link)

**When a disk is thrown out or repurposed, most of these techniques don't work.**

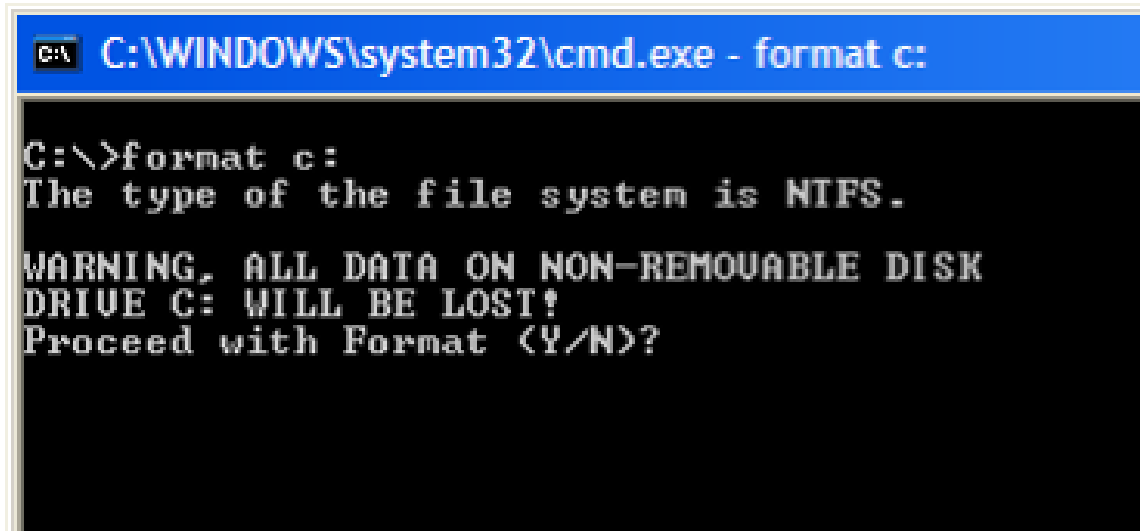
1. ~~Physical security~~
2. ~~Logical access controls (operating system)~~
3. Cryptography (disk & link)

**And most people don't encrypt their data.**



## More bad news.

FORMAT C: doesn't erase the hard drive.



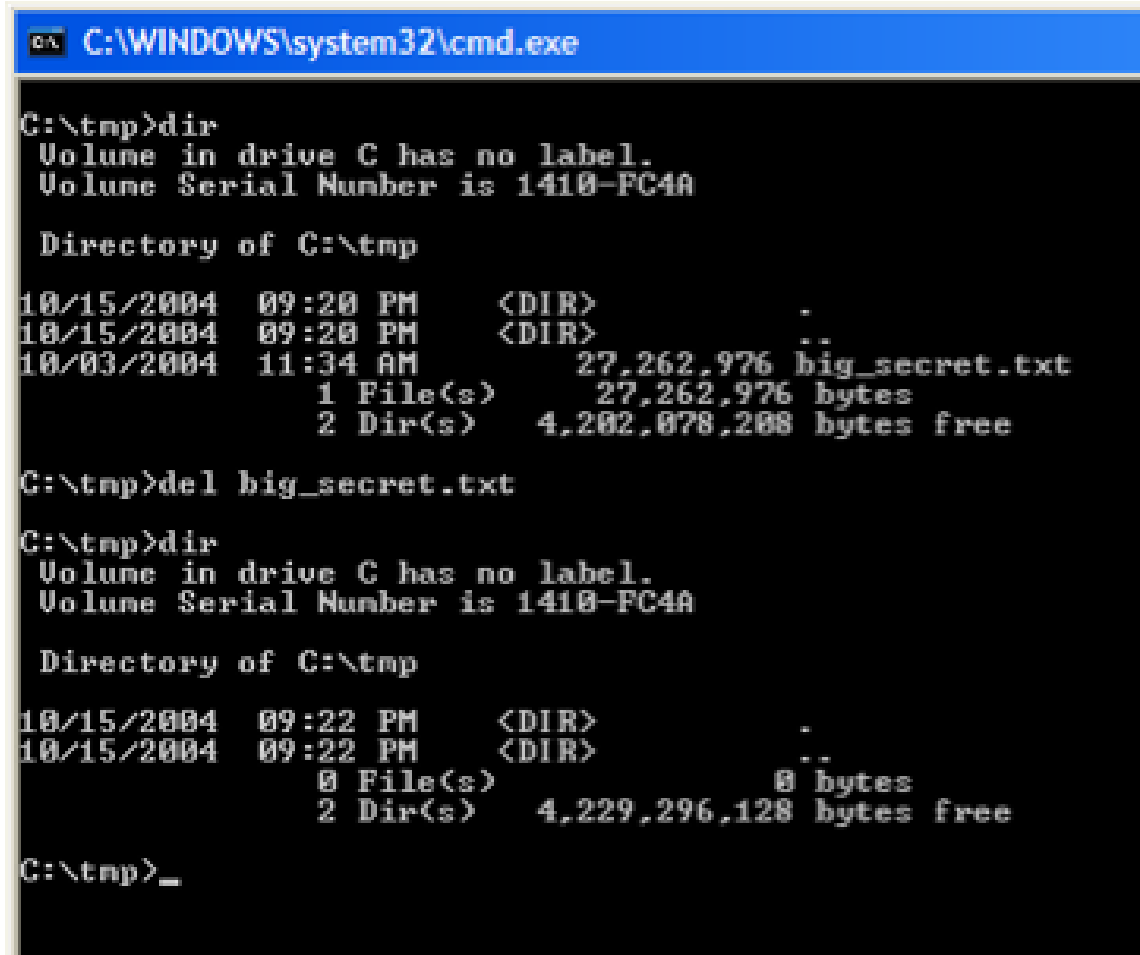
```
C:\WINDOWS\system32\cmd.exe - format c:

C:\>format c:
The type of the file system is NTFS.

WARNING, ALL DATA ON NON-REMOVABLE DISK
DRIVE C: WILL BE LOST!
Proceed with Format (Y/N)?
```

**FORMAT just writes a new root directory.**

## DEL doesn't delete files



```
C:\WINDOWS\system32\cmd.exe

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

Directory of C:\tnp

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:\tnp>del big_secret.txt

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

Directory of C:\tnp

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:\tnp>_
```

**DEL simply removes the file's name from the directory.**

**These failings are shared by all modern file systems.**

- FAT12 – DOS Floppy disks
- FAT16, FAT32 – DOS, Windows, USB Drives
- NTFS – Windows NT/XP/Longhorn
- UFS, FFS, EXT2/3 – Unix
- HFS, HFS+ – MacOS
- Novell

**Compressed and Encrypted file systems complicate recovery of data.**

# A typical hard disk

Factory-Fresh Hard disk: All Blank

0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

Each block is  
512 bytes

A 20G disk has  
40M blocks.

Disk blocks (not to scale)

# “All Blank”

Each block has 512 ASCII NULs:

[illegible]

# % format C:\*

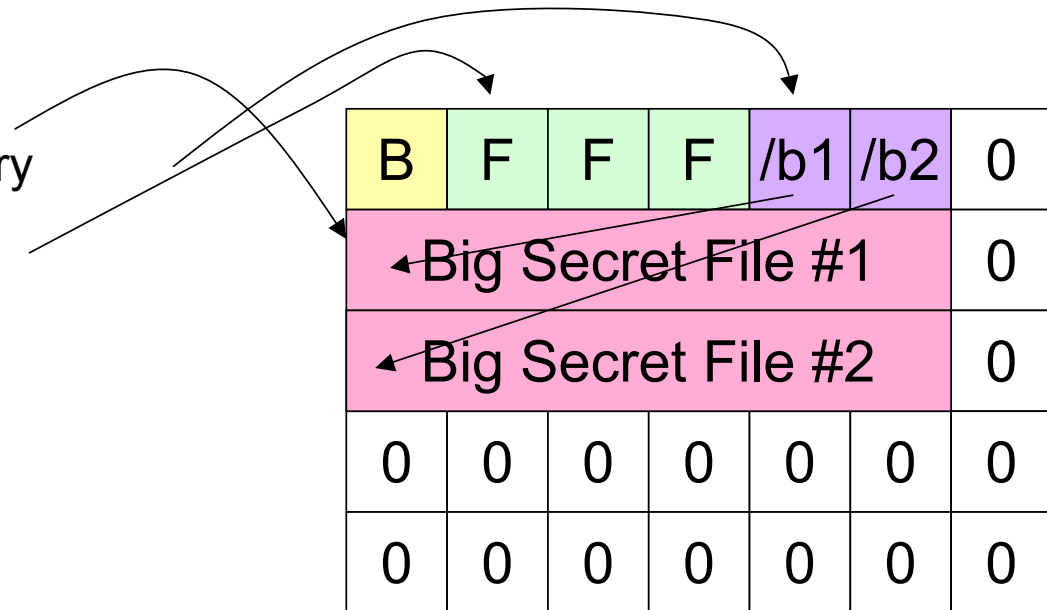
- Writes:
  - Boot blocks
  - Root directory
  - “File Allocation Table” (FAT)
  - Backup “superblocks” (UFS/FFS)
- May also:
  - Validate surface

B	F	F	F	/	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

\* Examples based on FAT32 running under Unix

```
% cp bfs1 /mnt/b1
% cp bfs2 /mnt/b2
```

- Writes:
  - File Contents
  - File Directory Entry
  - Bookkeeping



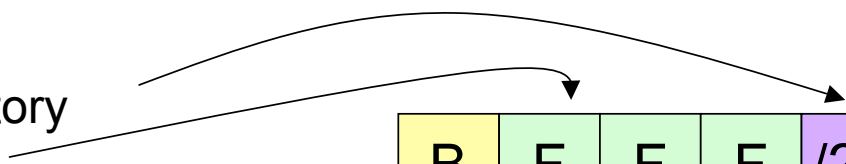
- root directory:
 

```
b1_____ ._____ jan 1 2004 block 7
b2_____ ._____ jan 1 2004 block 14
```

% rm /mnt/b1

% rm /mnt/b2

- Writes:
  - New root directory
  - Bookkeeping



B	F	F	F	/?1	/?2	0
Big Secret File #1						0
Big Secret File #2						0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

- new root directory:

?1	_____	_____	jan 1 2004	block 7
?2	_____	_____	jan 1 2004	block 14



# % cp Madonna.mp3 /mnt/mp3

- Writes:
  - New root directory
  - madonna.mp3
  - Bookkeeping

B	F	F	F	/mp3	/?2	0
Madonna et File #1						0
Big Secret File #2						0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

- new root directory:

```
Madonna_.mp3  jan 2 2004  block  7
```

```
?2_____.____  jan 1 2004  block 14
```

# What's on the disk?

- Madonna.mp3
- Madonna.mp3's directory entry
- All of B2
- Most of B2's directory entry
- Part of B1

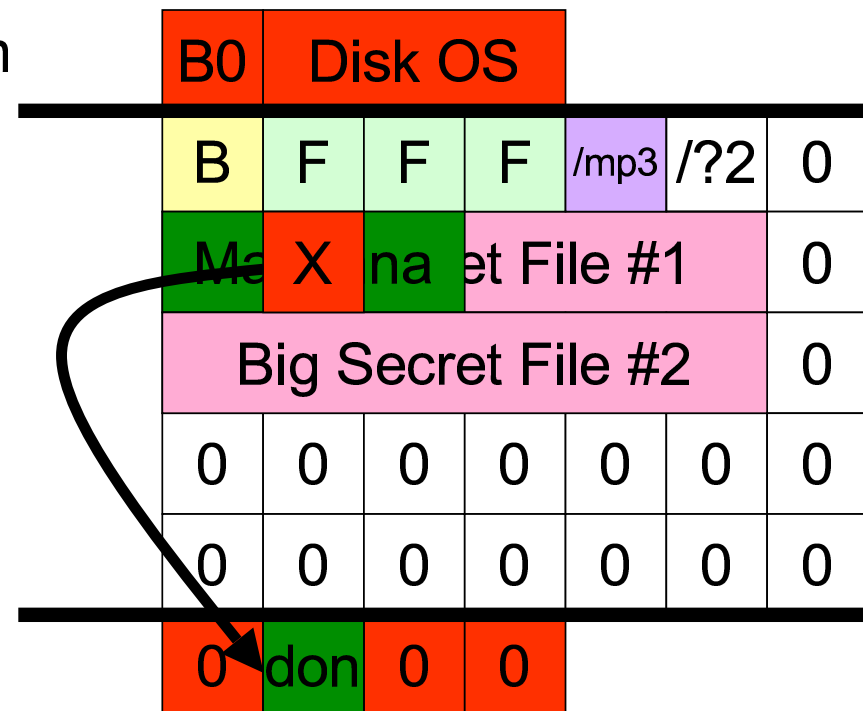
B	F	F	F	/mp3	/?2	0
Madonna			et File #1			0
Big Secret File #2						0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

# Taxonomy of hard disk data

Level 0	Files in file system
Level 1	Temp files (/tmp, /windows/tmp, etc)
Level 2	Recoverable deleted files
Level 3	Partially over-written files
Level 4	Data accessible by vendor commands
Level 5	Overwritten data

# Level 4 Data: Vendor Area

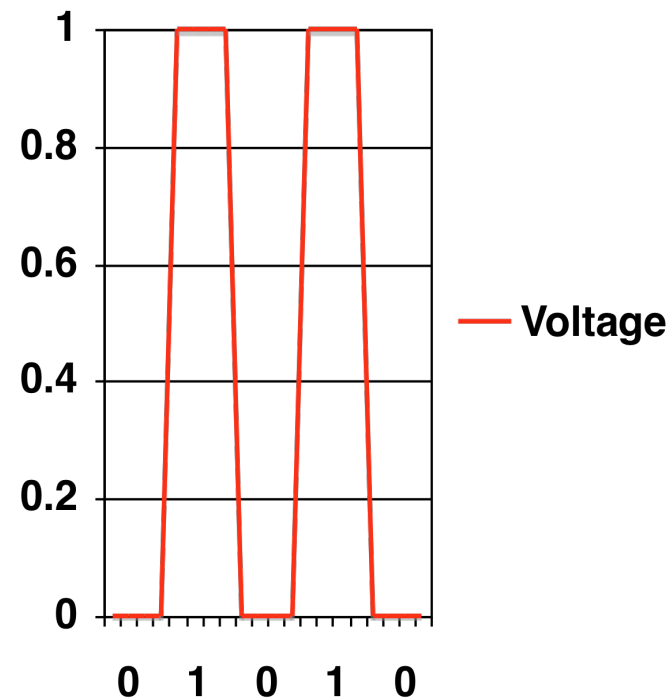
Disk operating system



Bad block regions

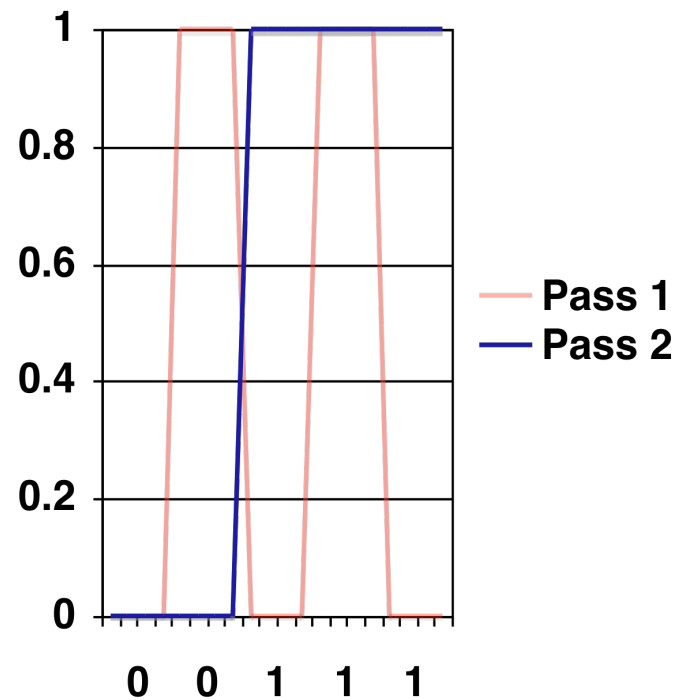
# Level 5: Overwritten Data

- Disk Drives are analog devices



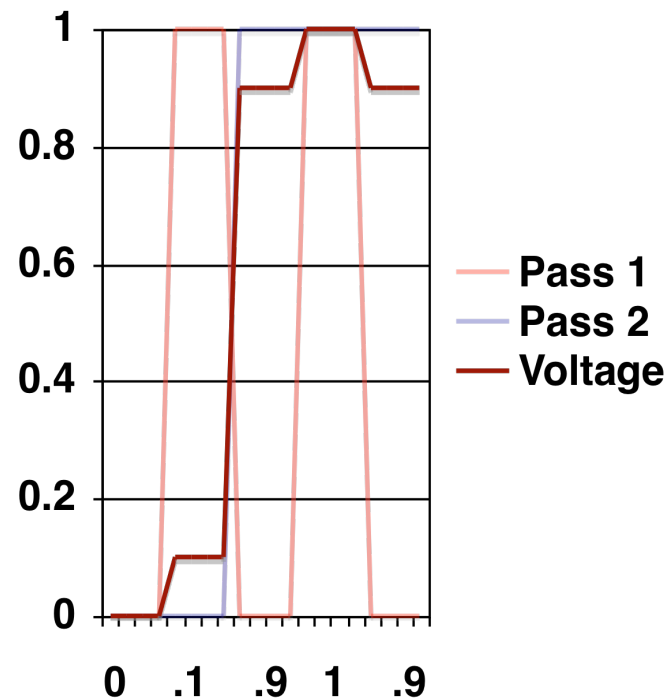
# Level 5: Overwritten Data

- Disk Drives are analog devices
- Overwritten data doesn't just die...



