

Security, Privacy and Usability



Simson L. Garfinkel
Center for Research on Computation and Society
Harvard University
February 21, 2006

The Tandy 200



Three memory banks (1 visible, 2 hidden)

Hidden information is a widespread Usability/Security problem today.



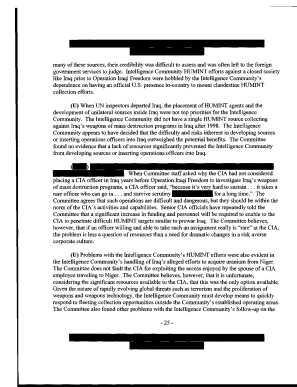
Tandy 200



USB drive



Hard Drive



PDF file

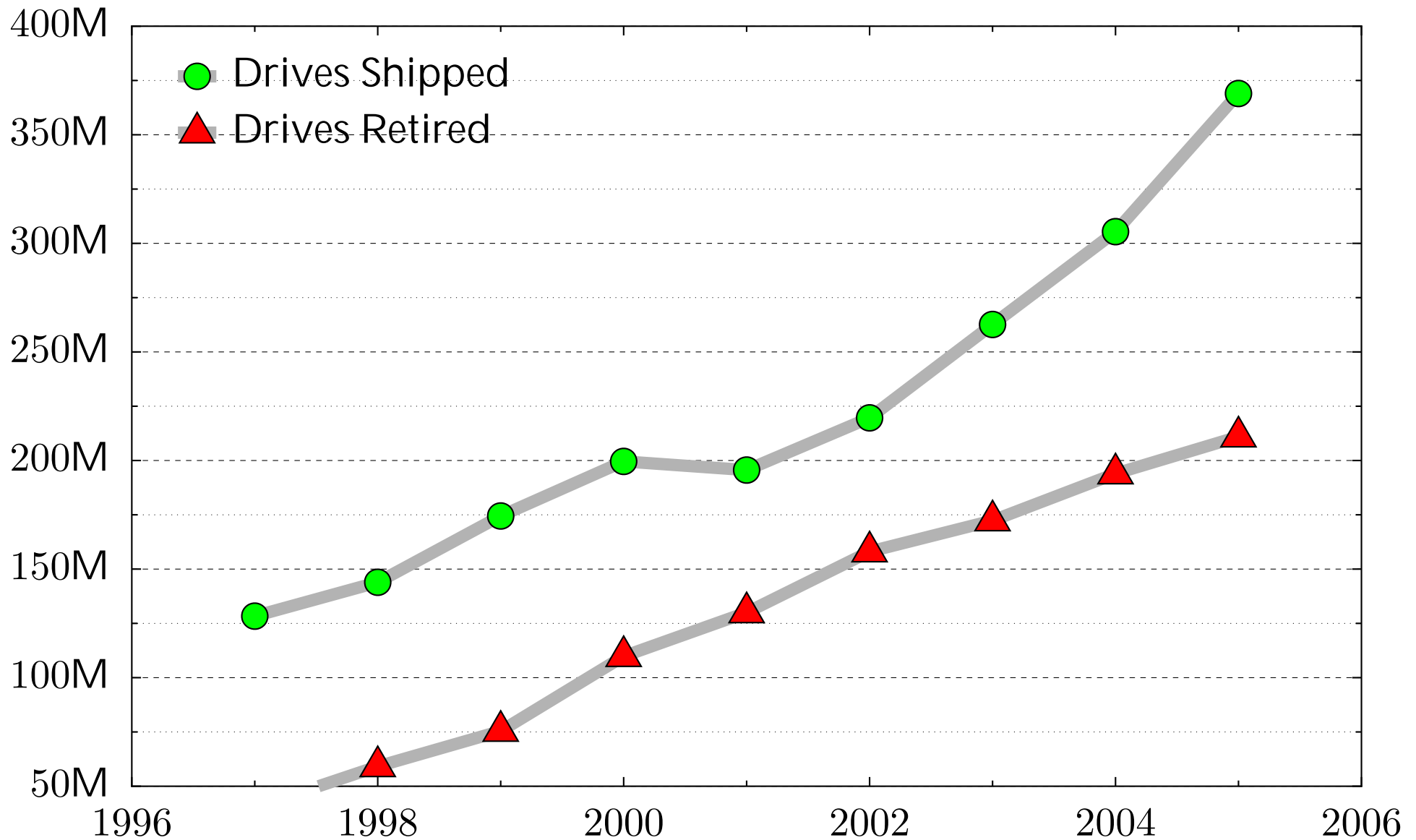
There are roughly a dozen documented cases of people purchasing old PCs and finding sensitive data.

- 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.[O&O 04]
- O&O Software GmbH – 200 drives.[O&O 05]

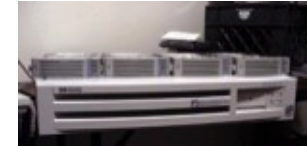


None of these are scientifically rigorous studies.

**This is a huge problem:
210 million drives were retired in 2005!**



There is a significant market for used disk drives.



Retired drives are:

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

All Categories [Save this search](#)

360 items found for hard drives

Sort by items: [ending first](#) | [newly listed](#) | [lowest priced](#) | [highest priced](#)

Picture	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
	CD 1.2 hard drives & (15) 10/100 network	\$25.00	1	1h 00m
	Lot of 3 Quantum 9.1 gig SCSI Hard Drives	\$26.00	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 <i>7 days left</i>	-	1h 50m
	QTY 3... IDE Hard Drives 2.5 Gg	\$20.50	5	2h 02m
	5 WESTERN DIGITAL 2.5 GIG HARD DRIVES	\$30.00	4	2h 03m
	QTY 3... IDE Hard Drives 1.0 Gg	\$9.99	1	2h 04m
	Western Digital 850 meg IDE Hard Drives each	\$6.00	1	2h 57m
	WINDOWS	\$6.00	-	3h 18m

About 1000 used drives/day sold on eBay.

In 1998 I decided to start purchasing hard drives on the secondary market.



2001: 100 drives



2003: 150 drives



2005: 500 drives



2006: 950 drives

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

With 150 hard drives purchased on eBay we found:

- Thousands of credit card numbers
- Financial records
- Medical information
- Trade secrets
- Highly personal information



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

There are three primary techniques for assuring data confidentiality.

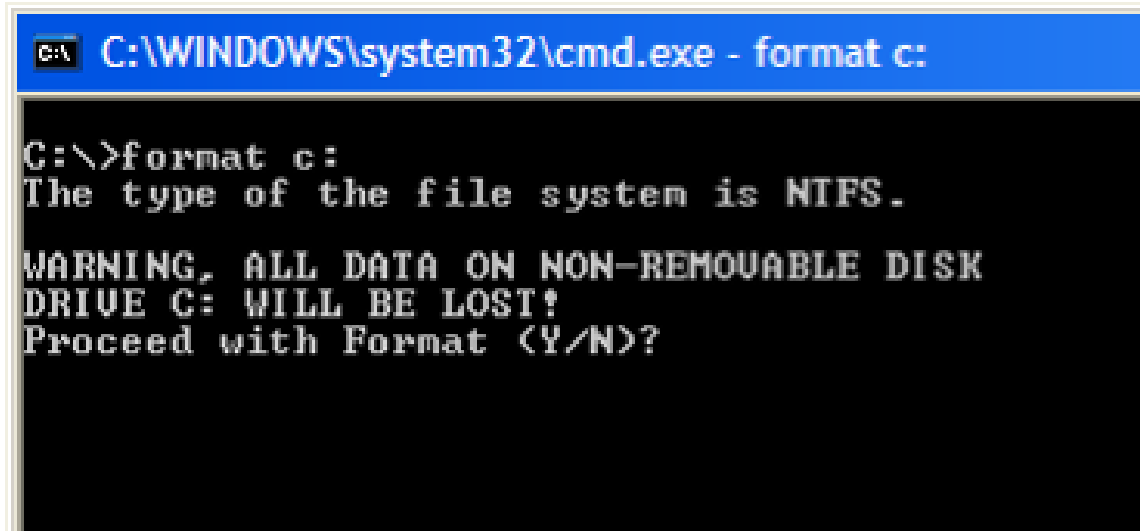
1. Physical security.
2. Logical access controls. (operating system)
3. Cryptography (disk & link)

These techniques don't work when a disk is thrown out or repurposed.

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

Most people don't encrypt their data.