# Level 5: Overwritten Data

- Disk Drives are analog devices
- Overwritten data doesn't just die...
- Read data *should* be a function of all previous data values...



# Level 5: What to do?

- DOD 5220.22-M
  - “Degauss with a Type I degausser”
  - “Degauss with a Type II degausser”
  - “Overwrite all locations with a character, it’s complement, then a random character and verify”
  - Destroy, Disintegrate, incinerate, pulverize, shred, or melt



# Type 1 Degausser

- Model HD-2000
  - 73 seconds cycle time
  - 260 lbs
  - \$13,995
  - Monthly rental \$1,400
- Note:
- Your hard disk won't work after it's been degaussed (why not?)



<http://www.datadev.com/v90.htm>

# Drive Slagging

- Melting down the drives works just fine



<http://driveslag.eecue.com/>

# Drive Slagging Cont...



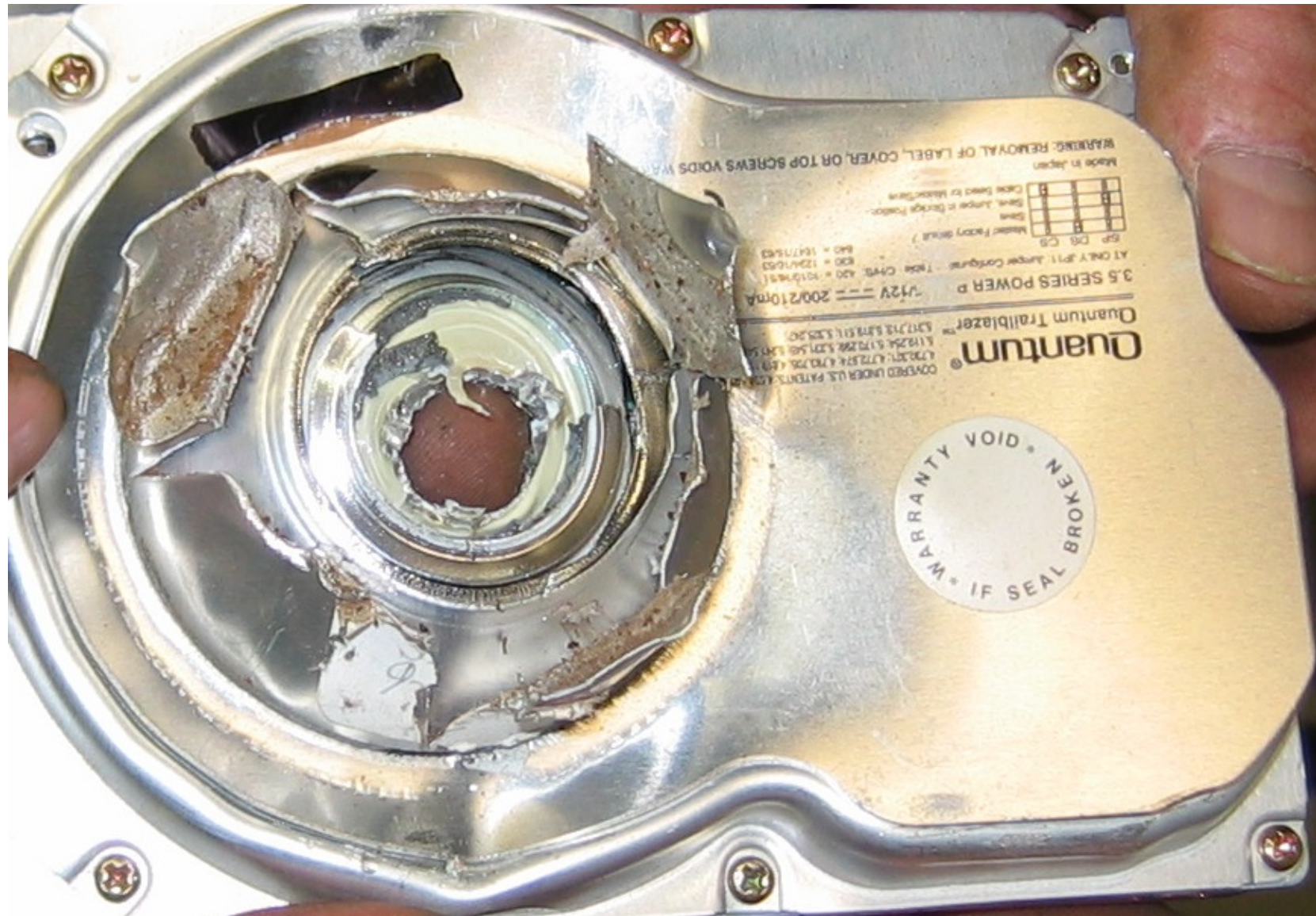
# Drive Slagging

- “Good luck removing data from this.”





**Punching a hole also works.**



**The bad news: Most people aren't using these techniques.**

**Purchased used from a computer store in August 1998:**



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

A law firm's file server...  
...with client documents!



Computers #2 through #10 had:

- Mental health records
- Home finances
- Draft of a novel...

**Was this a chance accident or common occurrence?**



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

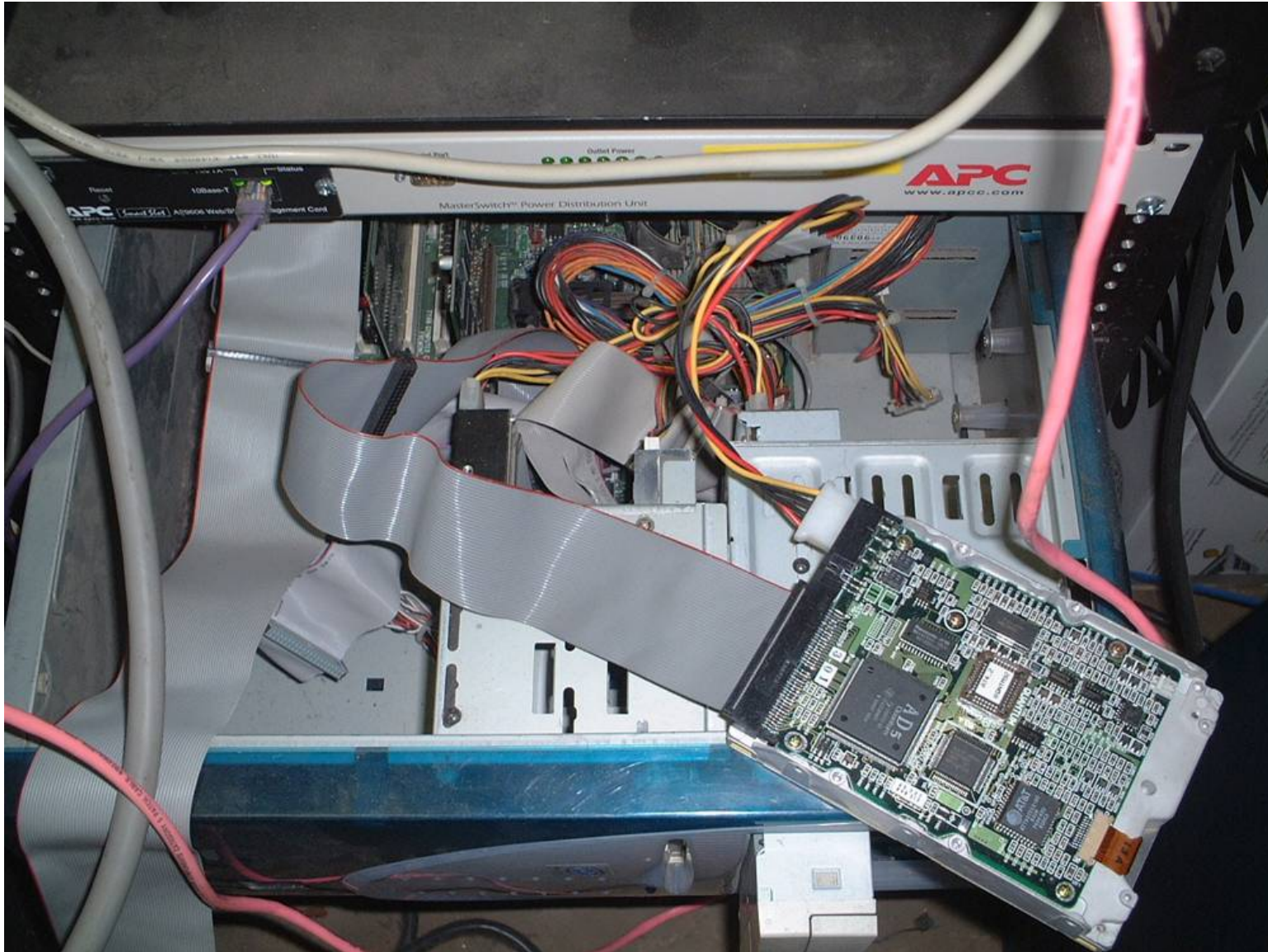




## Drives arrived by UPS



## Data on drives “imaged” using FreeBSD



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



## Images stored on a RAID



## For every drive, I cataloged:

- Disk SN, date of manufacture, etc.
- Every readable sector on the drive..
- All visible files.
- MD5 of every file.
- 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

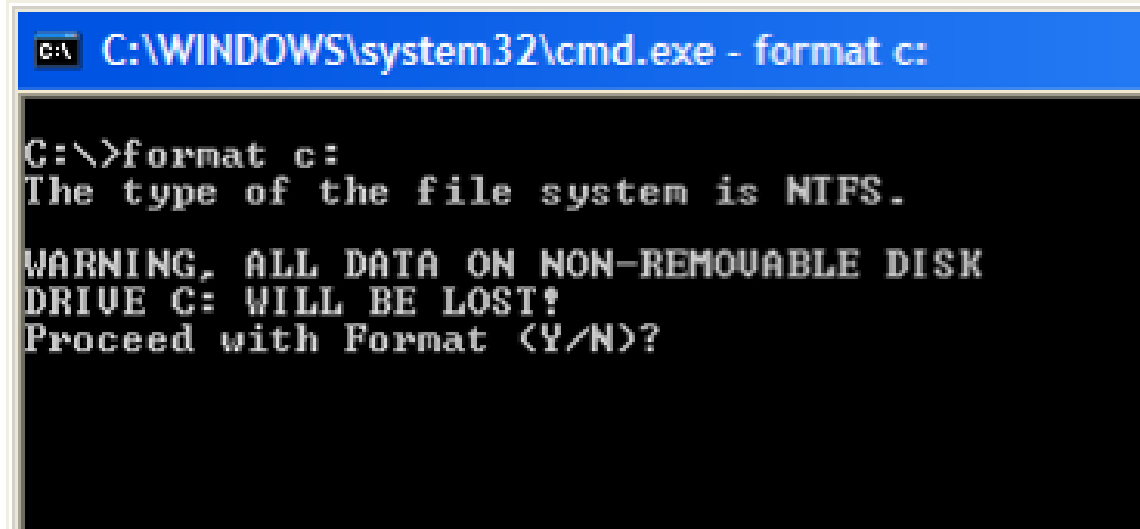
Initial analysis:

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 had been FORMATED...**



```
C:\WINDOWS\system32\cmd.exe - format c:

C:\>format c:
The type of the file system is NTFS.

WARNING, ALL DATA ON NON-REMOVABLE DISK
DRIVE C: WILL BE LOST!
Proceed with Format (Y/N)?
```

**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-limited evaluation version of the IBM AntiVirus Desktop Edition product. You may have received the Trial Edition on a promotional CD-ROM, a single-file installation program over a network. The Trial Edition is available in seven national languages, and each language is provided on a separate CC-ROM or as a separate installation program.

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 behind.**

## 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 – 200 drives.



**None of these cases are scientifically rigorous.**

## Why don't we hear more stories?

Hypothesis #1: Disclosure of “data passed” is exceedingly rare because most systems are properly cleared.

Hypothesis #2: Disclosures are so common that they are not newsworthy.

Hypothesis #3: Systems aren't properly cleared, but few people notice the data.

**I think that data left behind on hard drives is a serious social problem.**

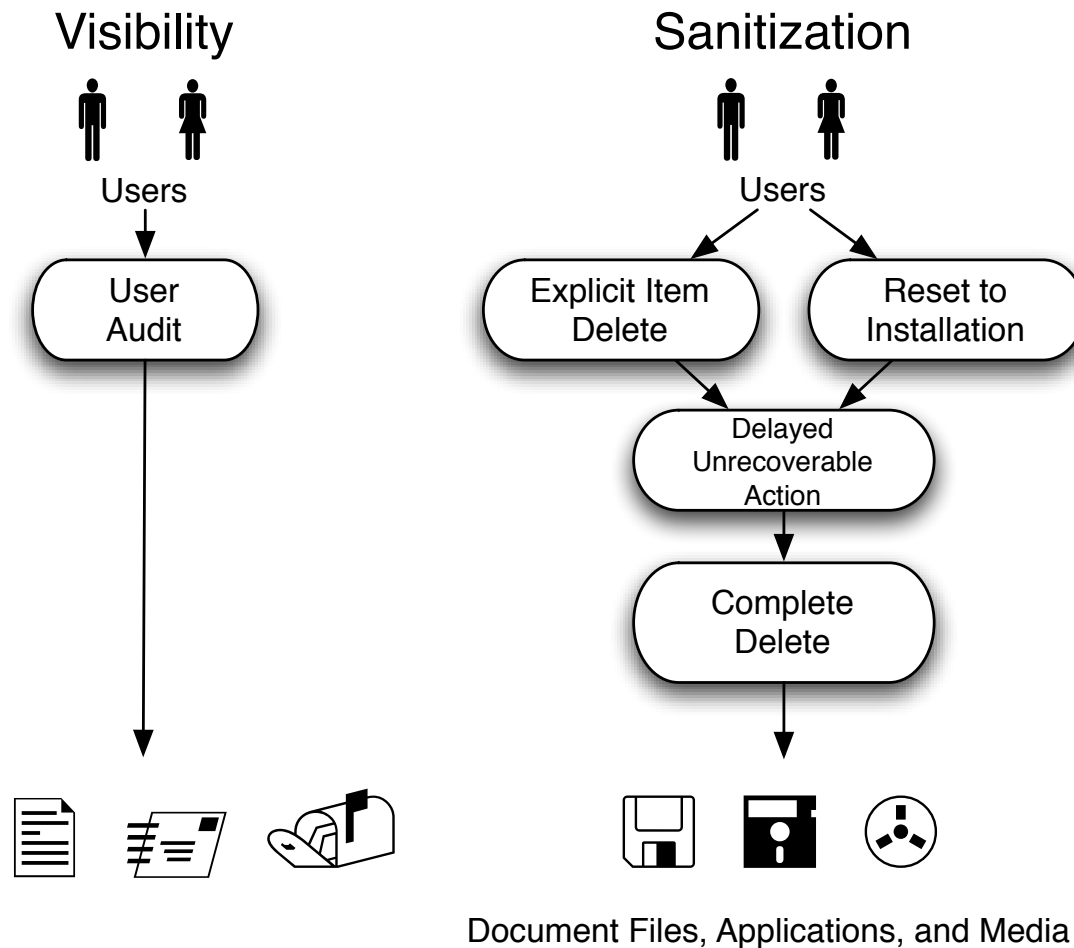
Large numbers of drives are being sold and given away.