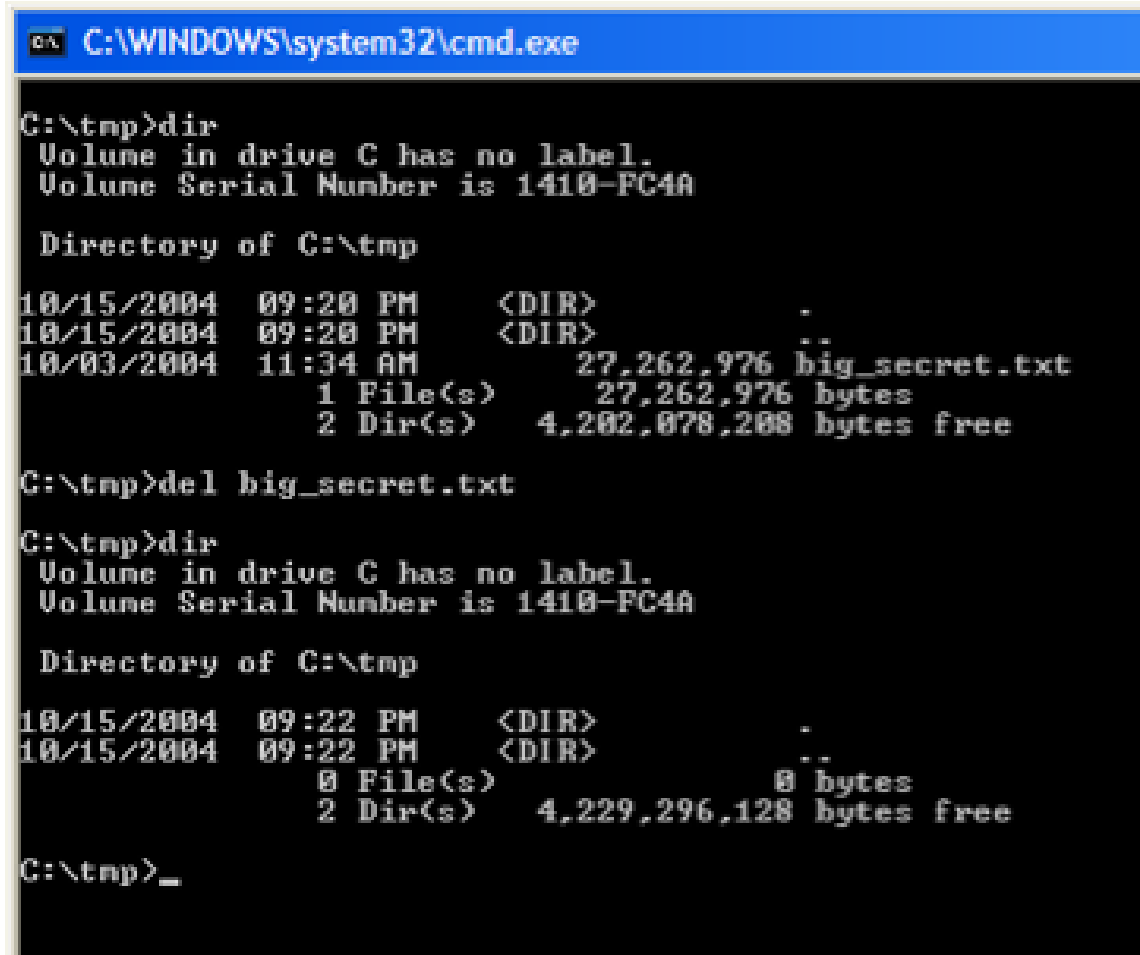
FORMAT C: doesn't erase the hard drive.



```
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 don't solve the problem
(but they do complicate data recovery.)**

Drives arrive by UPS and USPS



Drives are “imaged” with aimage.



Images stored on external firewire drives



900GB of storage holds 800 hard drive images

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

Purchased for \$5 from a Mass retail store on eBay

Copied the data off: 541MB

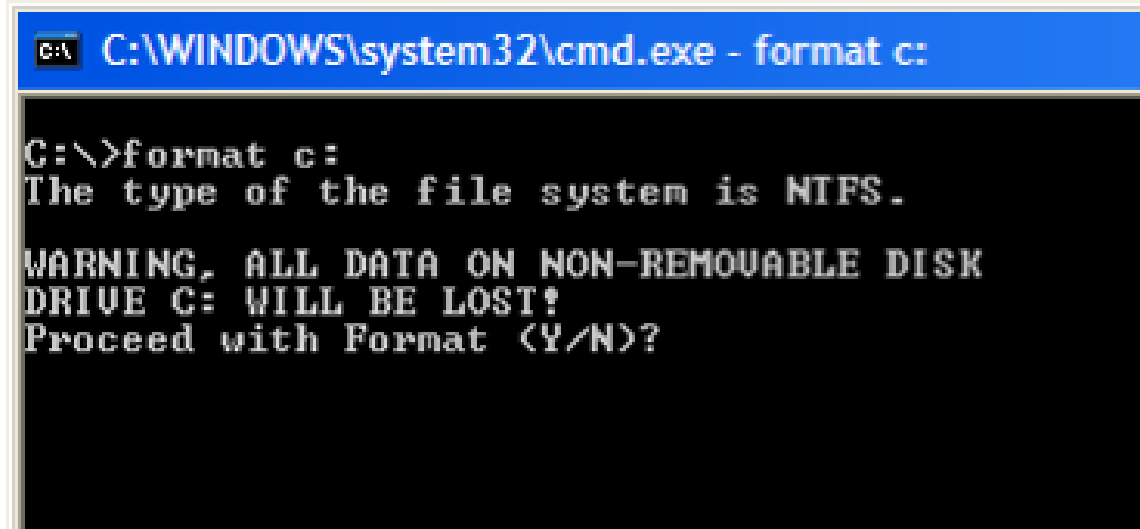
Initial analysis:

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 incompat

version of Windows. Insert a Startup diskette matching thi

OEMString = "NCR 14 inch Analog Color Display Enhanced 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

Data left behind in computer systems is a serious social problem.

Large numbers of drives are being sold and given away.

Many of them appear to have hidden confidential information.



Computer Science is morally obligated to solve this problem!

To be effective, a solution must address the root cause

Usability Problem:

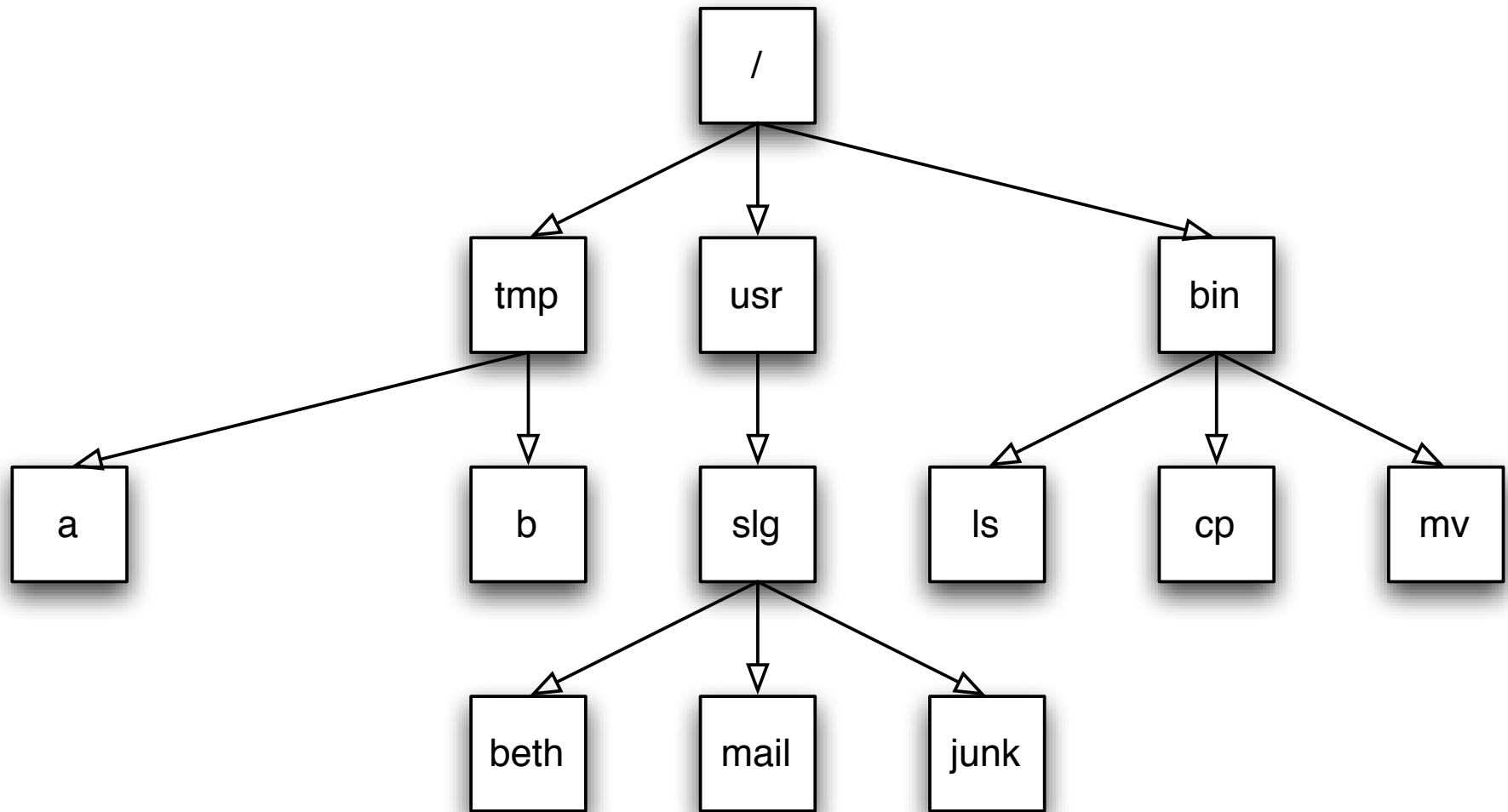
- Effective audit of information present on drives.
- Make DEL and FORMAT actually remove data.
[Bauer & Priyantha 01]
- Provide alternative strategies for data recovery.

Education Problem:

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

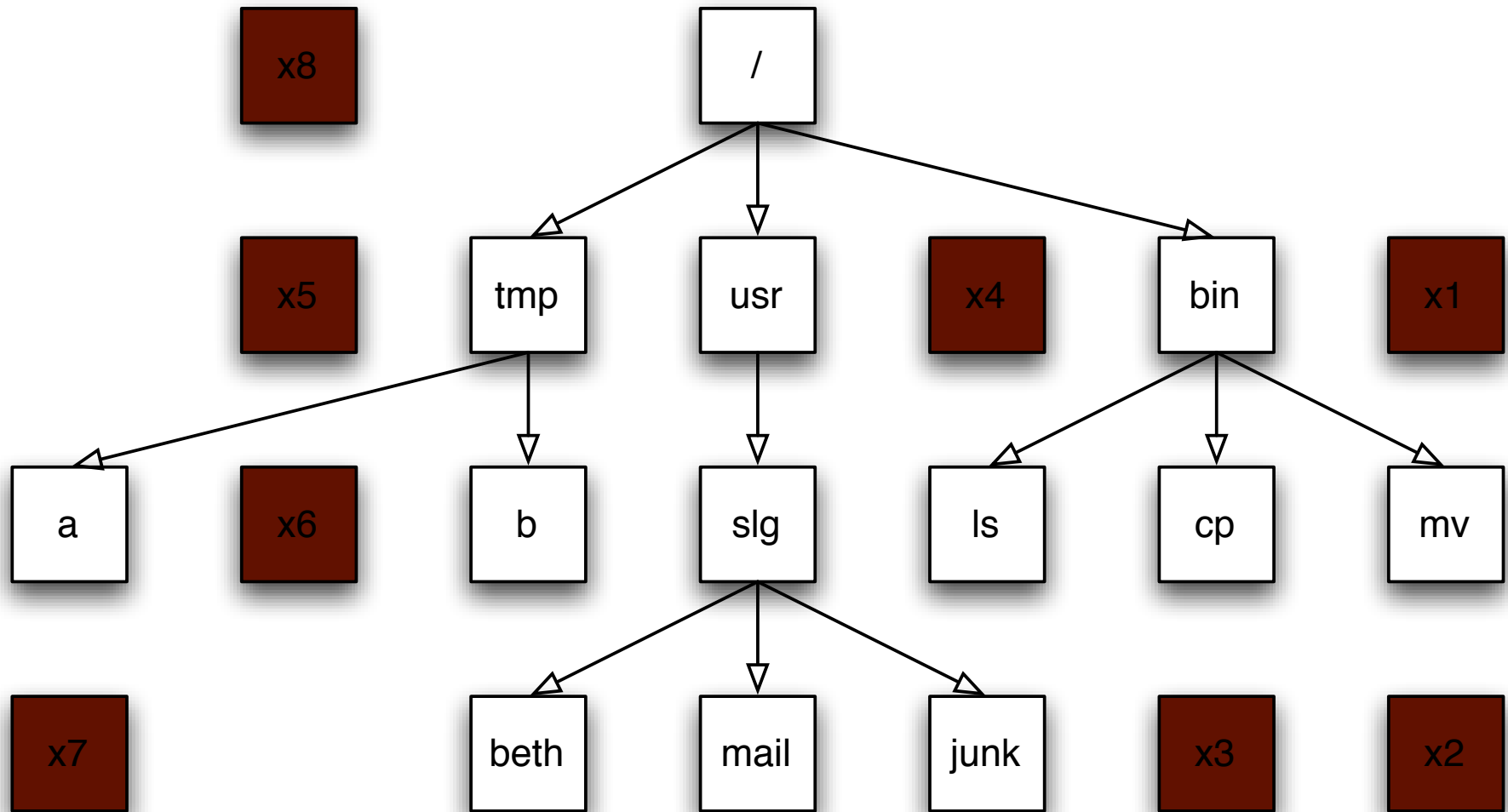
**To find that cause,
I looked *on the drives* and *contacted the data subjects*.**