Many of them appear to have hidden confidential information.



**We are morally obligated to solve this problem!**

## [Garfinkel '05] presents five distinct patterns for addressing the sanitization problem



<http://www.simson.net/thesis/>

## To be effective, a solution must address the root cause

### *Usability Problem:*

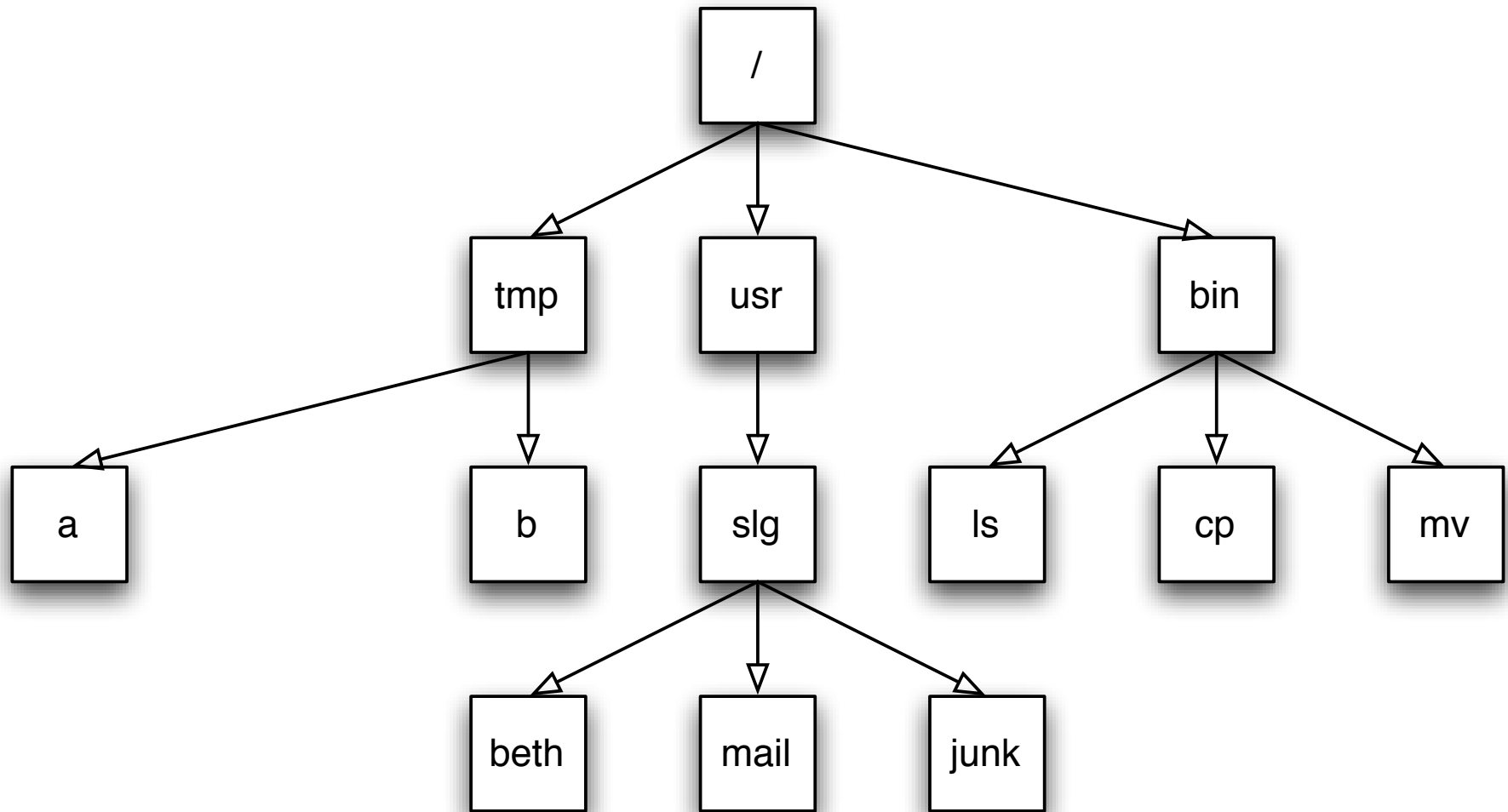
- Effective audit of information present on drives.
- Make DEL and FORMAT actually remove data.  
[Bauer & Priyantha 01]
- Provide alternative strategies for data recovery.

### *Education Problem:*

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

**To find that cause,  
I looked *on the drives* and *contacted the data subjects*.**

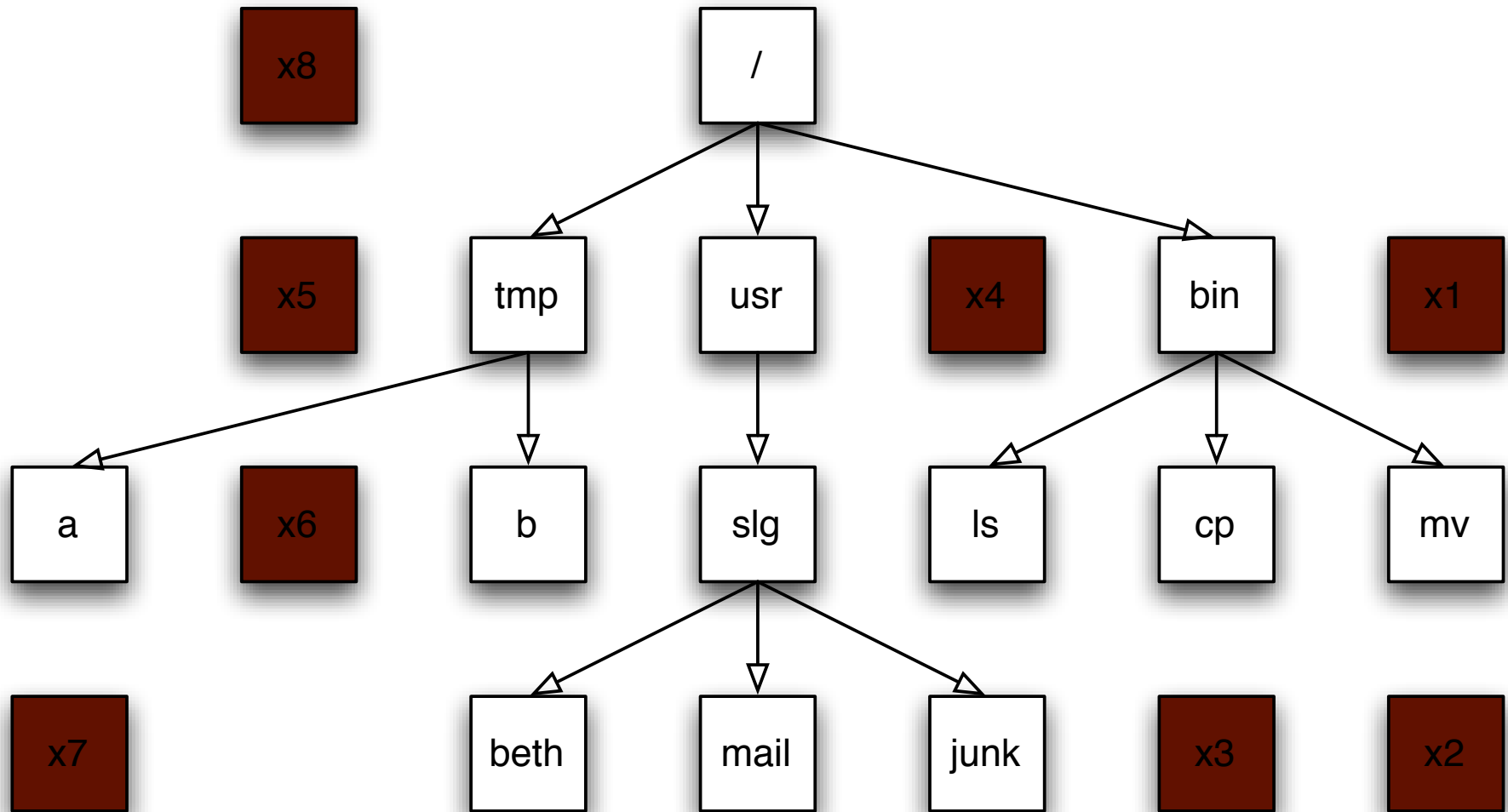
**Data on a hard drive is arranged in sectors.**



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

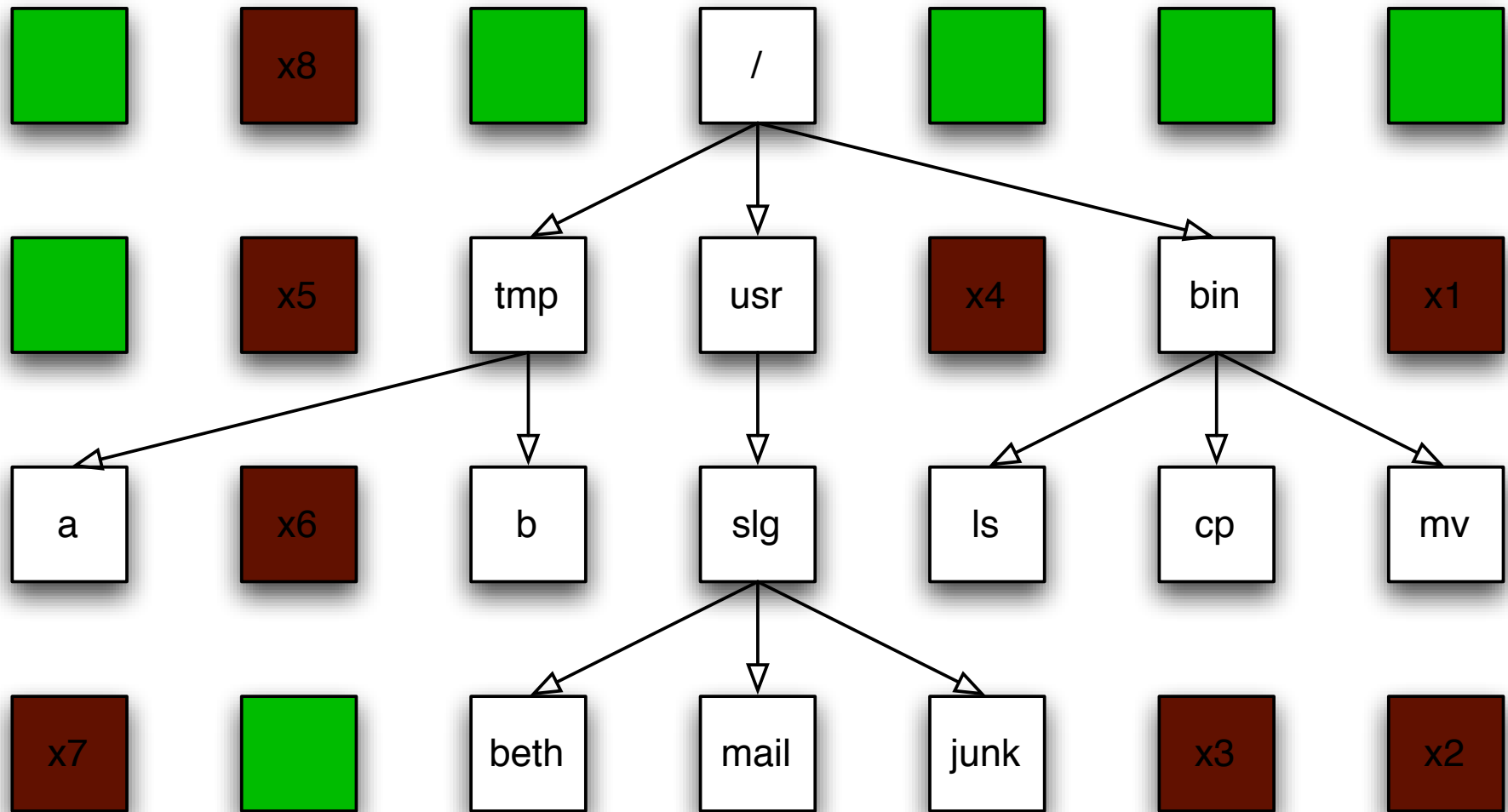


**Data on a hard drive is arranged in sectors.**



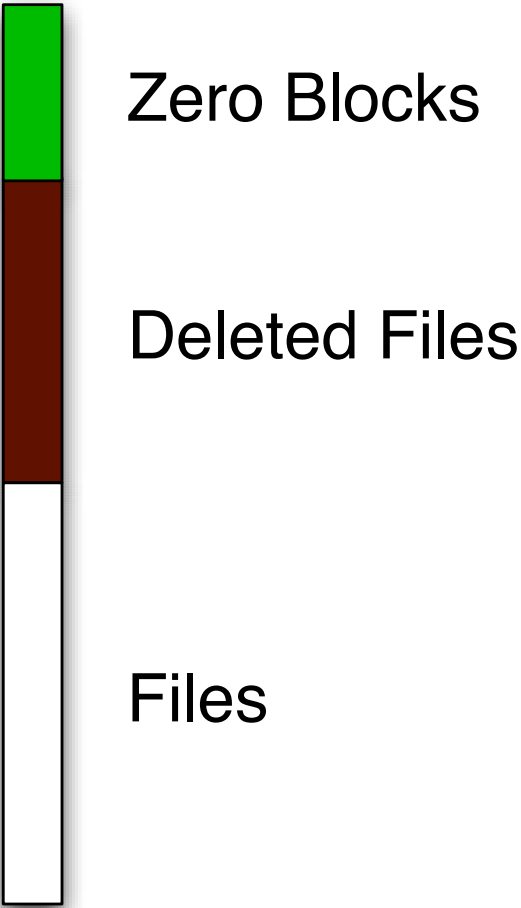
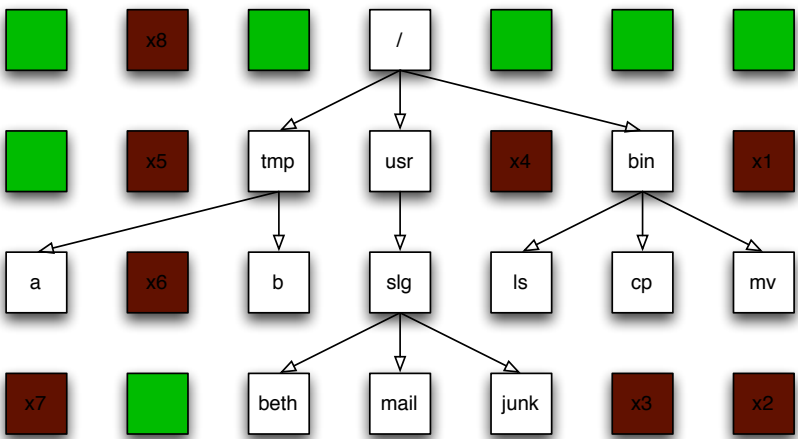
**The brown sectors indicate files that were deleted.**

**Data on a hard drive is arranged in sectors.**

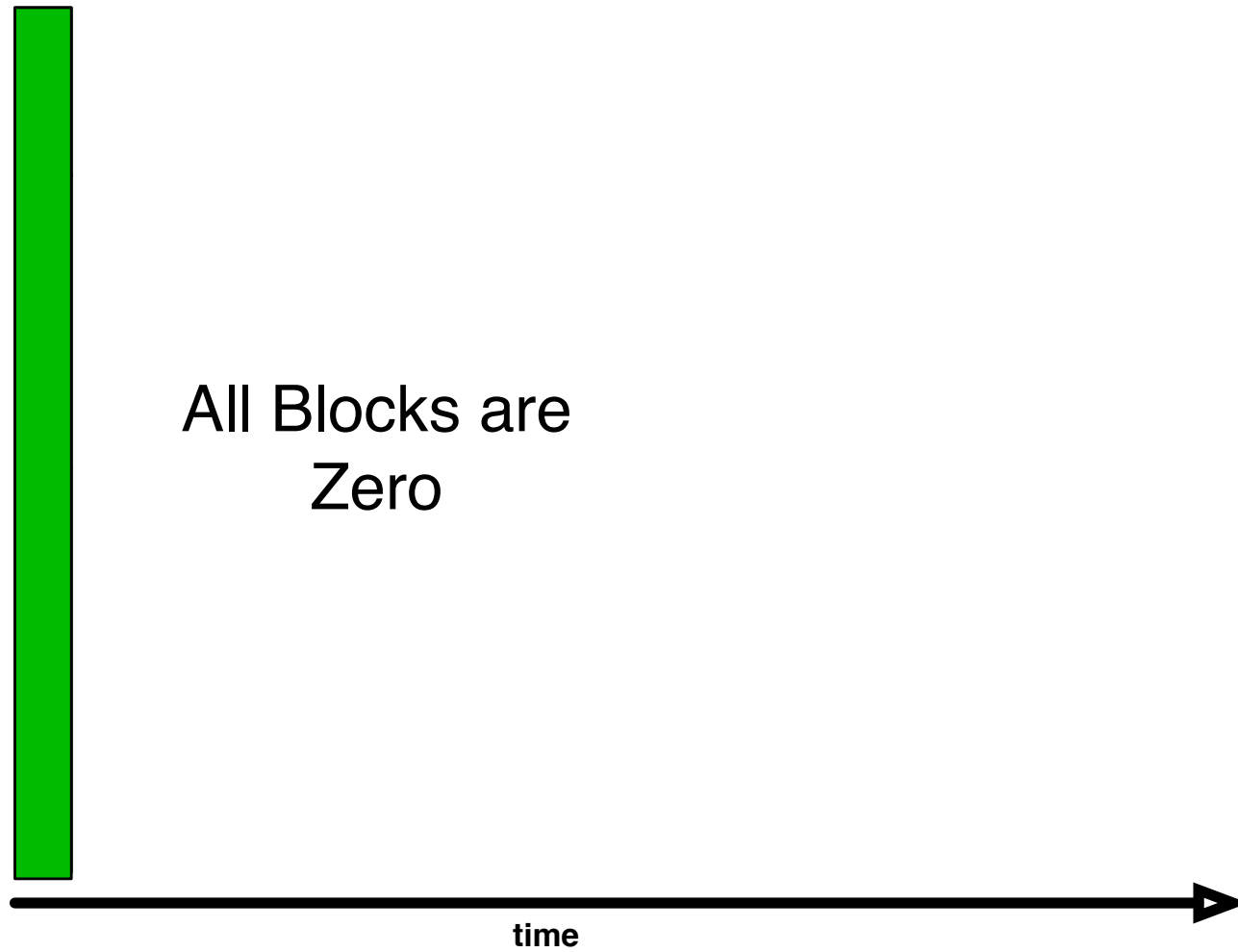


**The green sectors indicate sectors that were never used (or that were wiped clean).**

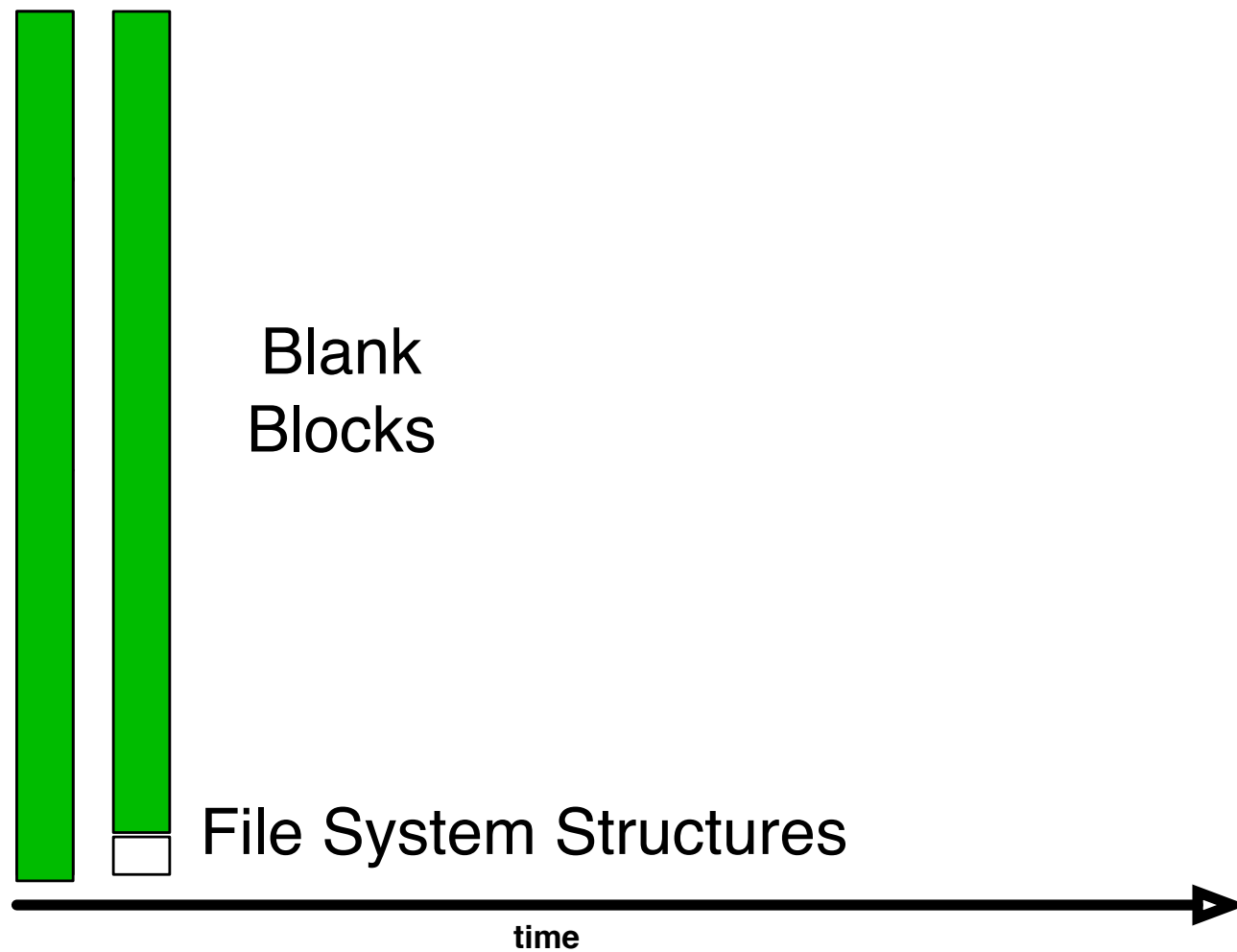
# Stack the disk sectors:



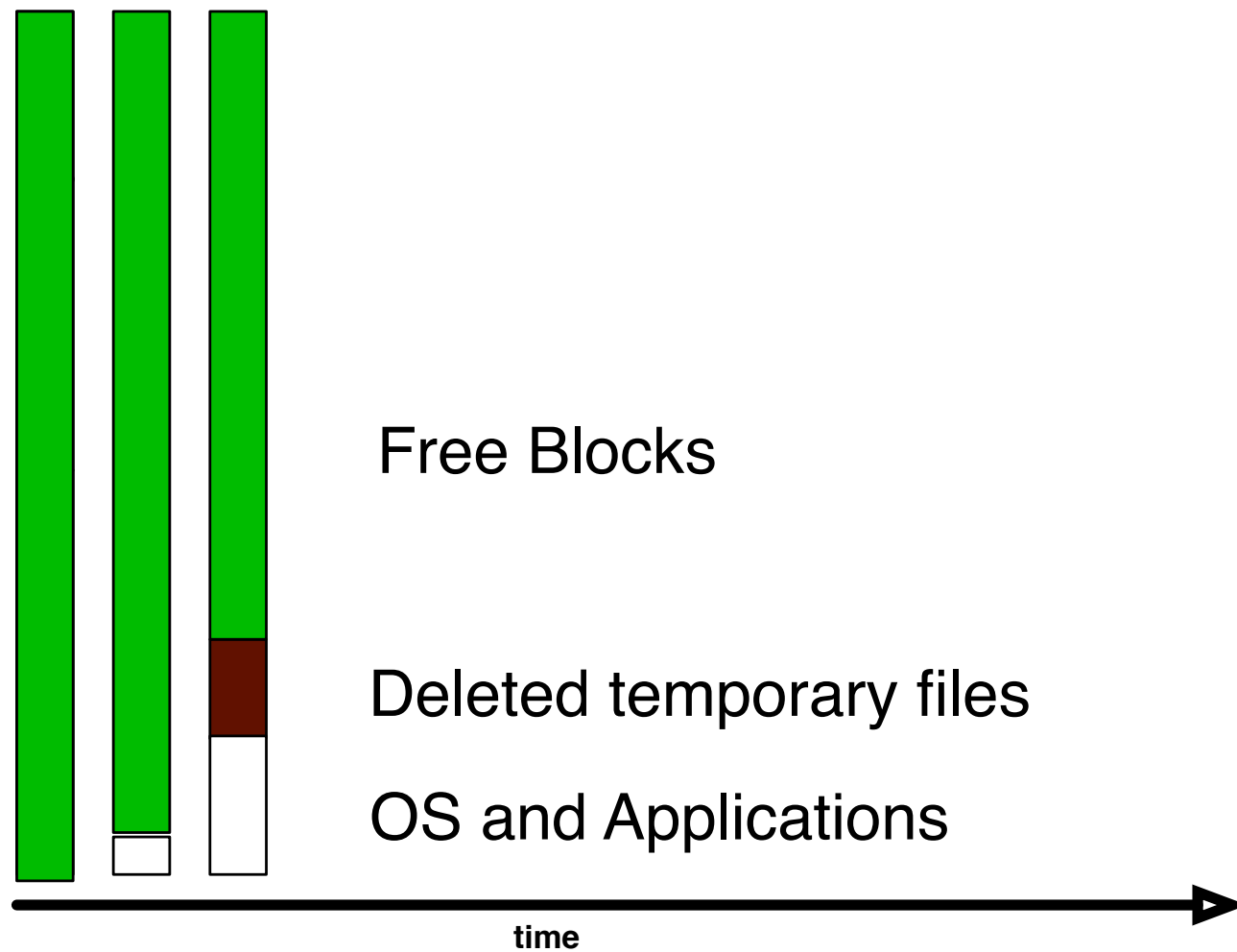
**NO DATA: The disk is factory fresh.**



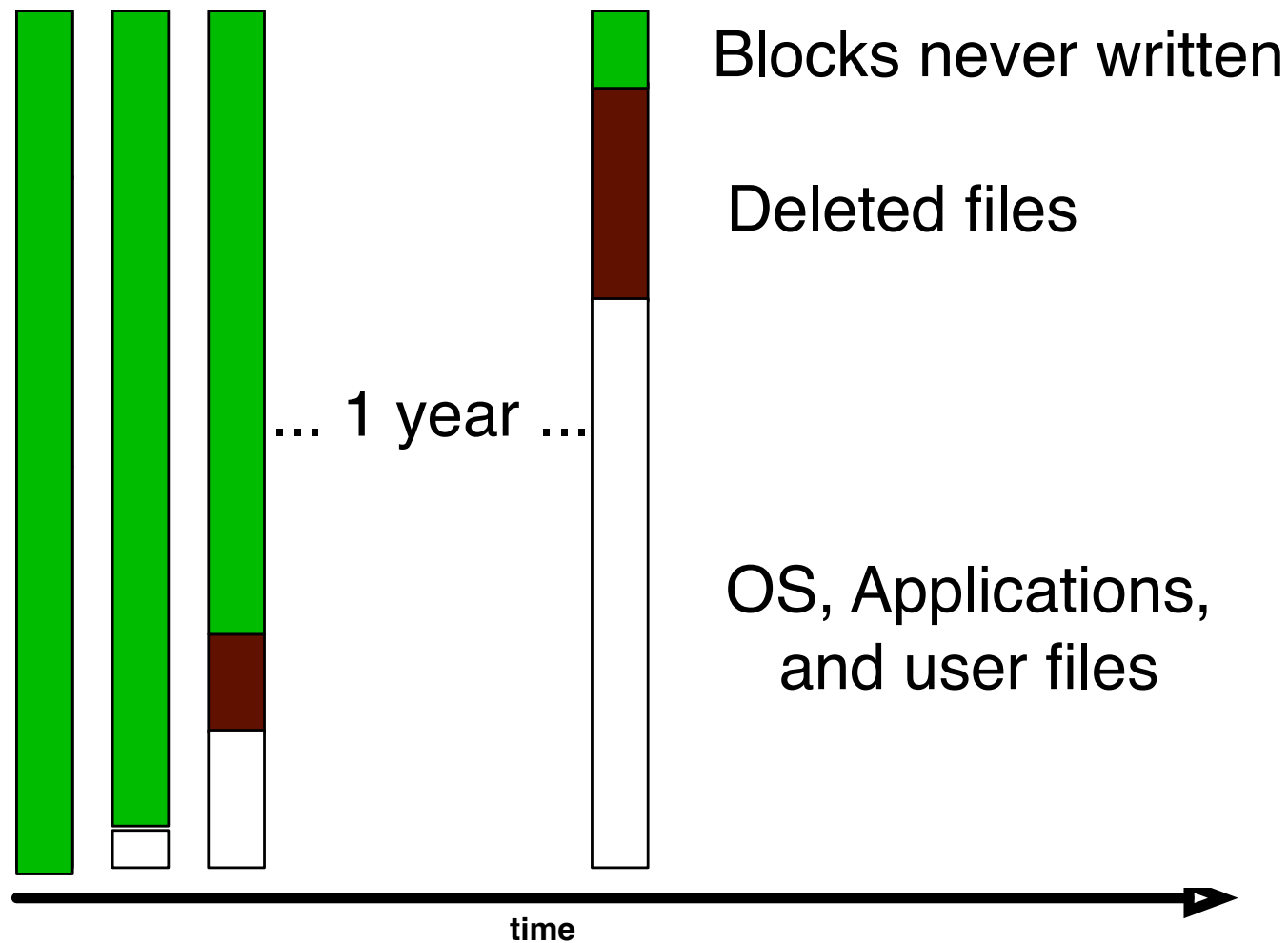
## FORMATTED: The disk has an empty file system