Data on a hard drive is arranged in sectors.



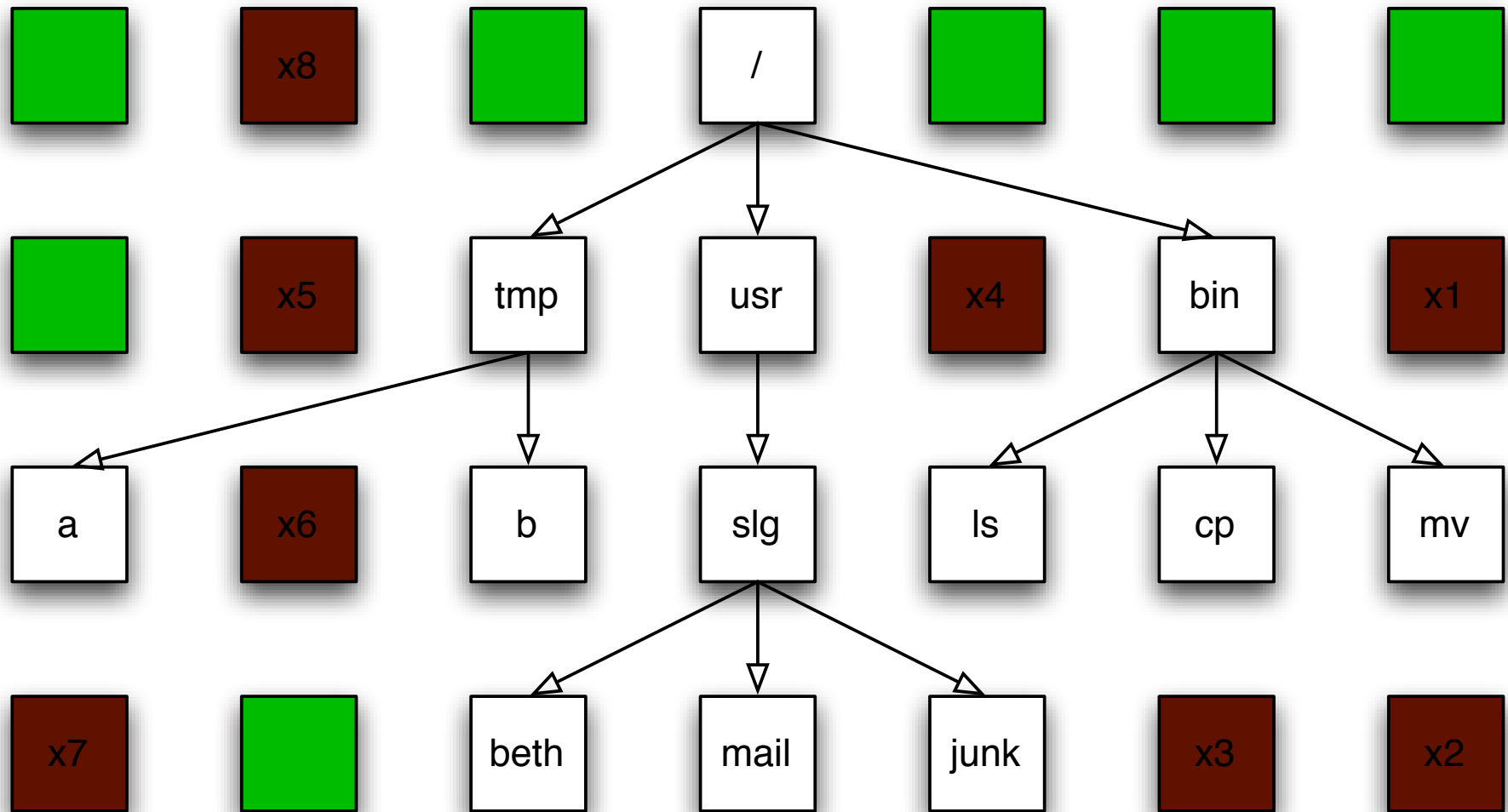
The white sectors indicate directories and files that are visible to the user.

Data on a hard drive is arranged in sectors.



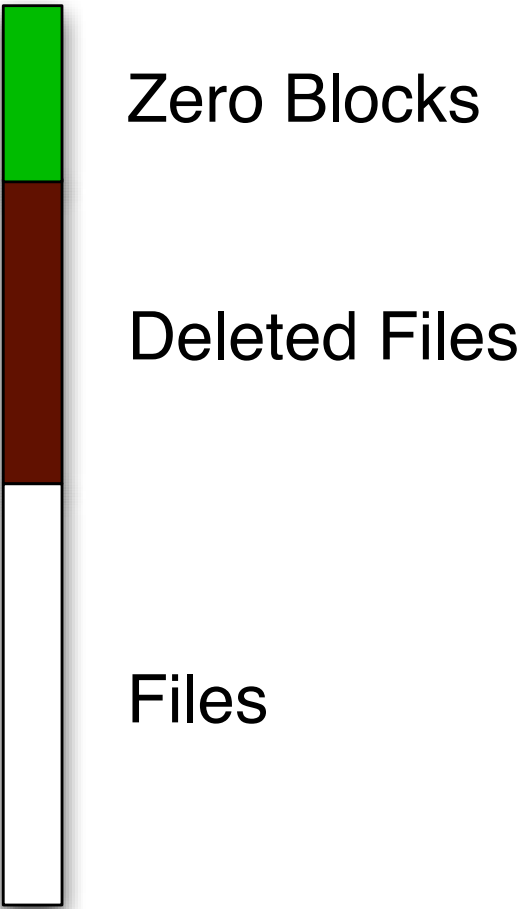
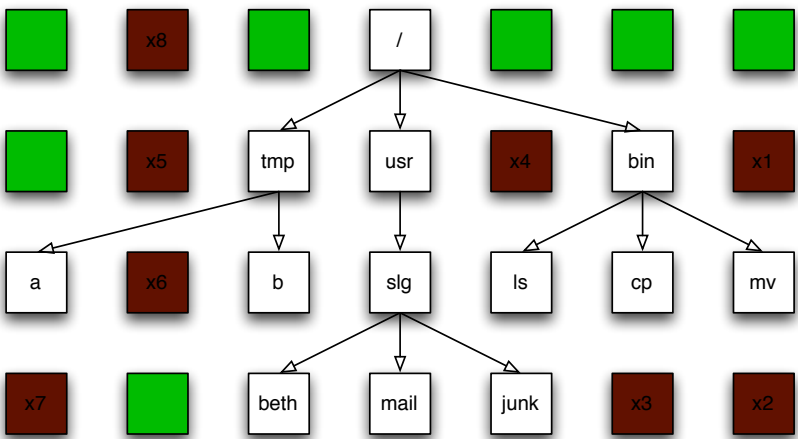
The brown sectors indicate files that were deleted.

Data on a hard drive is arranged in sectors.

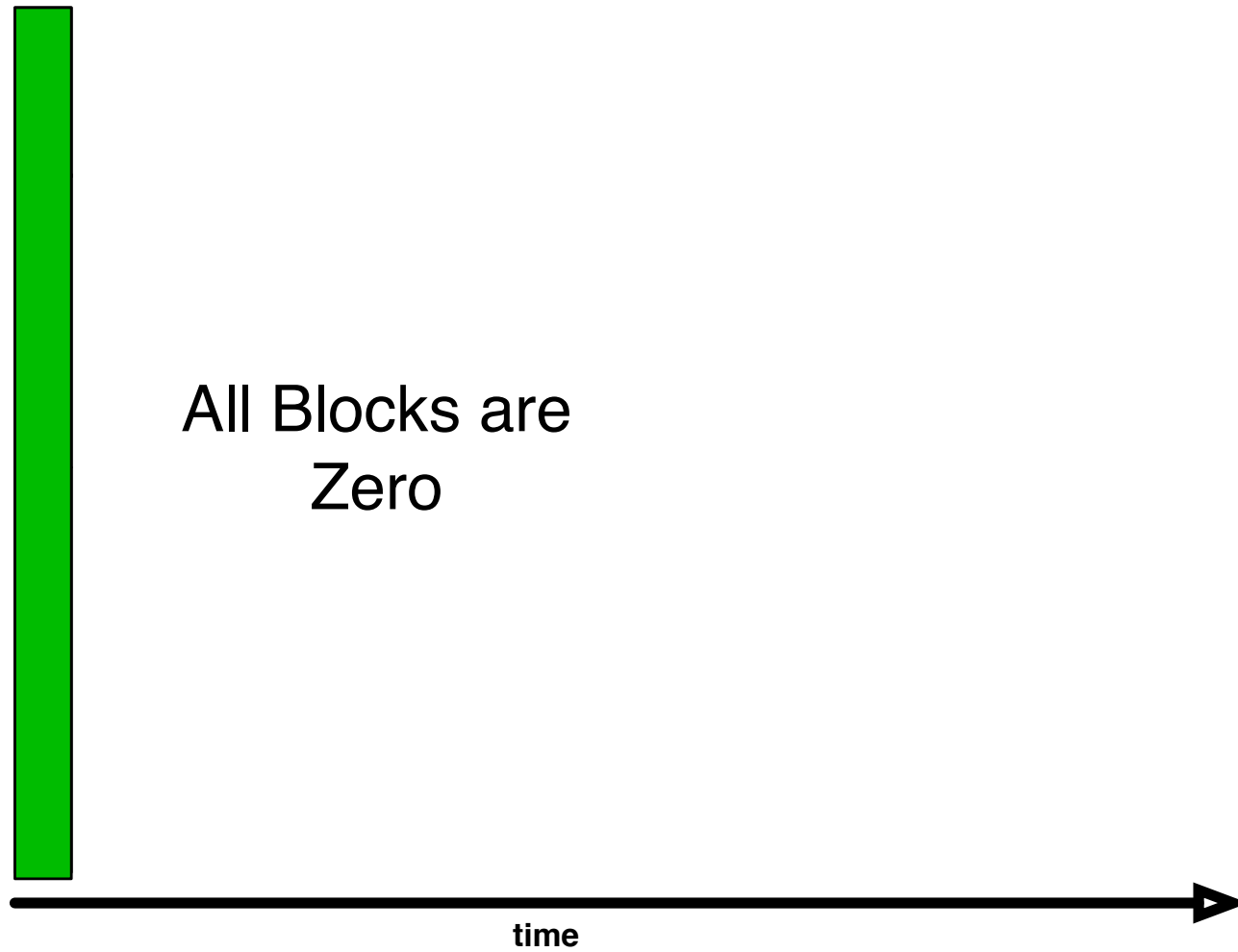


The green sectors indicate sectors that were never used (or that were wiped clean).

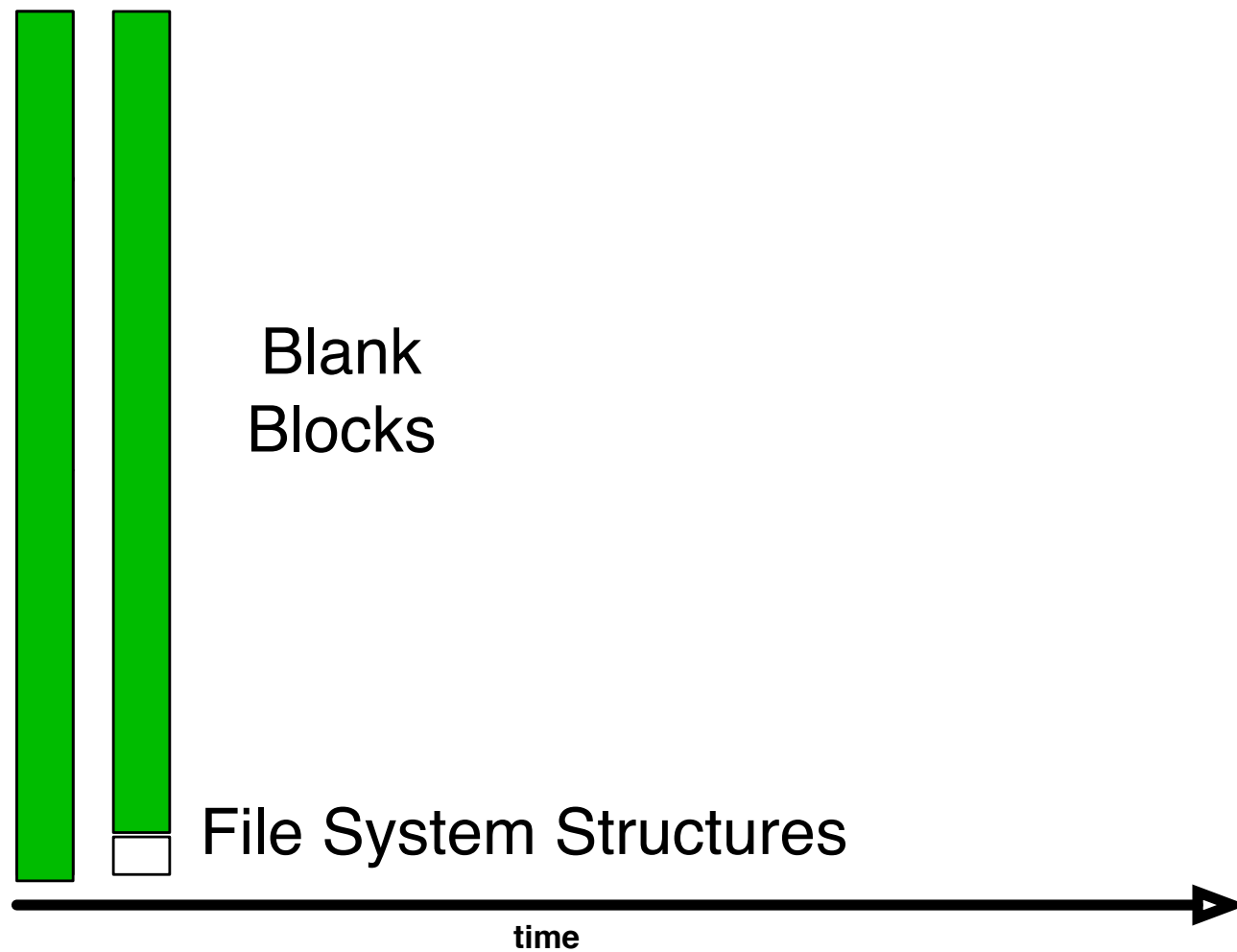
Stack the disk sectors:



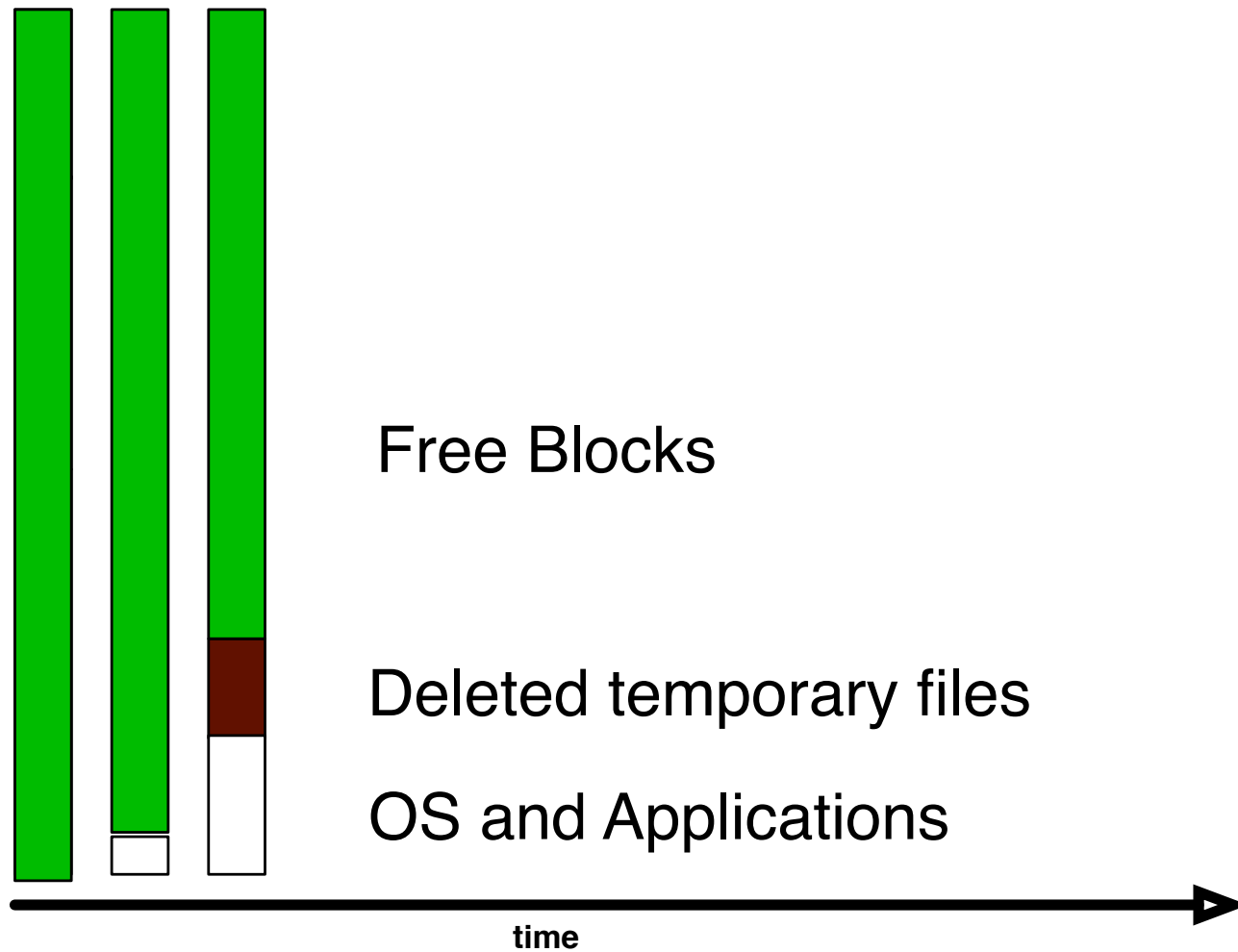
NO DATA: The disk is factory fresh.



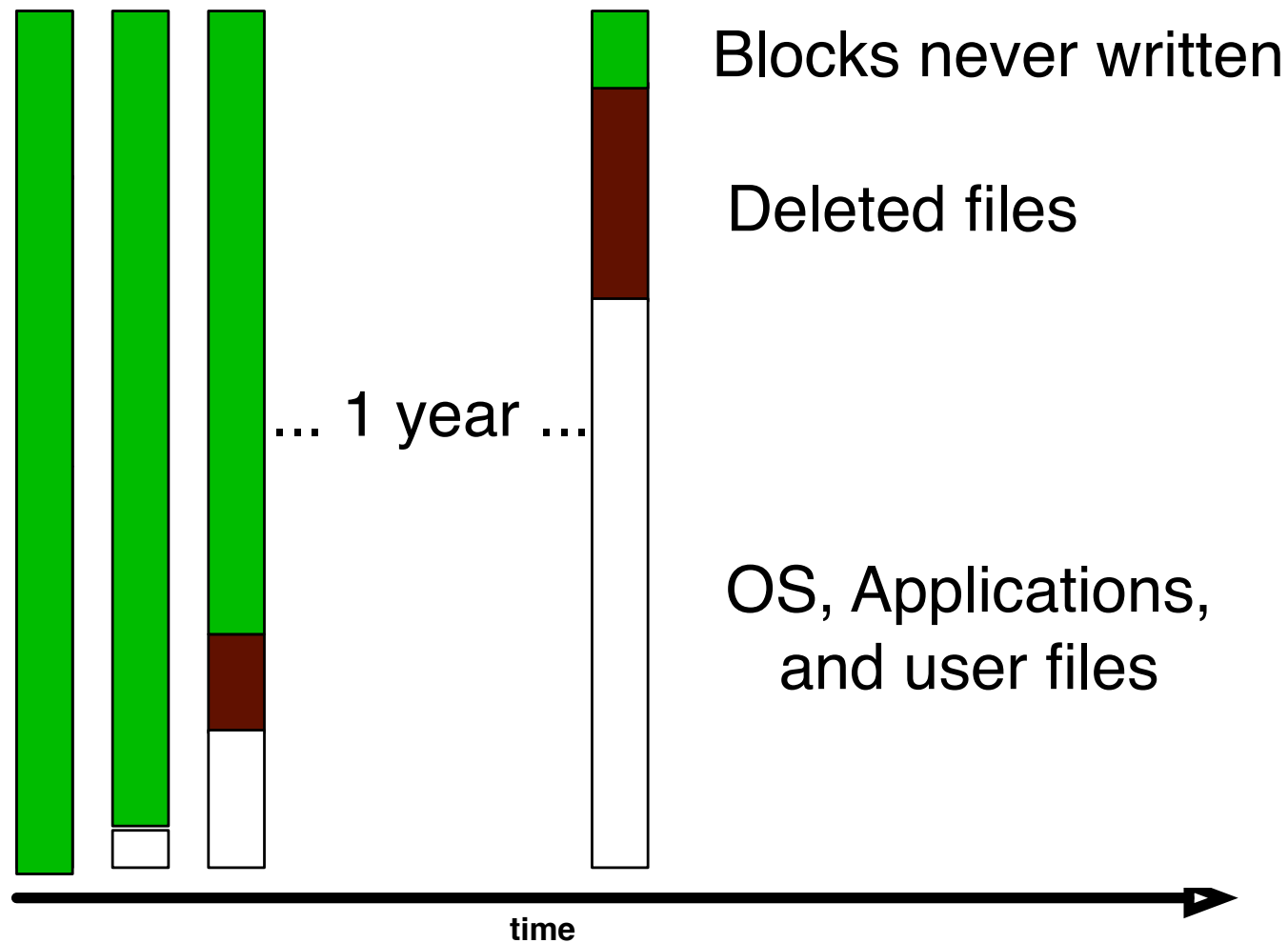
FORMATTED: The disk has an empty file system



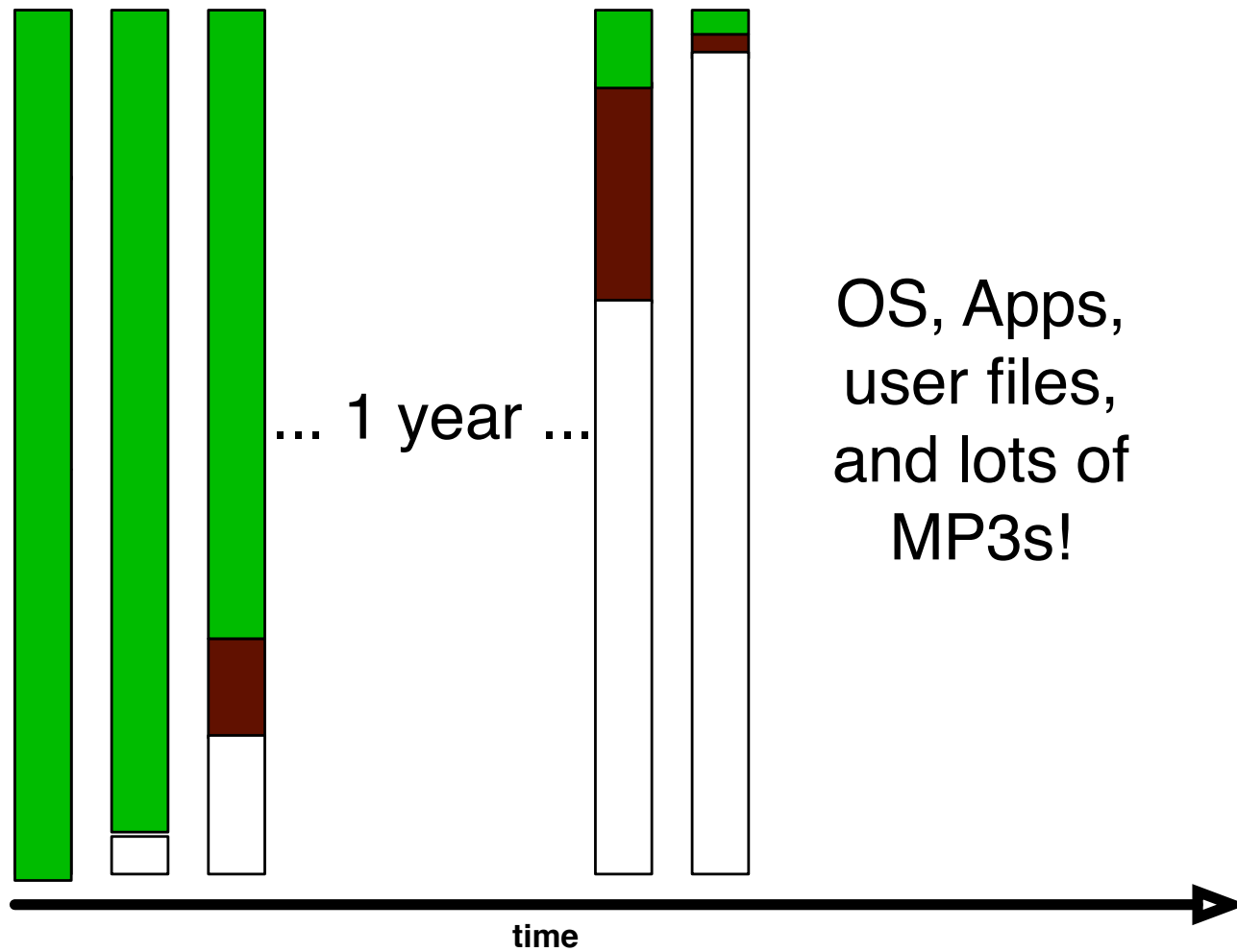
AFTER OS INSTALL: Temp. files have been deleted



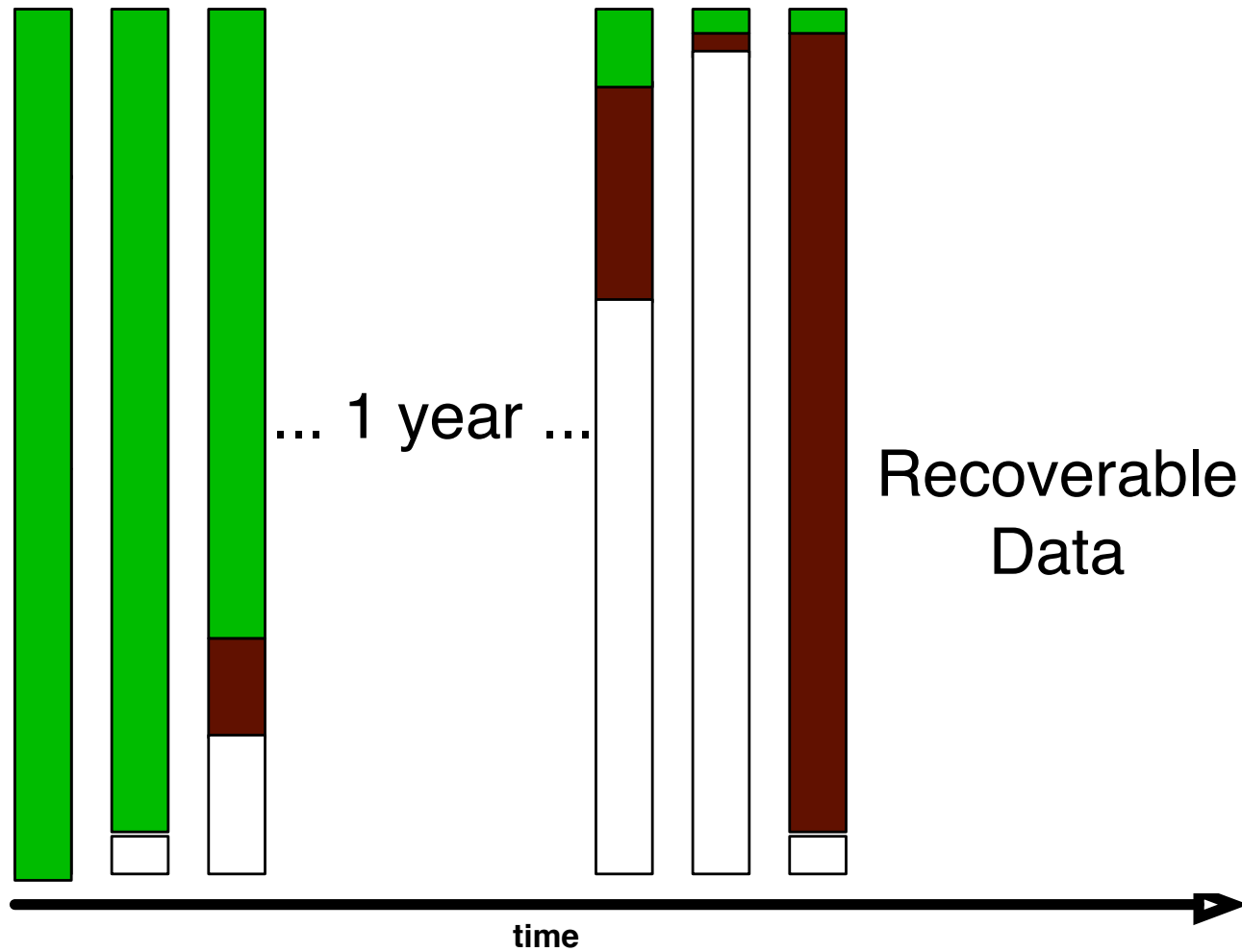
AFTER A YEAR OF SERVICE



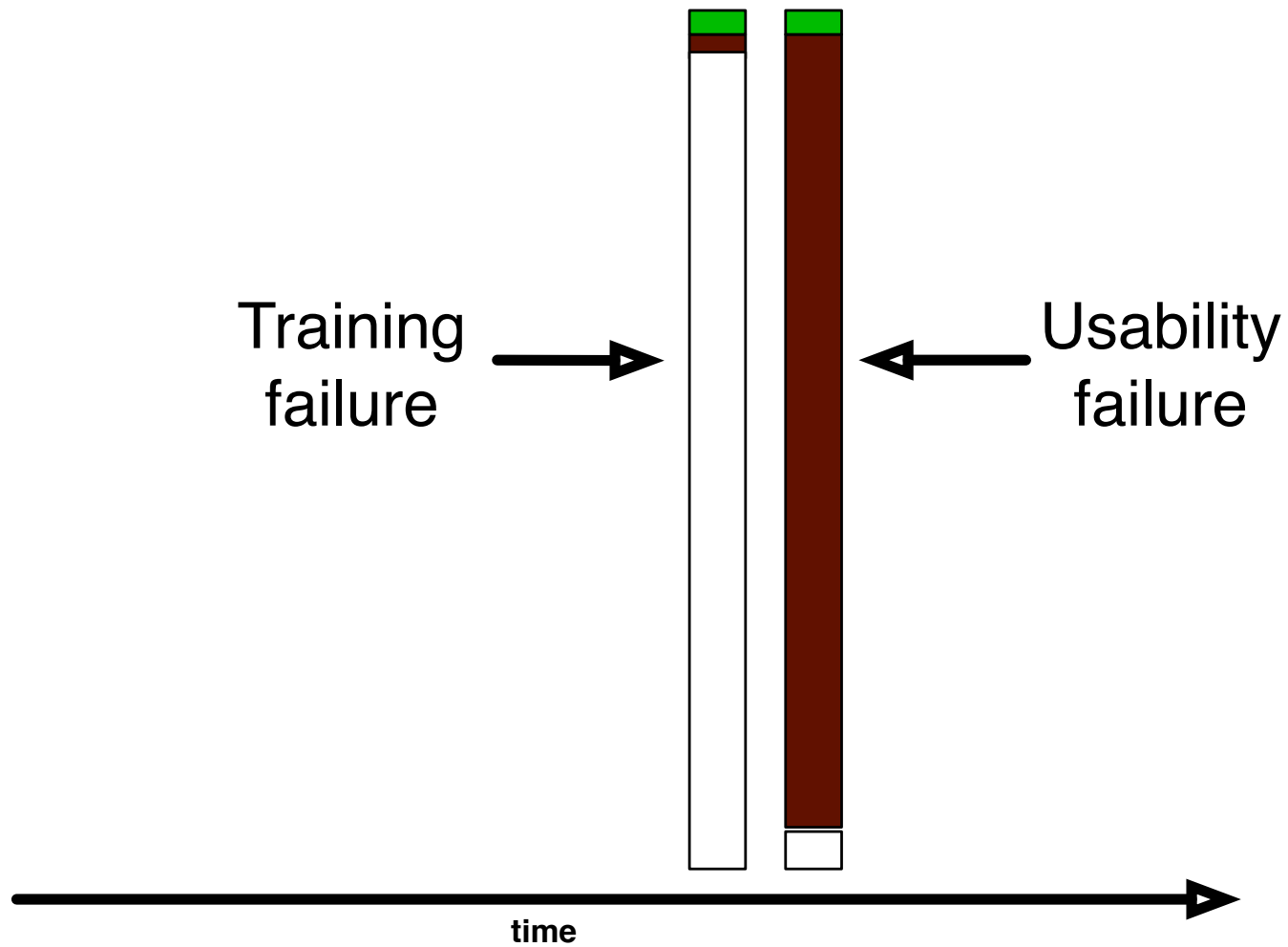
DISK NEARLY FULL!



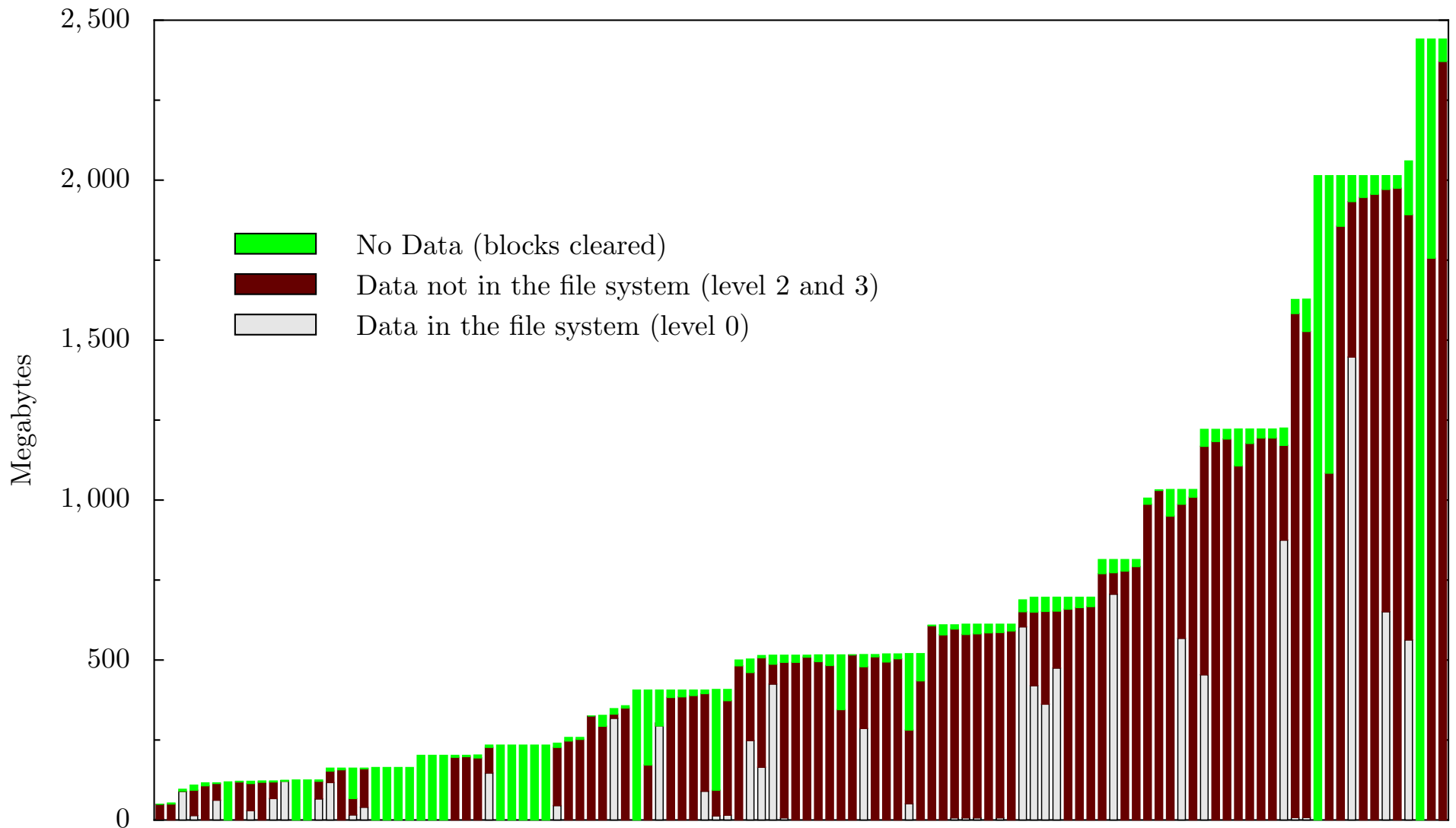
FORMAT C:\ (to sell the computer.)



We can use forensics to reconstruct motivations:



Drives I collected 1998-2003 are dominated by failed sanitization attempts...



..but training failures are also important.

But what *really* happened?



I needed to contact the original drive owners.

The *Remembrance of Data Passed Traceback Study.*

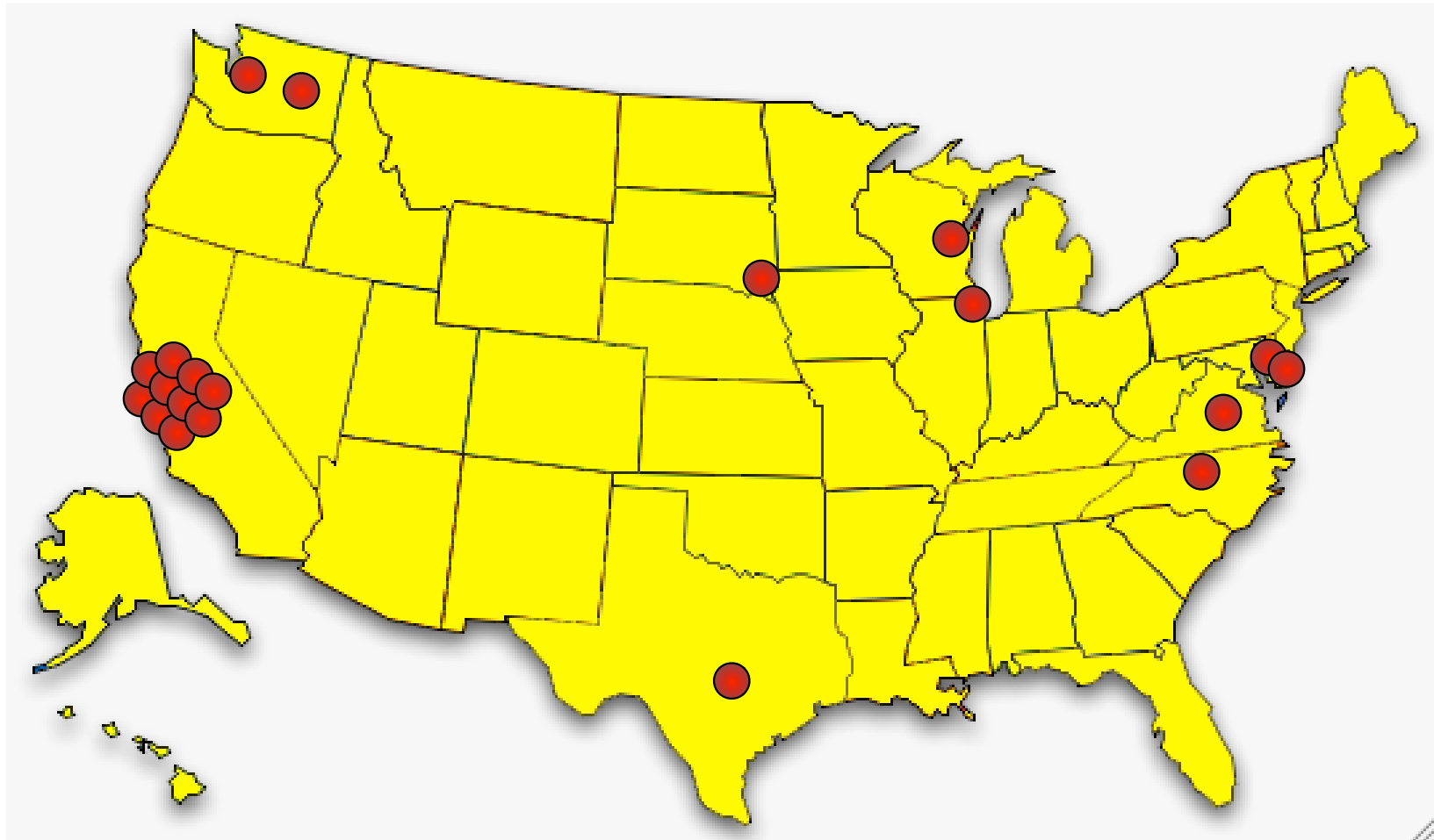
[Garfinkel 05]

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

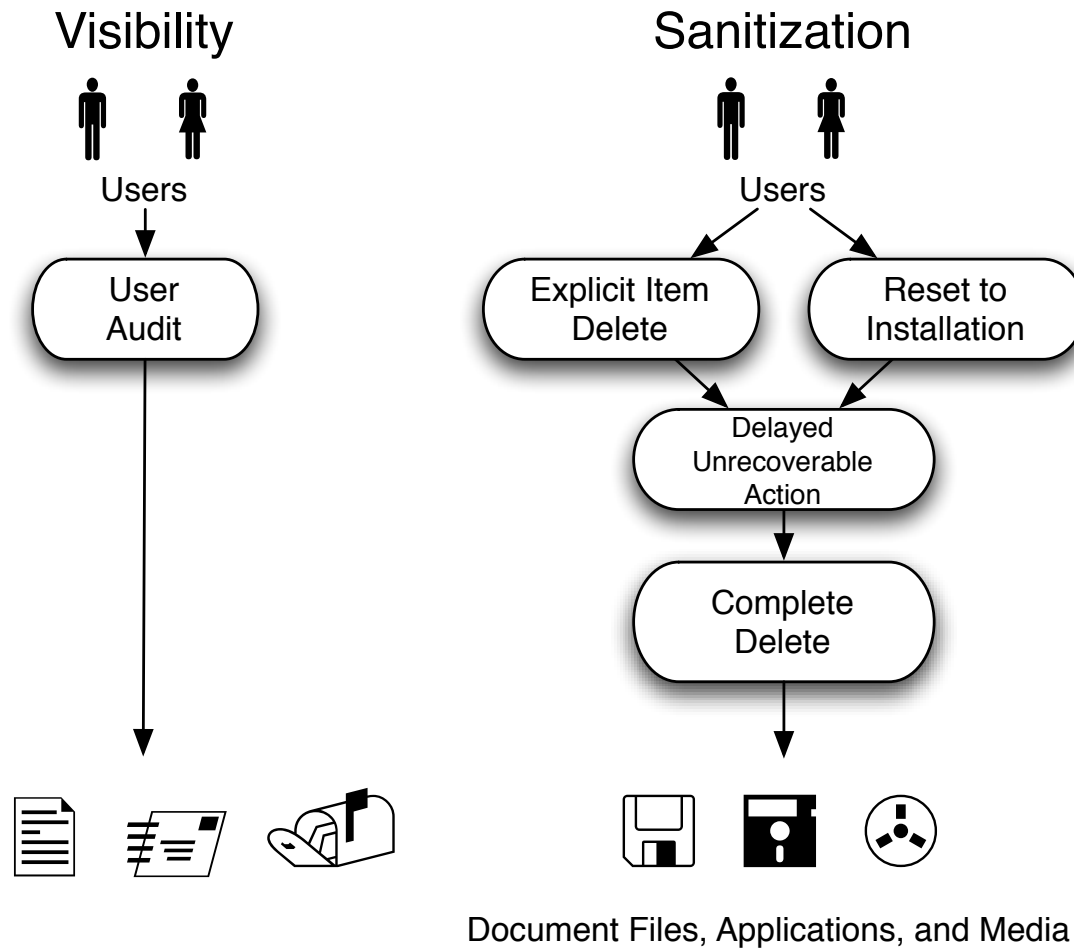
Lack of Concern: 2 cases

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

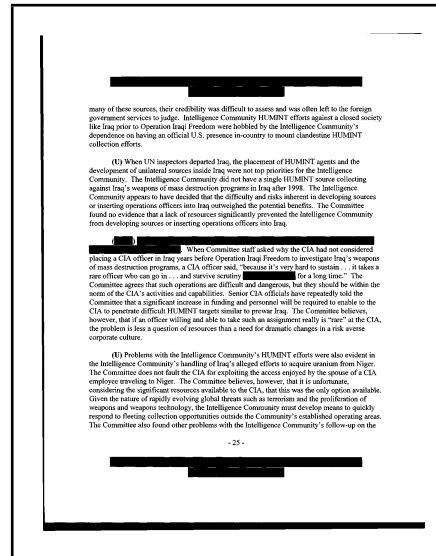
Regulation might have helped here, too.

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



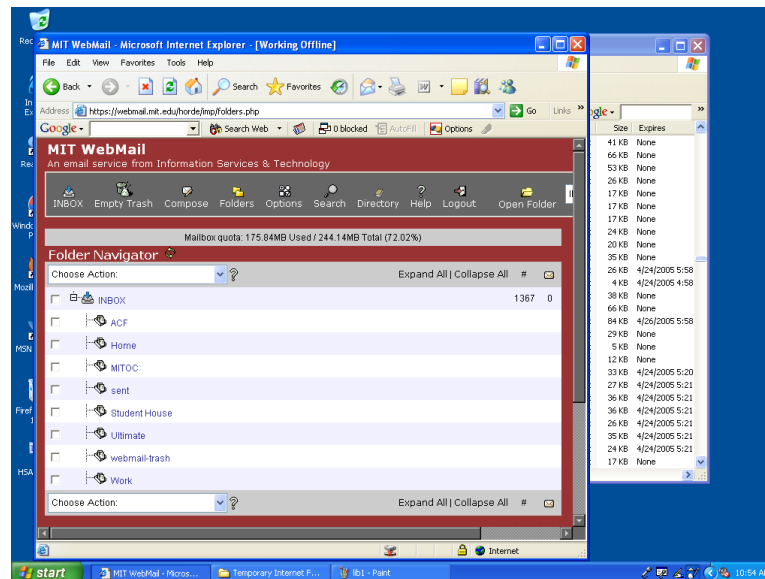
Naming these patterns is the first step to deployment.

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 forces in Iraq published classified information about insurgency methods.

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

Acrobat is literally a threat to national security.

UNCLASSIFIED

- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]

(Annex 11E).

3. (U) Insurgent TTPs for VBIEDs

(U) There are two basic types of car bombs, i.e., [REDACTED]. Both can be either command or remote-detonated. (Annex 8E).

[REDACTED]

(U) The techniques for employing VBIEDs continue to evolve. Some of the more commonly used techniques include:

6

UNCLASSIFIED

UNCLASSIFIED

easy to emplace by staging equipment in vehicles or near overpasses, and, in a matter of minutes, having the IED armed and in the desired location.

- (S//NF) Explosives wrapped in a brown paper bag or a plastic trash bag. This is a particularly easy method of concealment, easy to emplace, and has been used effectively against Coalition Forces and civilians along Route Irish.
- (S//NF) Explosives set on a timer. This technique is new to the Route Irish area, but is being seen more frequently.