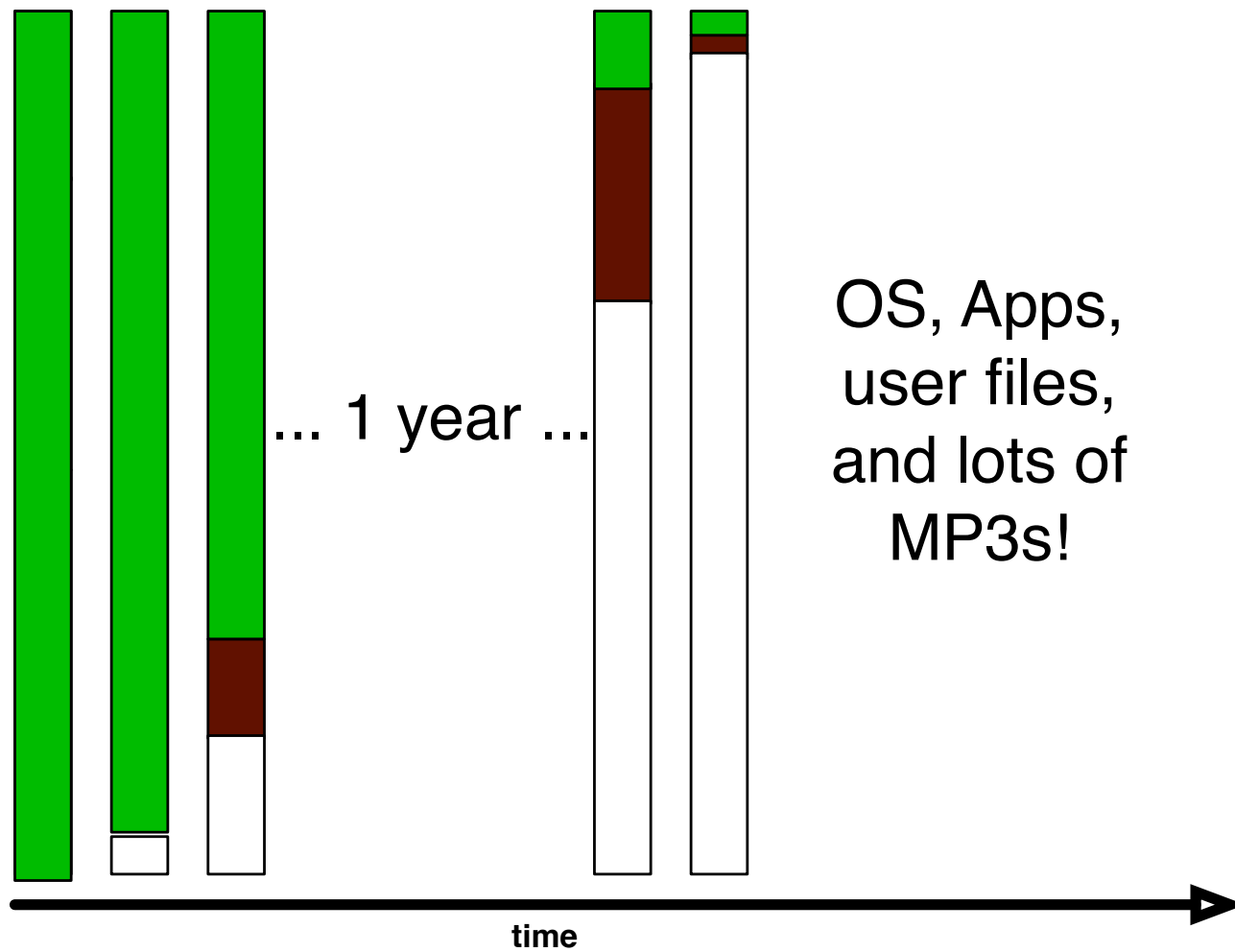
## AFTER OS INSTALL: Temp. files have been deleted



## AFTER A YEAR OF SERVICE

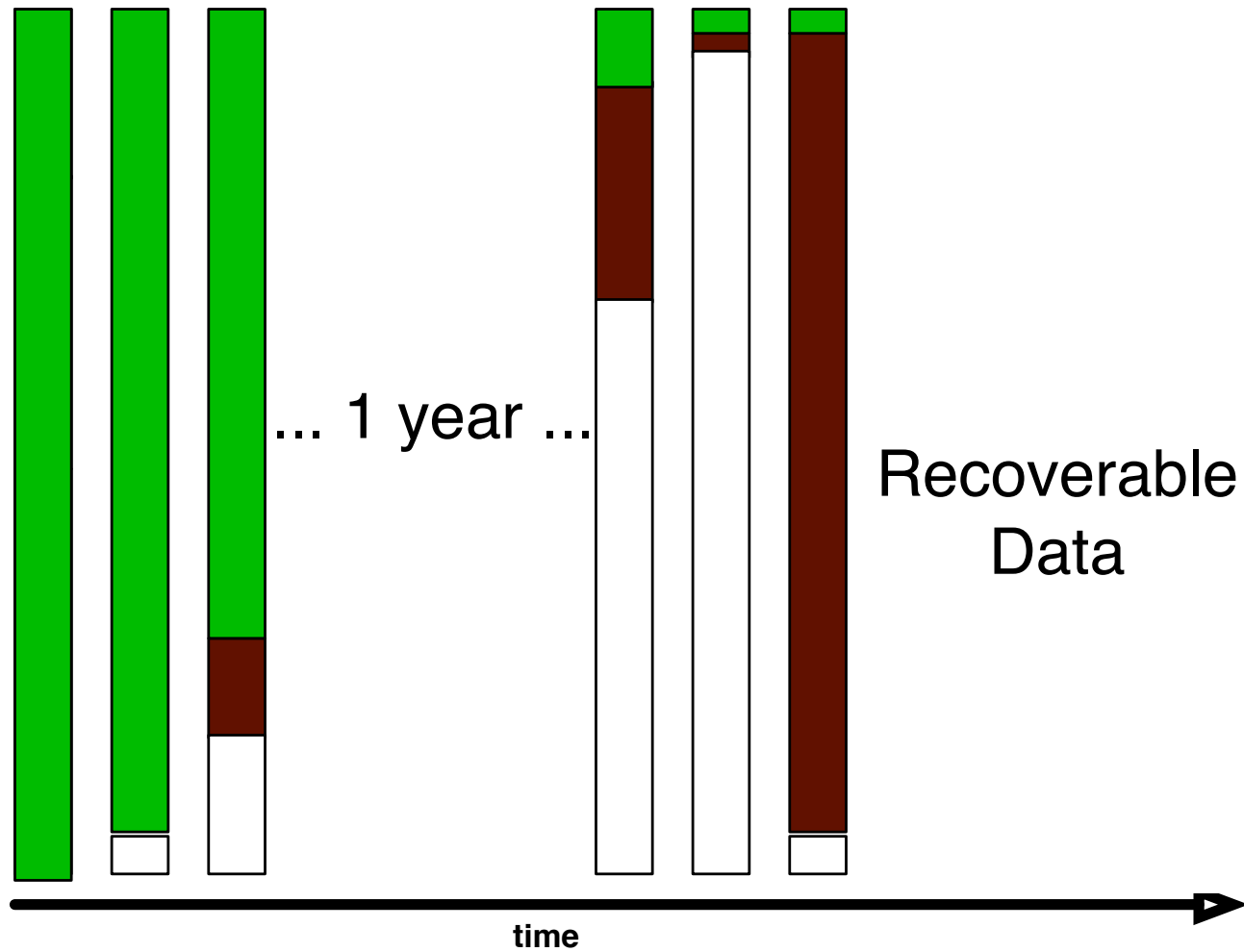


## DISK NEARLY FULL!

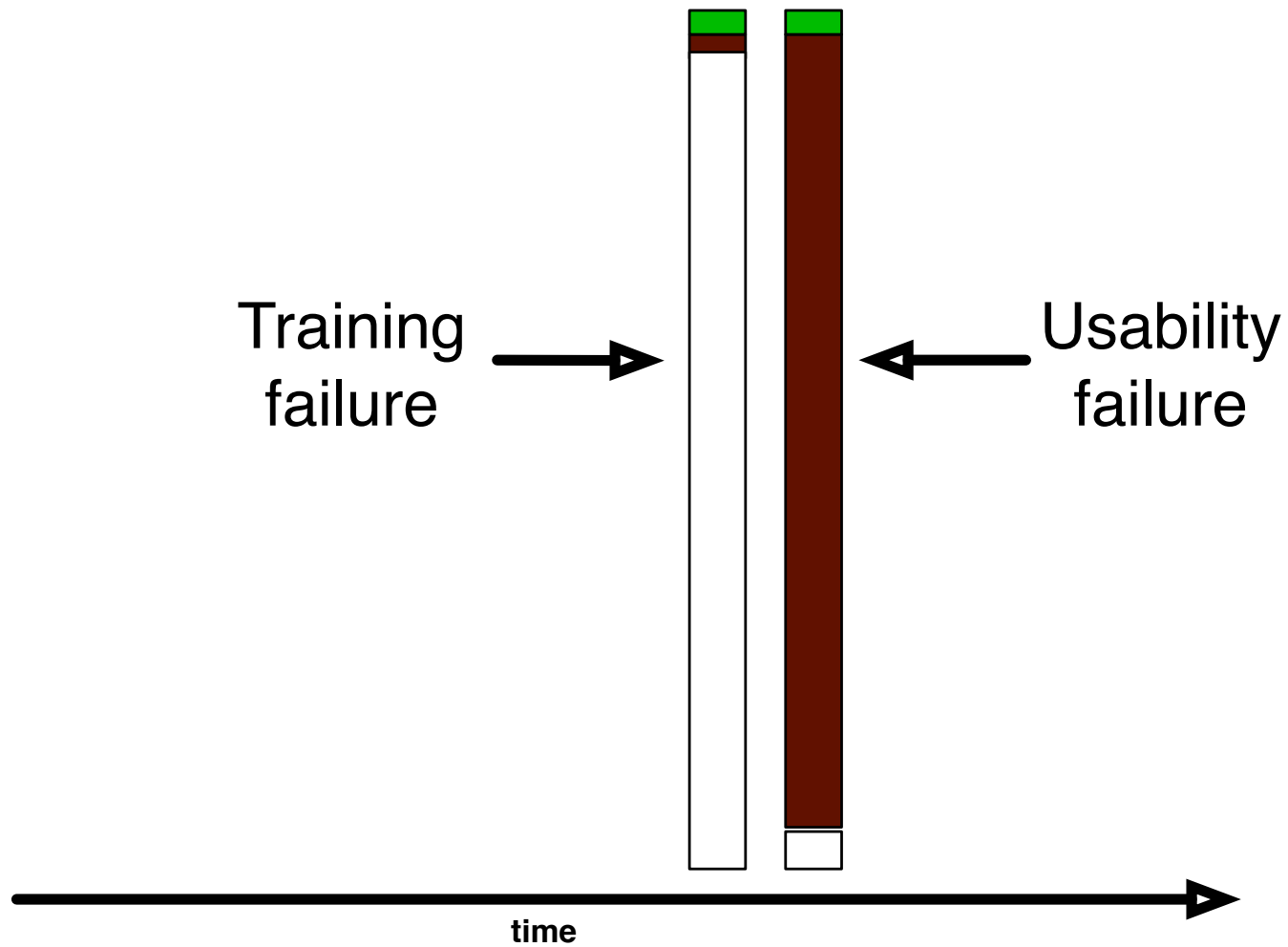




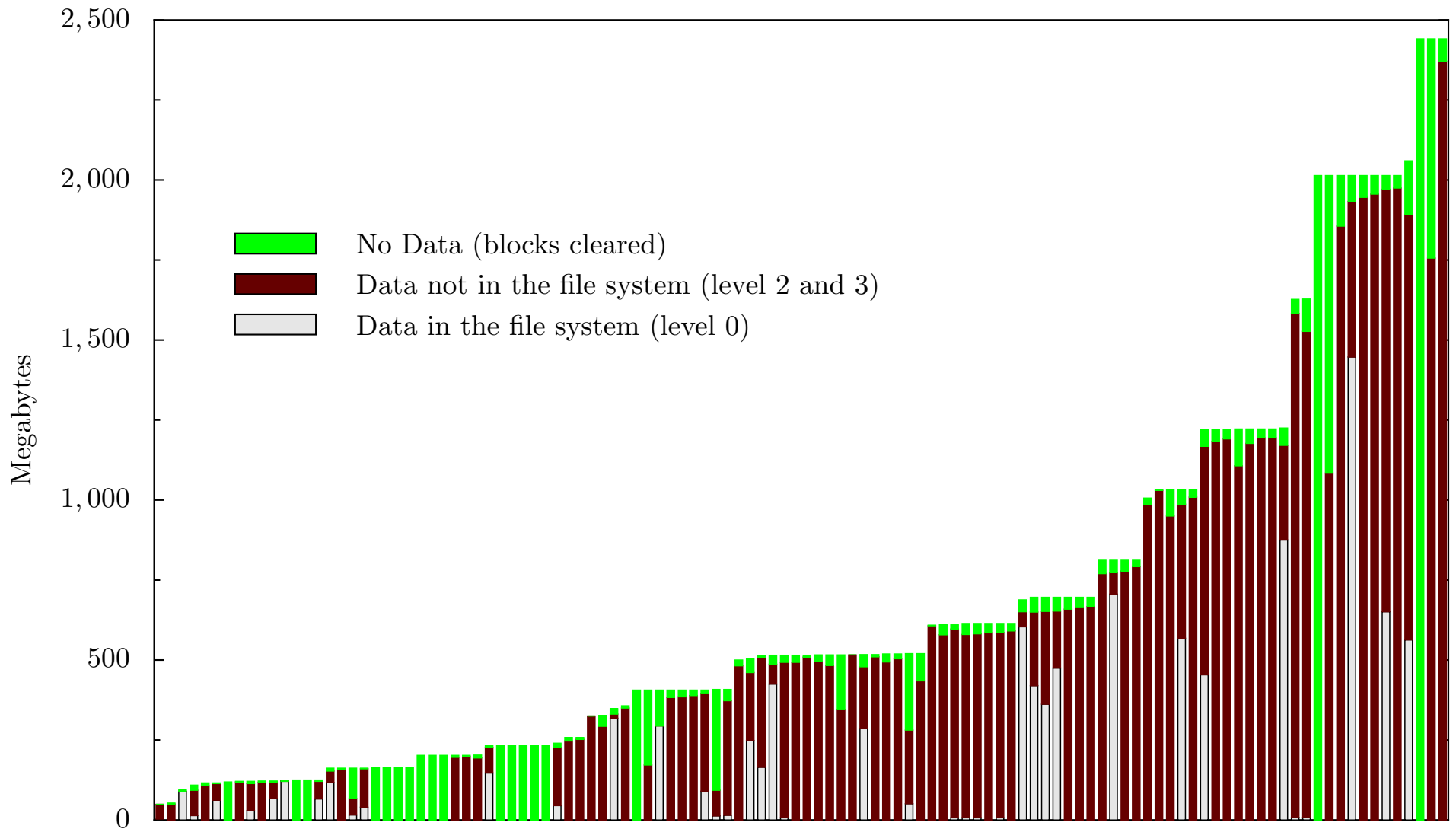
## FORMAT C:\ (to sell the computer.)



## We can use forensics to reconstruct motivations:



## The 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

## **Only 33 out of 176 working drives were properly cleared!**

- 1 from Driveguys — but 2 others had lots of data.
- 18 from pcjunkyard — but 7 others had data.
- 1 from a VA reseller — 1 DOA; 3 dirty formats.
- 1 from an unknown source — 1 DOA, 1 dirty format.
- 1 from Mr. M. who sold his 2GB drive on eBay.