- (S//NF) Use of the median. The 50 meter wide median of Route Irish provides a large area for emplacing IEDs. These can be dug in, hidden, and/or placed in an animal carcass or other deceptive container.
- (S//NF) Surface laid explosives. The enemy will drop a bag containing the explosive onto the highway and exit the area on an off-ramp with the detonation occurring seconds or minutes later depending on the desired time for the explosion.
- (S//NF) Explosives on opposite sides of the median. Devices have been found along both sides of the median that were apparently designed to work in tandem, to counter Coalition Force tactics to avoid the right side of the highway while traveling Route Irish.
- (S//NF) Explosives hidden under the asphalt. Insurgents pretend to do work on the pavement, plant the explosives, and repair the surface. These are usually remote-detonated devices.

(Annex 11E).

3. (U) Insurgent TTPs for VBIEDs

(U) There are two basic types of car bombs, i.e., suicide (where the car is moving) and stationary (where the car is parked). Both can be either command or remote-detonated. (Annex 8E).

(S//NF) The enemy is very skillful at inconspicuously packing large amounts of explosives into a vehicle. The most commonly used detonation materials are plastic explosives and 155mm artillery shells. When moving, these VBIEDs are practically impossible to identify until it is too late. (Annex 8E).

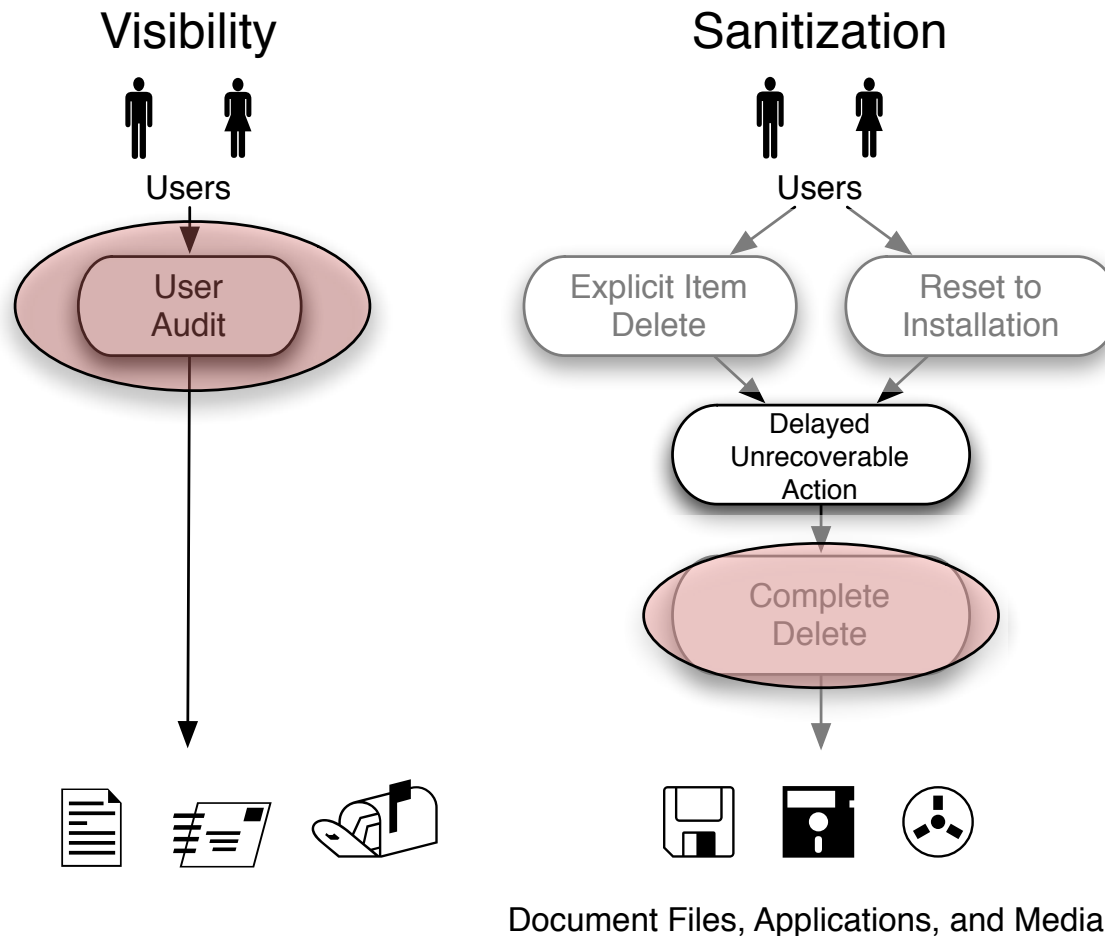
(U) The techniques for employing VBIEDs continue to evolve. Some of the more commonly used techniques include:

6

UNCLASSIFIED

NSA recently published a “how to sanitize” guide.

The information leaks because two patterns were not implemented.

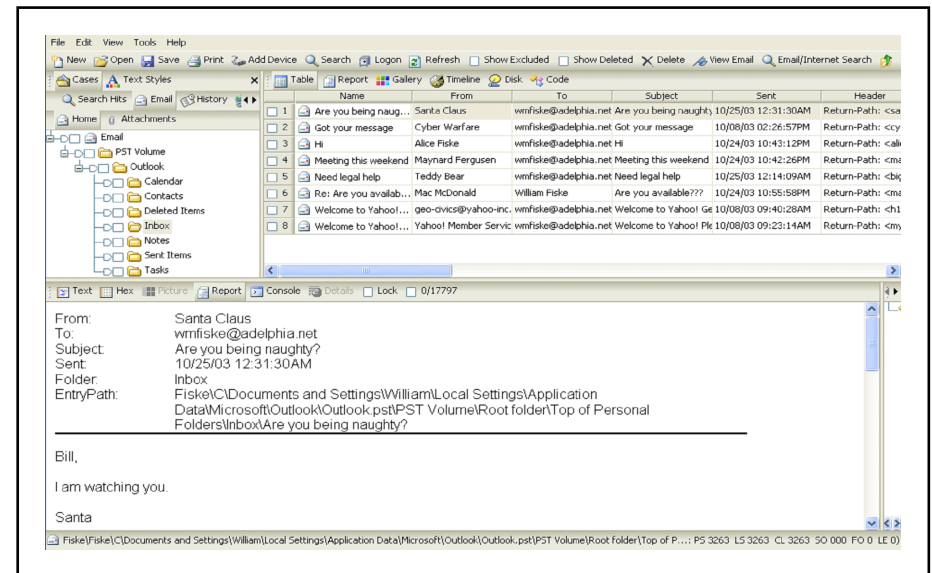


Current agenda: working with vendors to implement these patterns.

The techniques developed for [Garfinkel '05] are different than traditional forensics techniques.

Traditional forensics tools:

- Interactive user interface.
- Recovery of “deleted” files.
- Generation of “investigative reports” for courtroom use.
- Focus on one or a few disks.



In [Garfinkel '05], there were *hundreds* of disks to analyze.

Today's tools choke when confronted with thousands of disks.

- Has this drive been previously imaged?
- Which drives belong to my target?
- Do any drives belong to my target's associates?
- Where should I start?



**Today's tools are for criminal investigations.
Increasingly, we need tools for intelligence analysis.**

Intelligence objectives can be furthered by correlating information from multiple drives.

- Where any drives were used by the same organization?
- What names/places/email addresses are in common?
- Which drives were used in a place or at a time of interest?



Example problem: Who owned this disk drive?

Approach #1: Look for Microsoft Word files and try to determine the owner.

- Needs forensic skill.
- Requires complete documents.

Approach #2: Compute a histogram of all email addresses.

- Works with any file system.
- Works with incomplete data.