## MD5 hashing allows the identification of files.

Interestingly, few unique files that had not been deleted:

File type	Unique Files
Microsoft Word files:	783
Microsoft Excel files:	184
Microsoft PowerPoint files:	30
Outlook PST files:	11
audio files:	977

**Conclusion: *most users* DELETED their files before discarding their drives.**

**But what *really* happened?**



**I needed to contact the original drive owners.**

# The *Remembrance of Data Passed Traceback Study.*

## [Garfinkel 05]

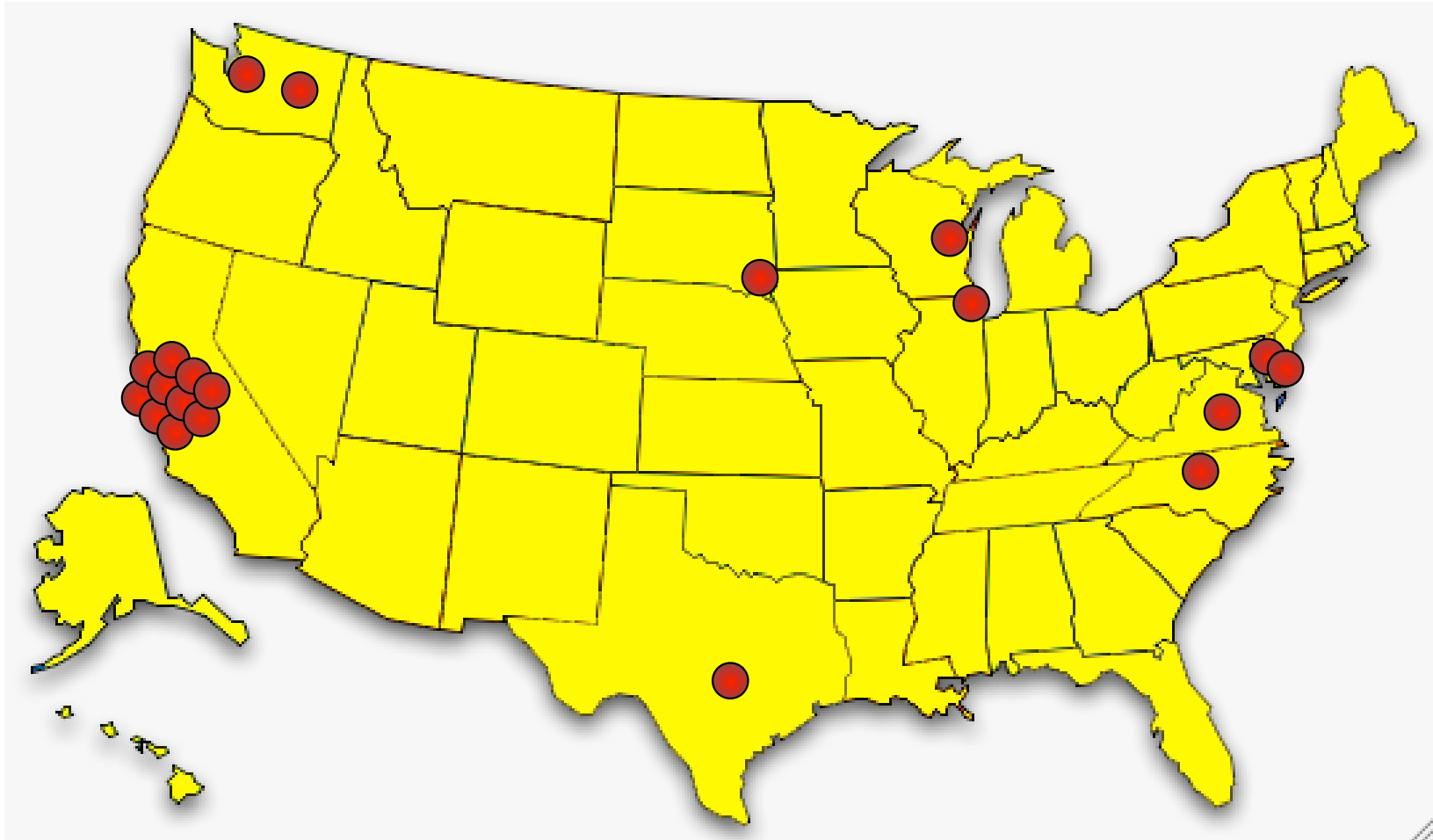
1. Find data on hard drive
2. Determine the owner
3. Get contact information for organization
4. Find the right person *inside* the organization
5. Set up interviews
6. Follow guidelines for human subjects work

```
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: 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.**

## **Second leading cause: Poor training and supervision**

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.**

## **Sometimes the data custodians just don'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

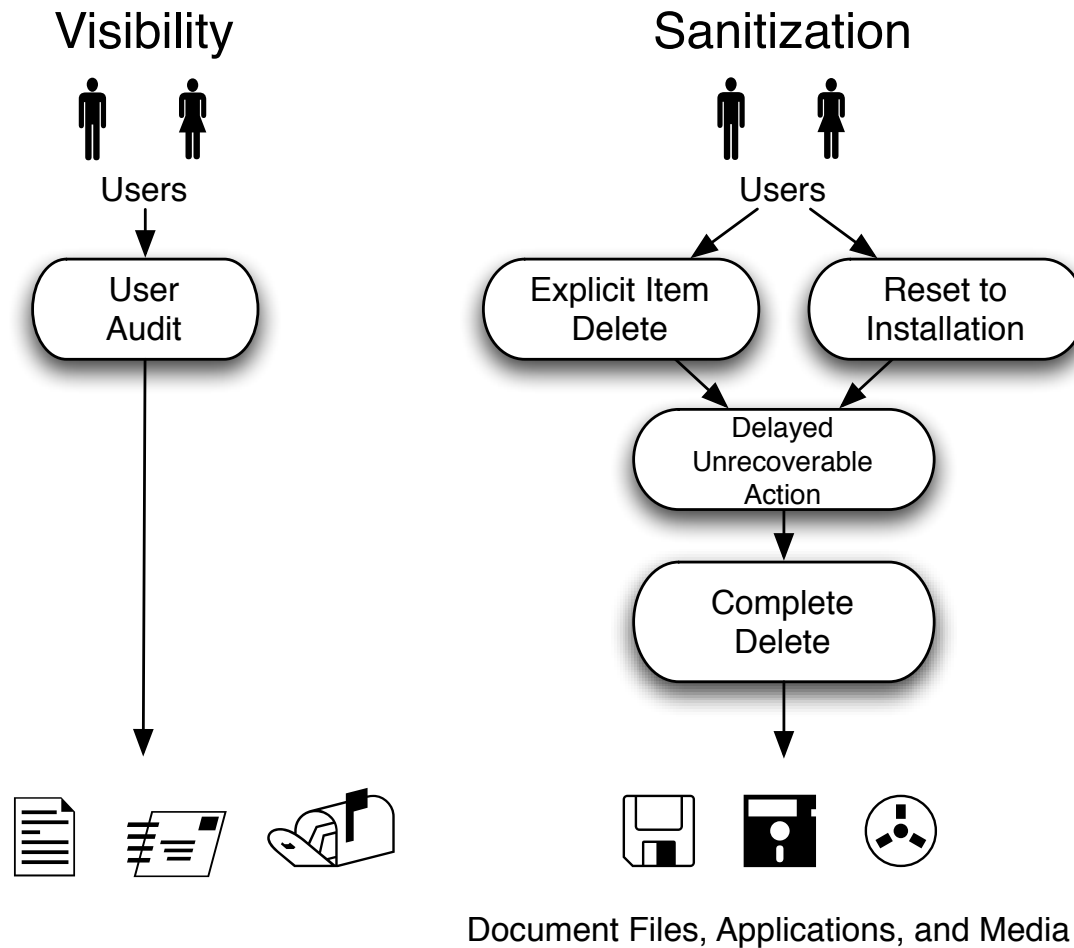
Lack of Concern: 2 cases

### **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.

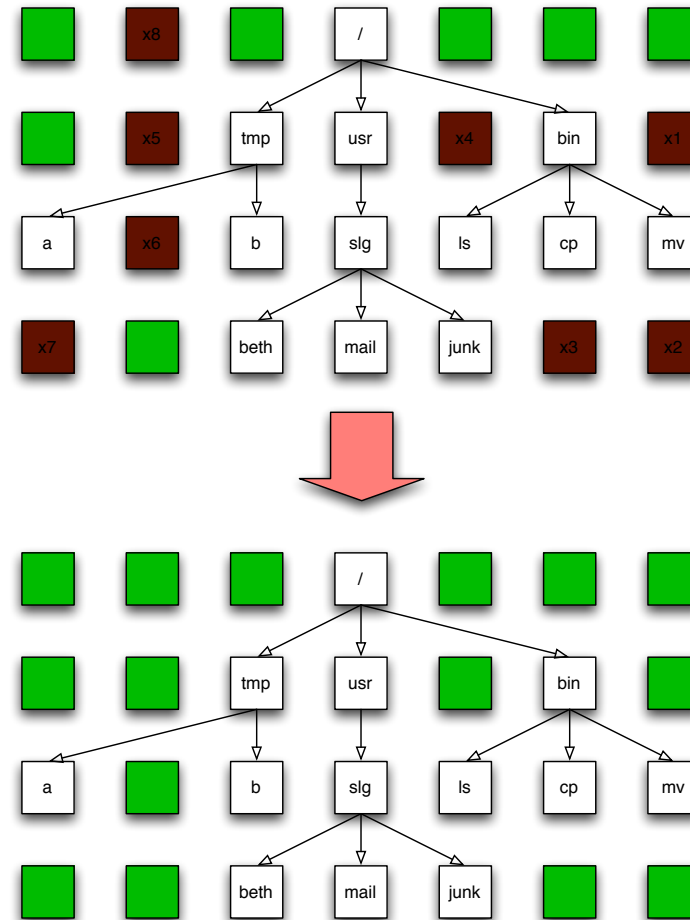
Sanitization



Complete  
Delete



Document Files, Applications, and Media



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

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

Sanitization



Users

Delayed  
Unrecoverable  
Action

Complete  
Delete



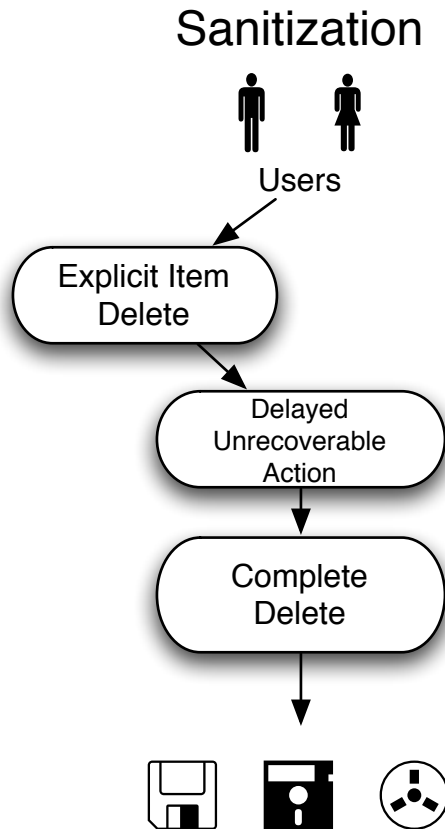
Document Files, Applications, and Media



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



# Two ways to delete information. #1: *Explicit Item Delete*



Document Files, Applications, and Media

```
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

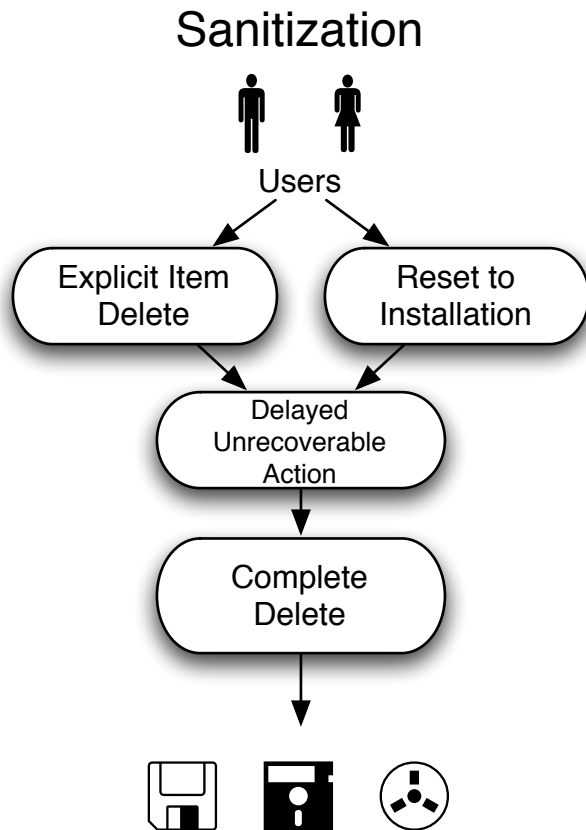
10/15/2004  09:22 PM    <DIR>          .
10/15/2004  09:22 PM    <DIR>          ..
               0 File(s)                   0 bytes
               2 Dir(s)         4,229,296,128 bytes free

C:\tmp>_
```

The screenshot shows a Windows command prompt window titled 'C:\WINDOWS\system32\cmd.exe'. It displays the directory contents of 'C:\tmp' before and after deleting the file 'big\_secret.txt'. The initial 'dir' command shows the file's size (27,262,976 bytes) and the total free space (4,202,078,208 bytes). After running 'del big\_secret.txt', the second 'dir' command shows that the file has been removed, leaving 0 bytes of file space and increasing the free space to 4,229,296,128 bytes.

**“Provide a means for deleting information where the information is displayed.”**

## ***Reset to Installation: Get rid of everything***



Document Files, Applications, and Media

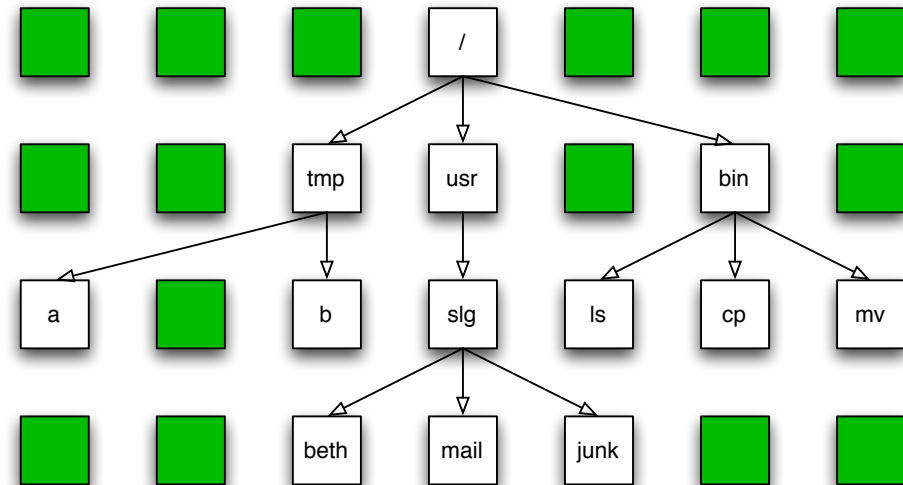
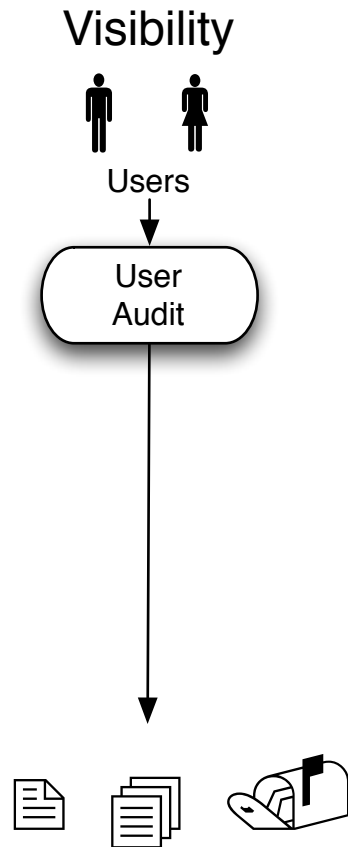
```
C:\>format c:
The type of the file system is NTFS.
WARNING, ALL DATA ON NON-REMOVABLE DISK
DRIVE C: WILL BE LOST!
Proceed with Format (Y/N)?
```

A screenshot of a Windows command prompt window. The title bar shows 'C:\> C:\WINDOWS\system32\cmd.exe - format c:'. The command prompt shows the command 'format c:' being entered, followed by the output 'The type of the file system is NTFS.' and a warning message: 'WARNING, ALL DATA ON NON-REMOVABLE DISK DRIVE C: WILL BE LOST! Proceed with Format (Y/N)?'.

**Reset/reinstall functionality is common (Windows; PalmOS; etc.).**

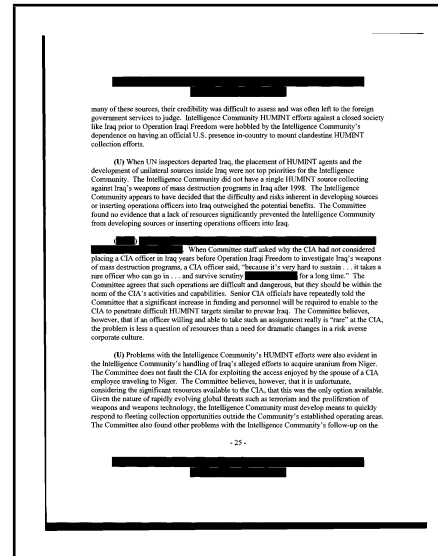
**This pattern framework clarifies *Reset's* security property.**

**User Audit:** If the information is present, make it visible.



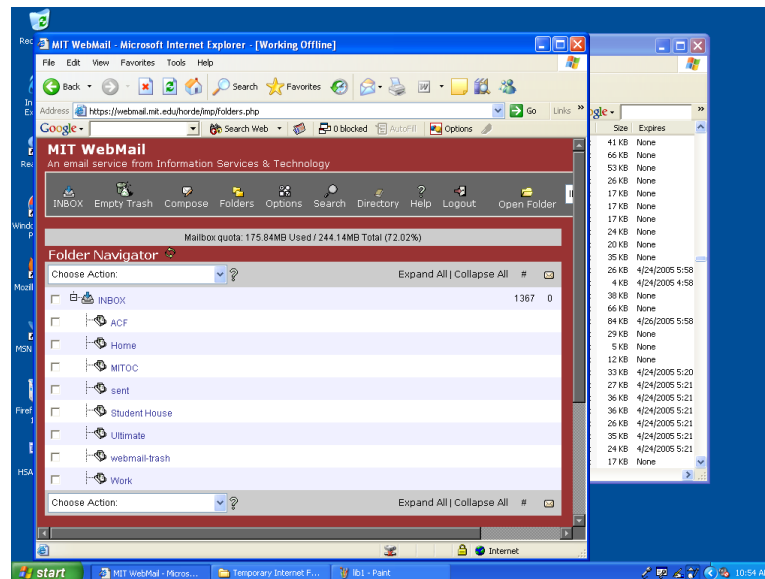
**With files, this happens automatically  
when the *Complete Delete* pattern is implemented.**

# The power of these patterns is that they apply equally well to other sanitization problems.



- Document Files

- Web Browsers

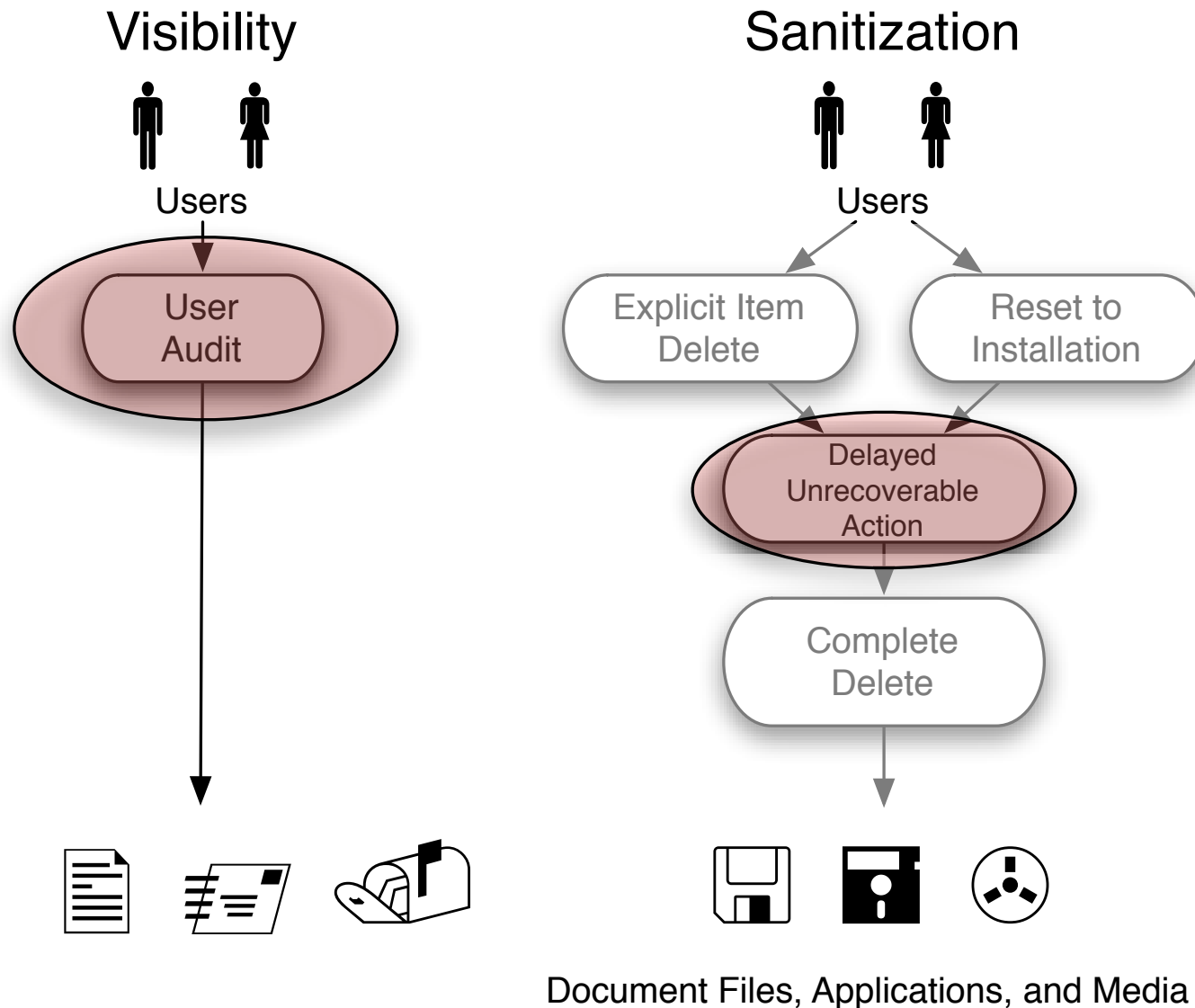


# Information is left in document files.

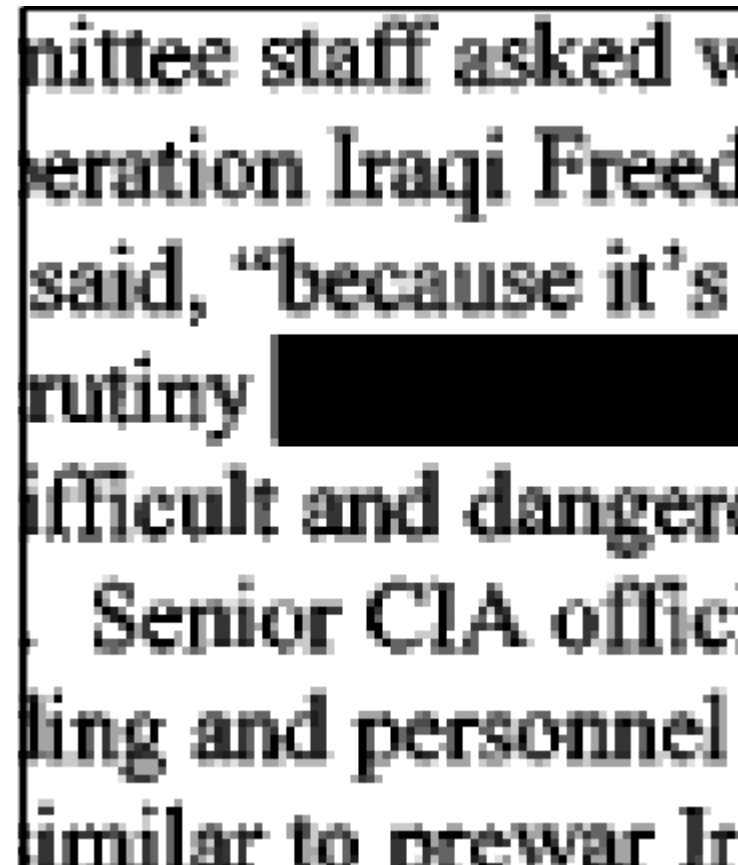
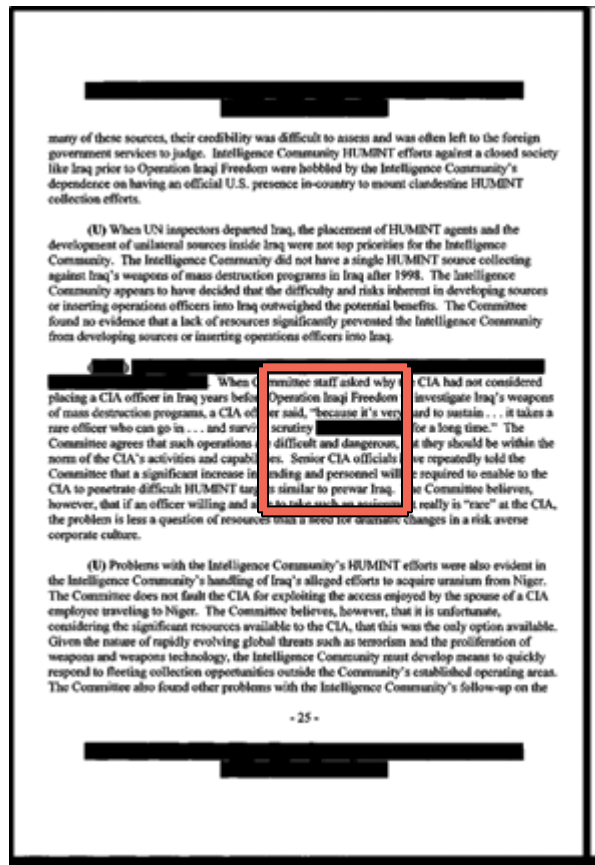
- The *New York Times* published a **PDF file** containing the names of Iranians who helped with the 1953 coup. [Young 00]
- US DoJ published a **PDF file** “diversity report” containing embarrassing redacted information. [Poulsen 03]
- SCO gave a **Microsoft Word file** to journalists that revealed its Linux legal strategy. [Shankland 04]
- Multinational Force-Iraq report

UNCLASSIFIED	
TABLE OF CONTENTS	
I. (U) BACKGROUND .....	1
A. (U) Administrative Matters .....	1
1. (U) Appointing Authority .....	1
2. (U) Brief Description of the Incident .....	1
B. (U) Constraints and Limitations .....	2
C. (U) Format of the Report .....	2
II. (U) ATMOSPHERICS .....	4
A. (U) Introduction .....	4
B. (U) Local Security Situation .....	4
1. (U) Iraq .....	4
2. (U) Baghdad .....	4
3. (U) Route Irish .....	4
C. (U) Known Insurgent Tactics, Techniques, and Procedures .....	5
1. (U) Methods of Attack .....	5
2. (U) Insurgent TTPs for IEDs .....	5
3. (U) Insurgent TTPs for VBIEDs .....	6
4. (U) Effectiveness of Attacks .....	7
D. (U) Recent Incidents in the Vicinity of Checkpoint 541 .....	8
E. (U) Unit Experience in the Baghdad Area of Responsibility .....	8
1. (U) [REDACTED] Division .....	8
2. (U) [REDACTED] Brigade, [REDACTED] Division .....	9
3. (U) [REDACTED] Battalion .....	9
4. (U) [REDACTED] Battalion .....	10
F. (U) Findings .....	10
III. (U) TRAFFIC CONTROL POINTS, BLOCKING POSITIONS, AND TRAINING .....	12
i	
UNCLASSIFIED	

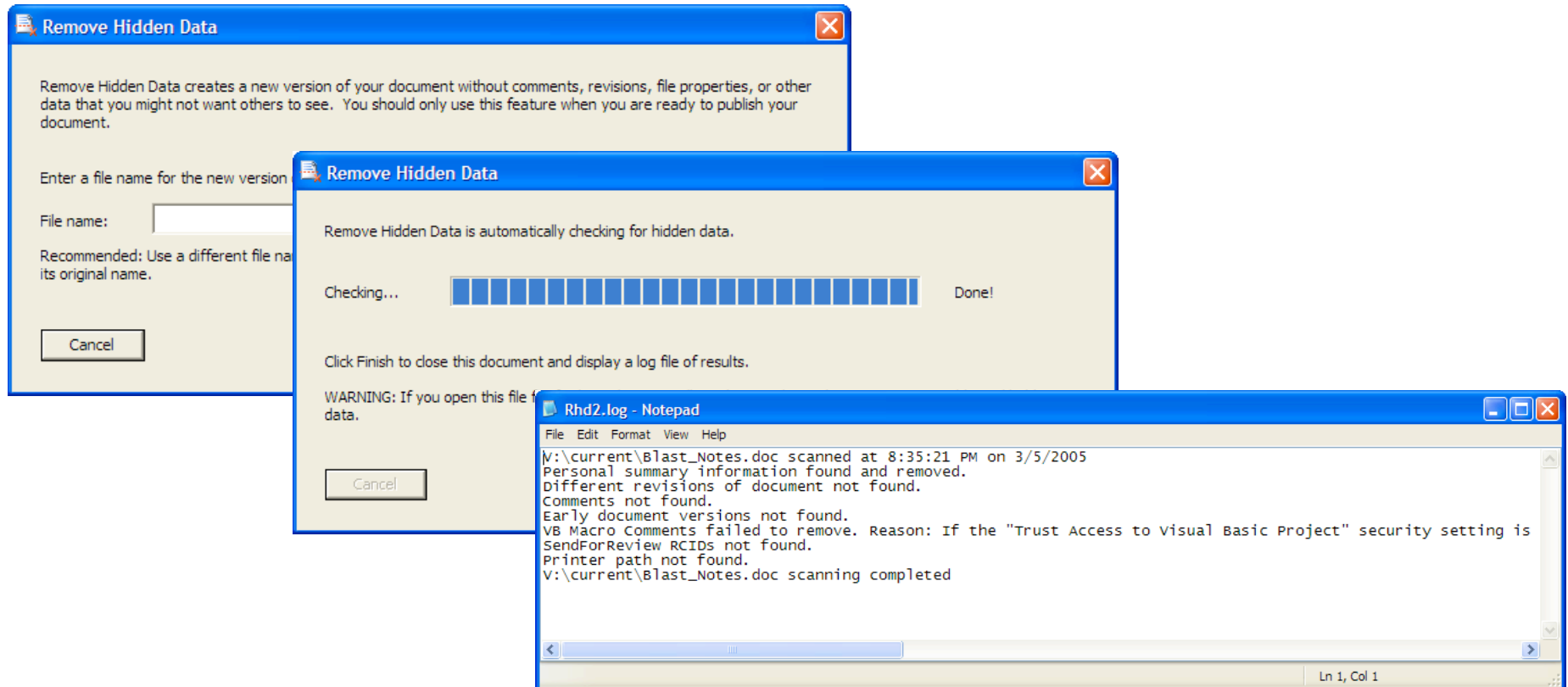
# The information leaked because two patterns were not implemented.



The Senate Foreign Intelligence Committee accomplished this goal by *scanning* the redacted report on pre-war Iraq intelligence to create the PDF that it distributed.



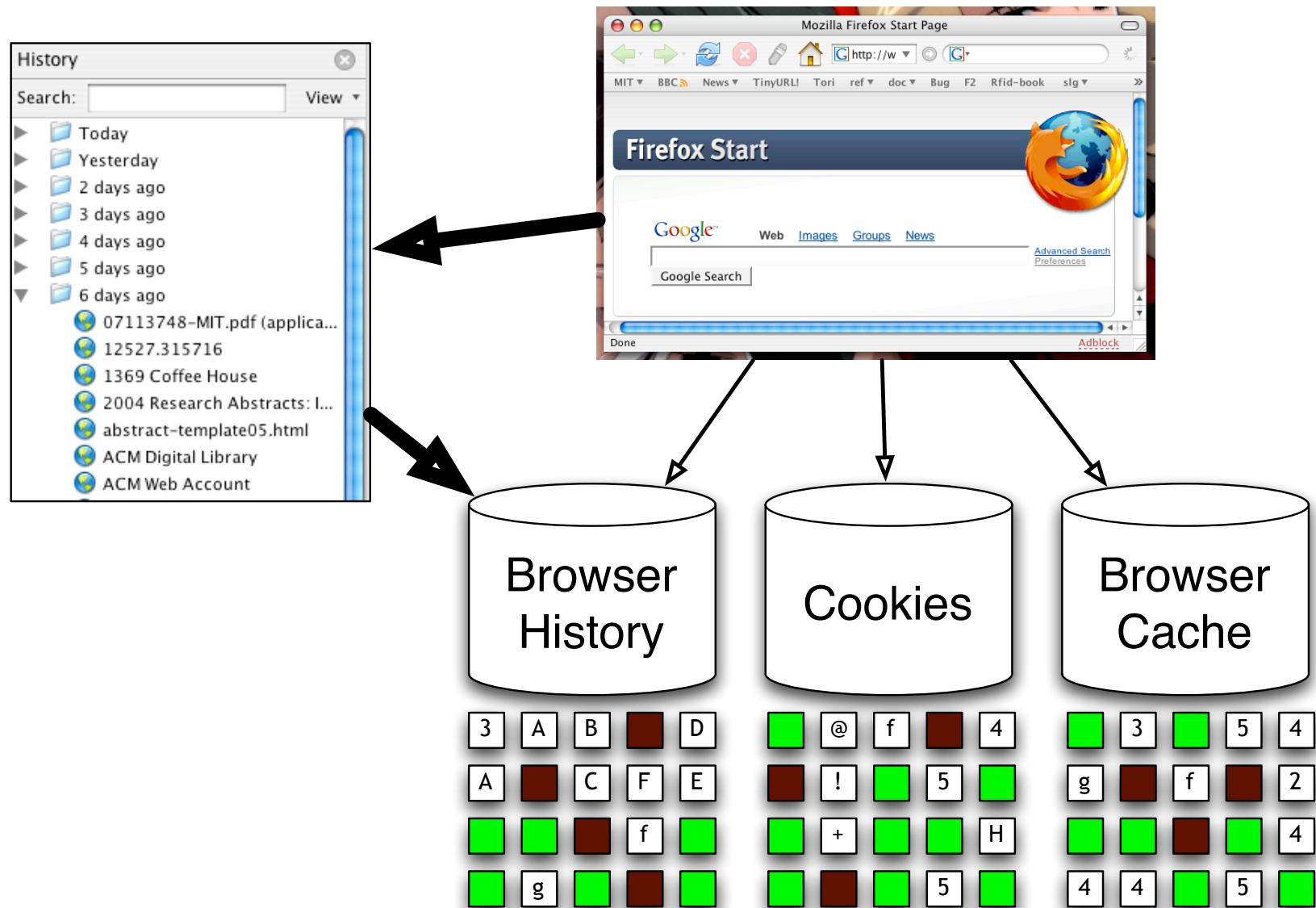
## Microsoft has tried to solve this problem with “Remove Hidden Data” tool.



**RHD doesn't integrate into the flow of document preparation. The patterns-based analysis predicts that RHD will fail in many cases.**

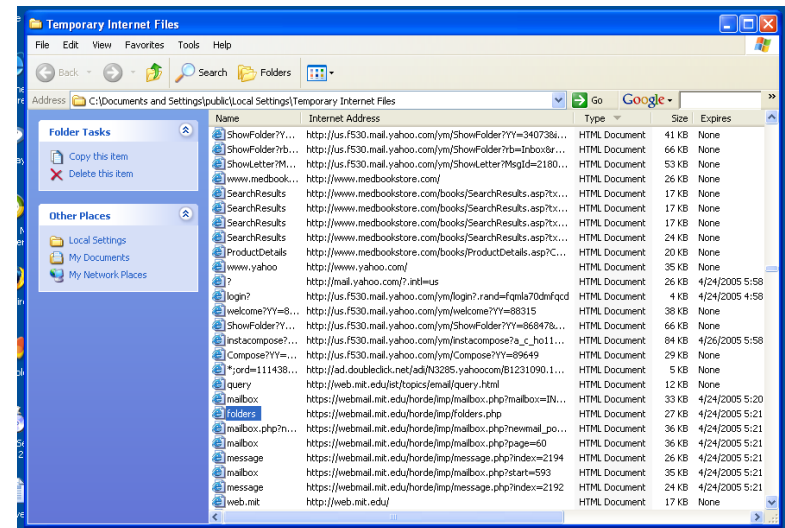


# Information is left behind in web browsers.



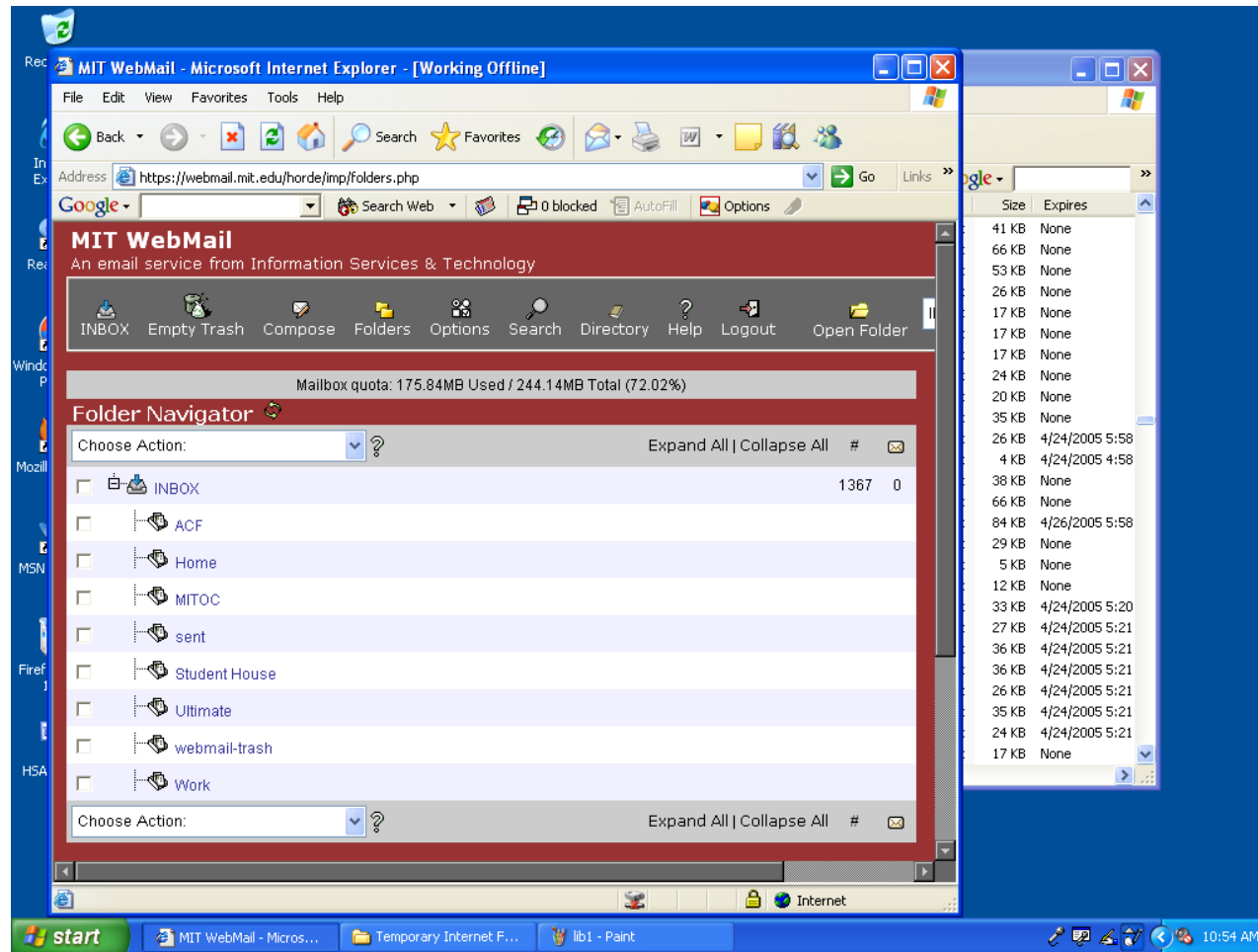
Two key problems: ① Deleted files; ② The cache

**In fact, a lot of information is left behind in web browsers.**



**MIT Humanities Library, April 25, 2005**

**4 out of 4 computers inspected had significant quantities of personal email in their browser caches.**



**The American Library Association recommends software that automatically purges caches on a *daily* basis.[ALA 05] (It would be better to purge after each use.)**

## **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.

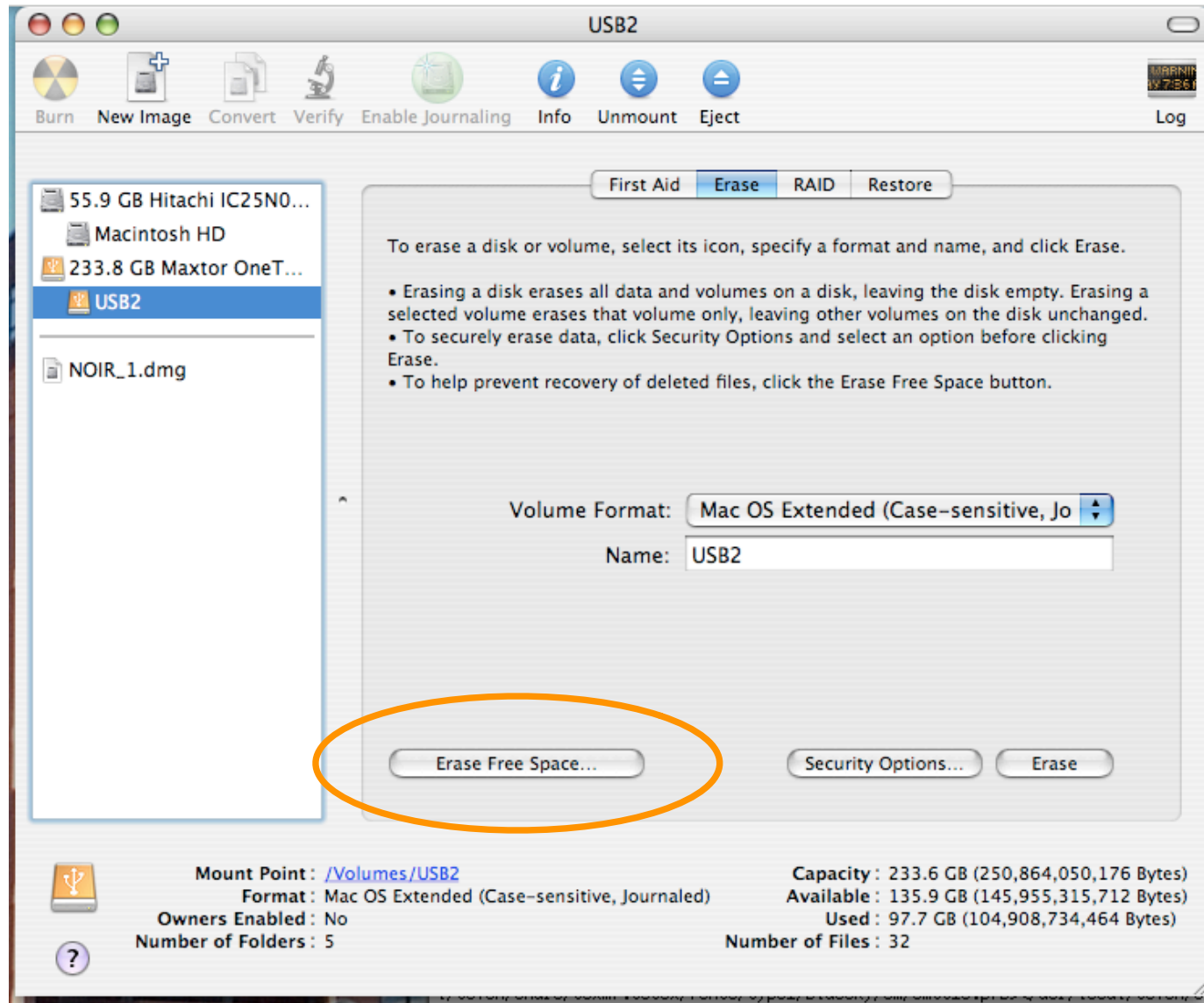


# Unfortunately, “Secure Empty Trash” is incomplete.

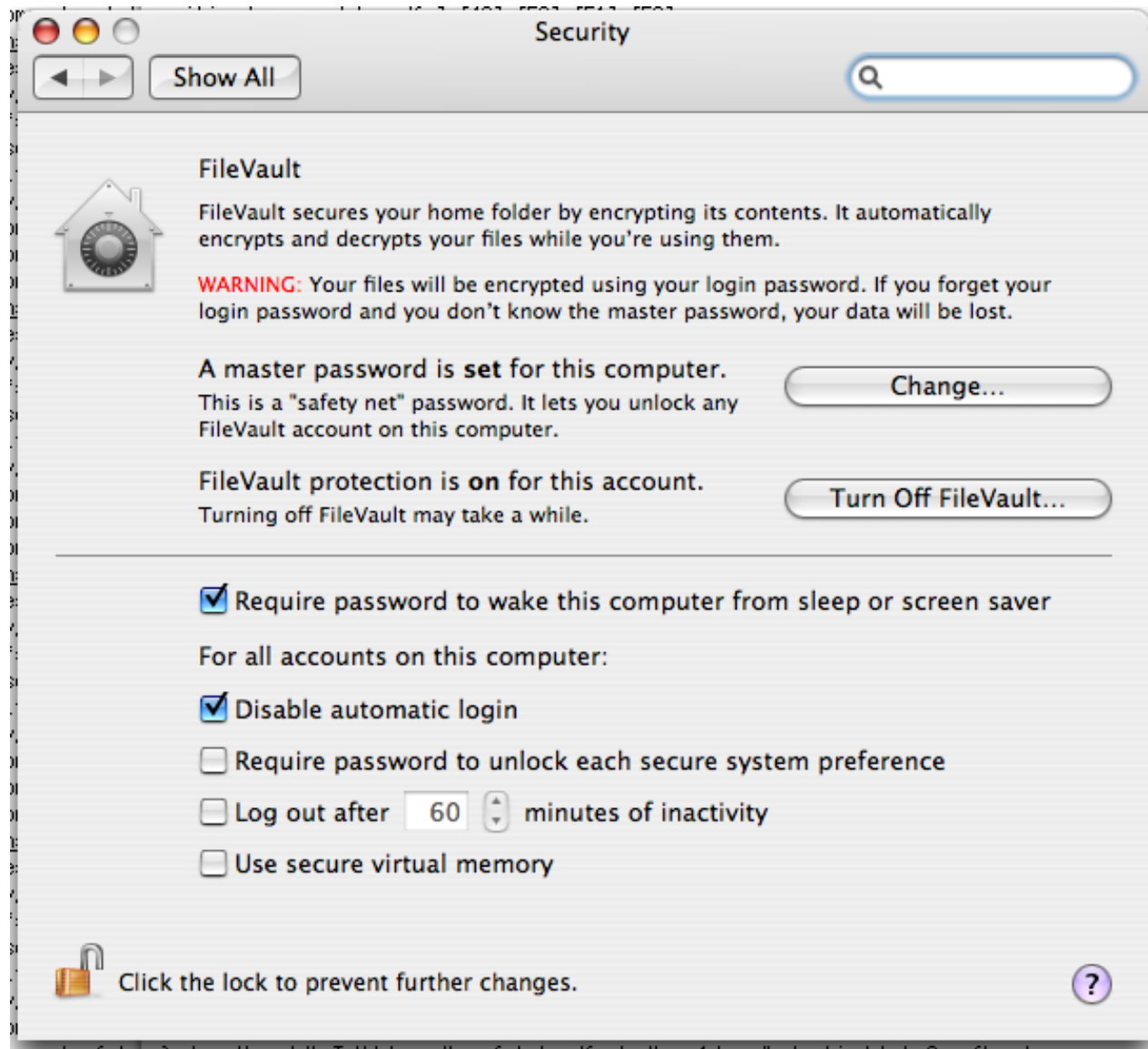
- Implemented in Finder (inconsistently)
- Locks trash can
- Can't change your mind



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



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





## Future Work: Deploying Compete Delete

- Make FORMAT actually erase the disk.
- Make “Empty Trash” actually overwrite data.
- Integrate this functionality with web browsers, word processors, operating systems.
- Address usability dangers of clean delete.
- Analysis of “one big file” technique.

**Let's put this in Linux!**

## **Future Work: 2500 Drive Corpus**

- Automated construction of stop-lists.
- Detailed analysis of false positives/negatives in CCN test.
- Explore identifiers other than CCNs.
- Support for languages other than English.

**More than 500 drives are standing by...**

## **Future Work: Toolkit**

- Easy-to-use, reliable, disk imaging software.
- New file format for disk images.
- Web-based database of hash codes.

**Initial version is available for download.**

## **Future Work: Economics and Society**

- Who is buying used hard drives and why?
- Hard drive honeypot.
- Compliance with FACT-A

**This is a lot of work...**

## Future Work: Summary

- Improved cross-drive forensics
- 2500 Drive Corpus
- Open-Source Toolkits
- Economics and Society

**Questions?**