The email histogram works even if you can't find any files.

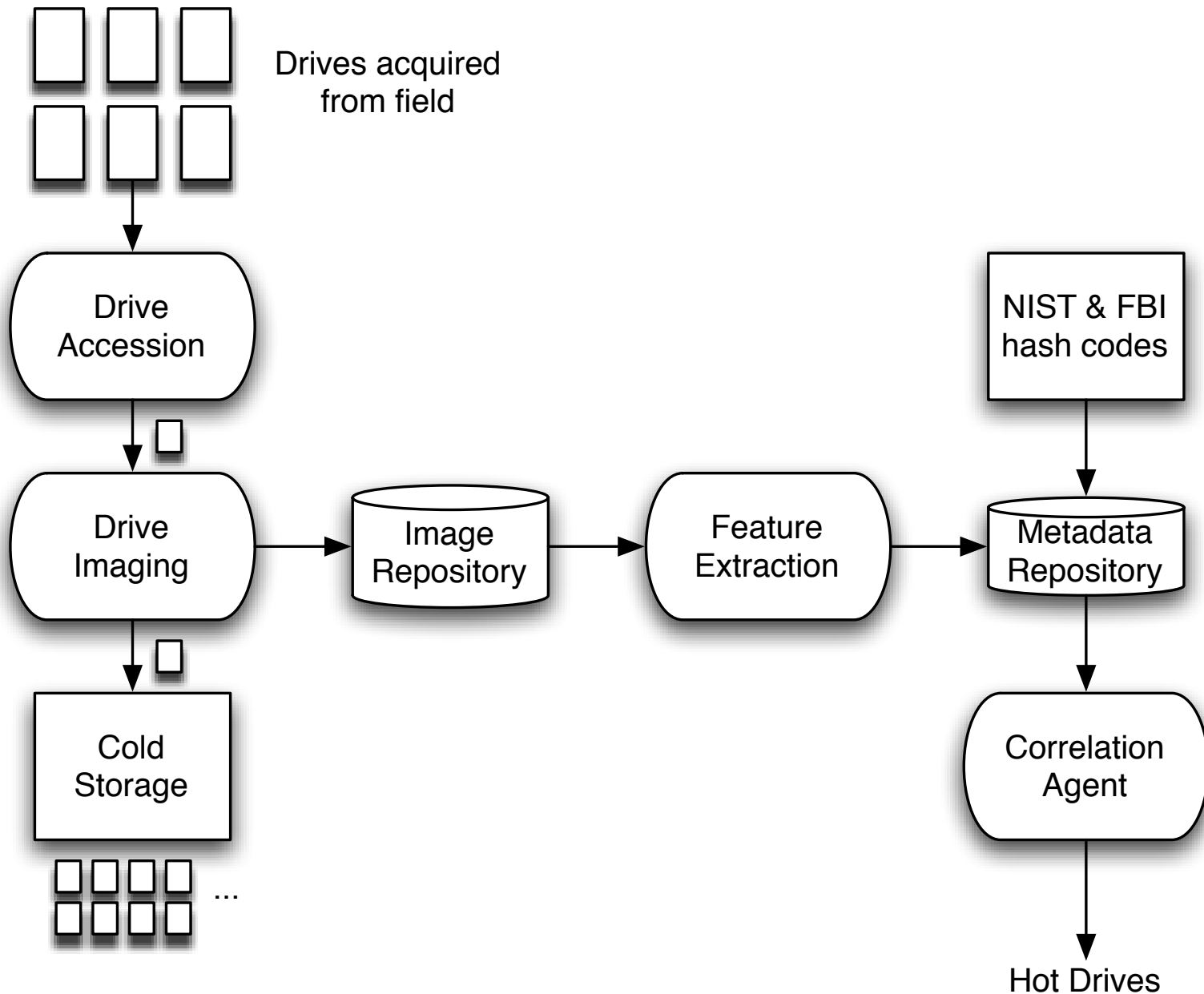
The email histogram approach works quite well.

Drive #51: Top email addresses (sanitized)

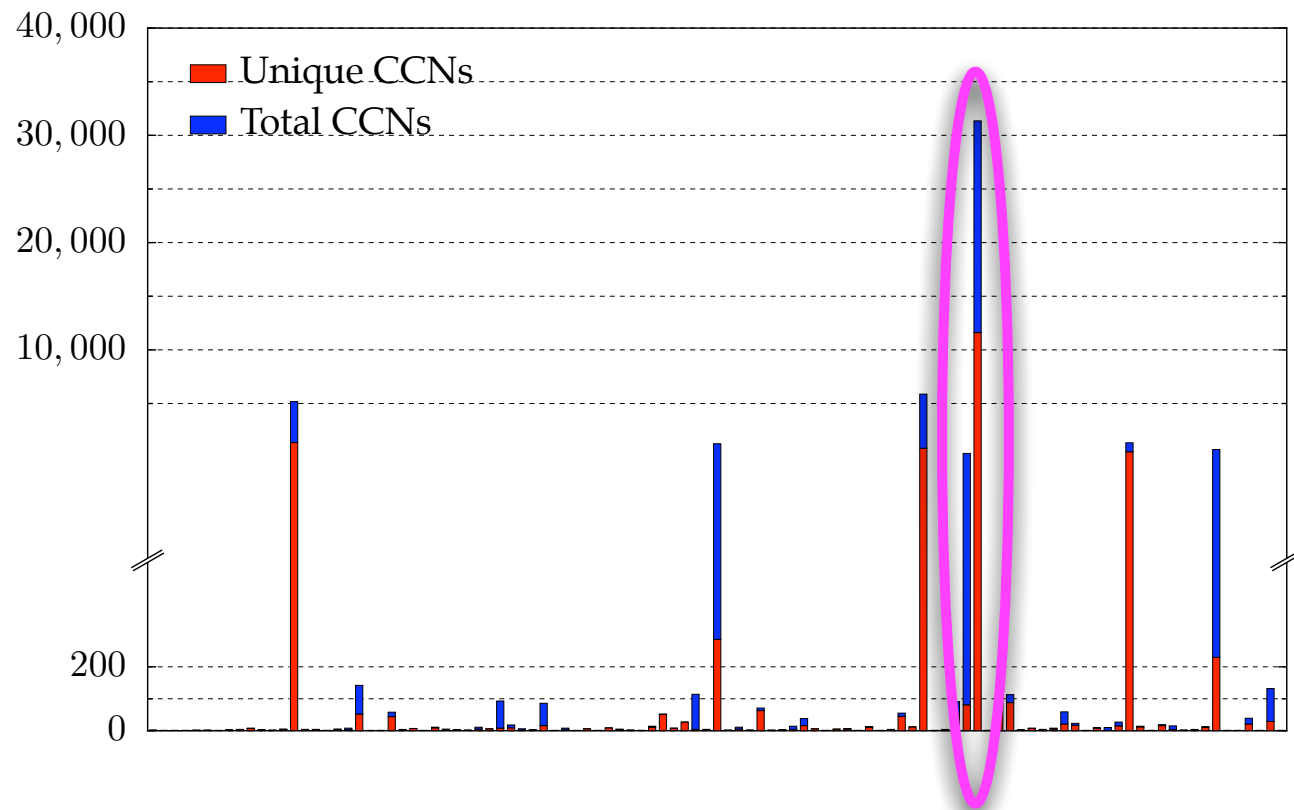
Count	Address(es)
8133	ALICE@DOMAIN1.com
3504	BOB@DOMAIN1.com
2956	ALICE@mail.adhost.com
2108	JobInfo@alumni-gsb.stanford.edu
1579	CLARE@aol.com
1206	DON317@earthlink.net
1118	ERIC@DOMAIN1.com
1030	GABBY10@aol.com
989	HAROLD@HAROLD.com
960	ISHMAEL@JACK.wolfe.net
947	KIM@prodigy.net
845	ISHMAEL-list@rcia.com
802	JACK@nwlink.com
790	LEN@wolfenet.com
763	natcom-list@rcia.com

(Can we automatically sanitize this kind of information?)

Cross-Drive Forensics systematizes this approach.



“First Order Cross-Drive Forensics” analyzes each drive with a filter.



Drives with high response warrant further attention.

Example: The Credit Card Number Detector.

The CCN detector scans bulk data for ASCII patterns that look like credit card numbers.

- CCNs are found in certain typographical patterns.
(e.g. XXXX-XXXX-XXXX-XXXX
or XXXX XXXX XXXX XXXX
or XXXXXXXXXXXXXXXXXXXX)
- CCNs are issued with well-known prefixes.
- CCNs follow the Credit Card Validation algorithm.
- Certain numeric patterns are unlikely.
(e.g. 4454-4766-7667-6672)

CCN detector: written in flex and C++

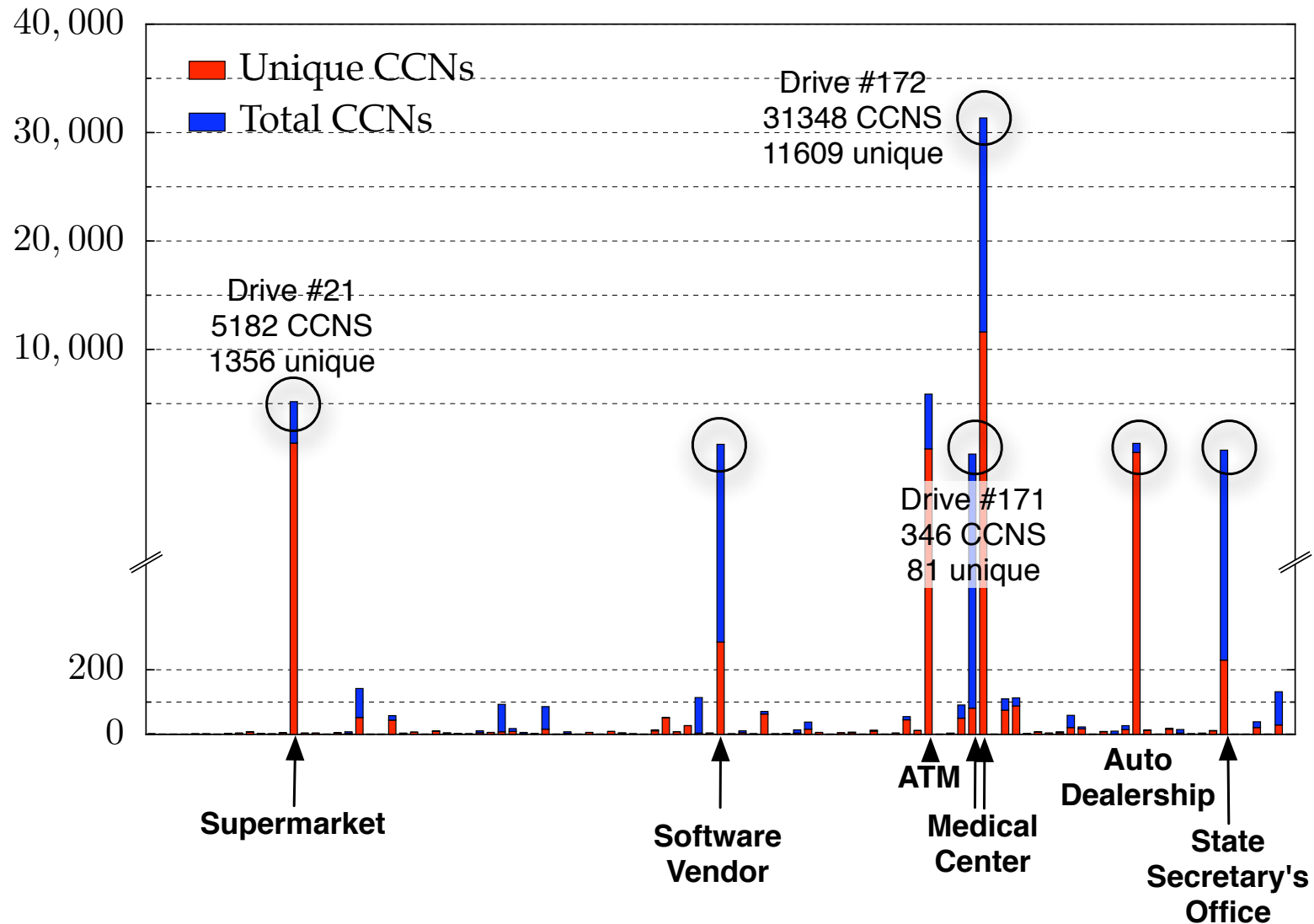
Scan of disk #105: (642MB)

Test	# pass
typographic pattern	3857
known prefixes	90
CCV1	43
numeric histogram	38

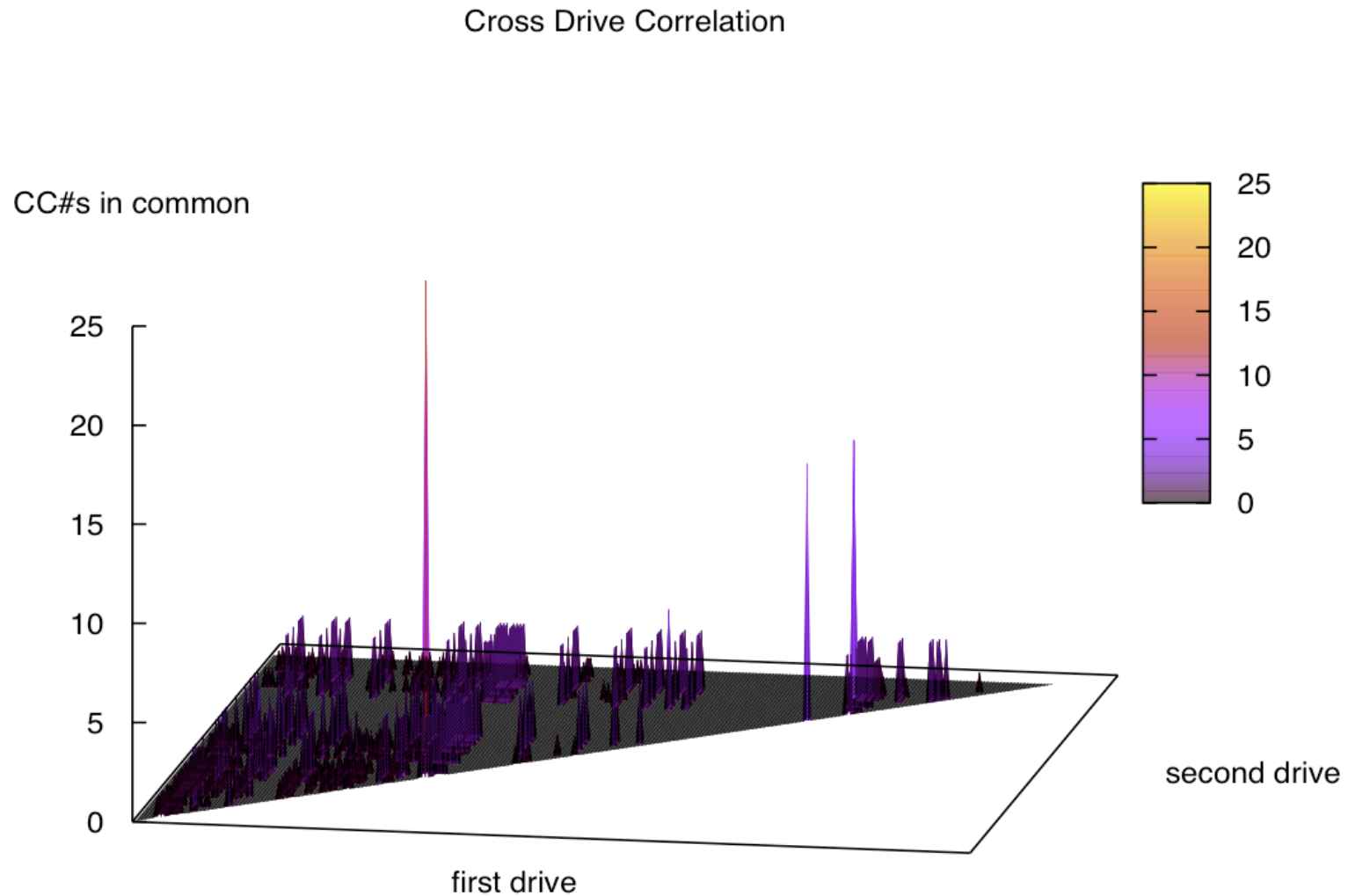
Sample output:

```
'CHASE NA|5422-4128-3008-3685|    pos=13152133
'DISCOVER|6011-0052-8056-4504|    pos=13152440
.'GE CARD|4055-9000-0378-1959|    pos=13152589
BANK ONE |4332-2213-0038-0832|    pos=13152740
.'NORWEST|4829-0000-4102-9233|    pos=13153182
'SNB CARD|5419-7213-0101-3624|    pos=13153332
```

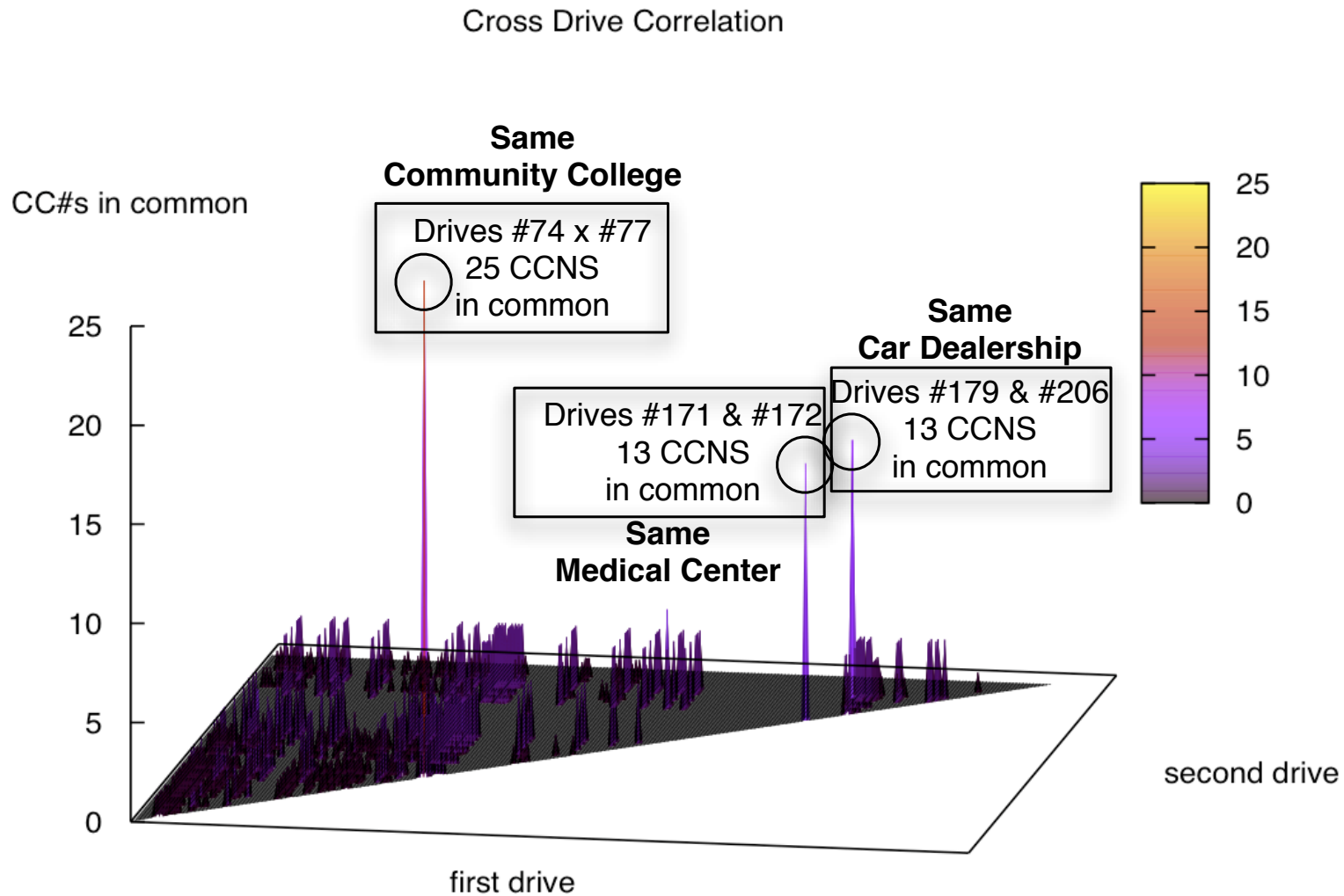
With a “credit card number detector,” we can rapidly identify drives with leaked consumer information.



Second-order analysis uses correlation techniques to identify drives of interest.

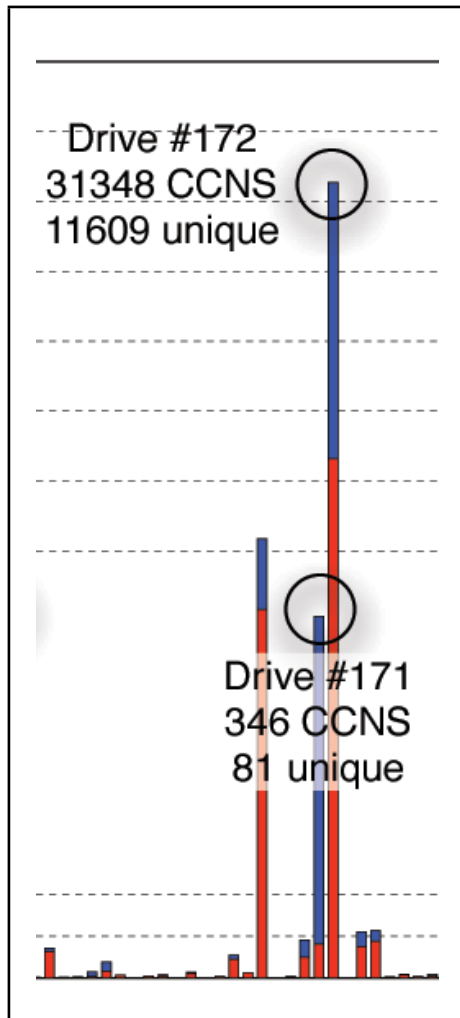


Second-order analysis uses correlation techniques to identify drives of interest.



In this example, three pairs of drive appear to be correlated.

Let's look at drives #171 and #172 again.



Cross-drive analysis tells us that #171 and #172 are from the same medical center.

Drive #171: Development drive

- Has source code.
- 346 CCNS; 81 unique.

Drive #172: Production system.

- 31,348 CCNS; 11,609 unique
- Oracle database (hard to reconstruct).

The programmers used live data to test their system.

Second-order analysis:

Identifiers:

- CCNs
- Email addresses
- Message-IDs
- sector hashes

Possible Uses:

- Identifying new social networks
- Testing for inclusion in an existing network.
- Measuring dissemination of information

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.

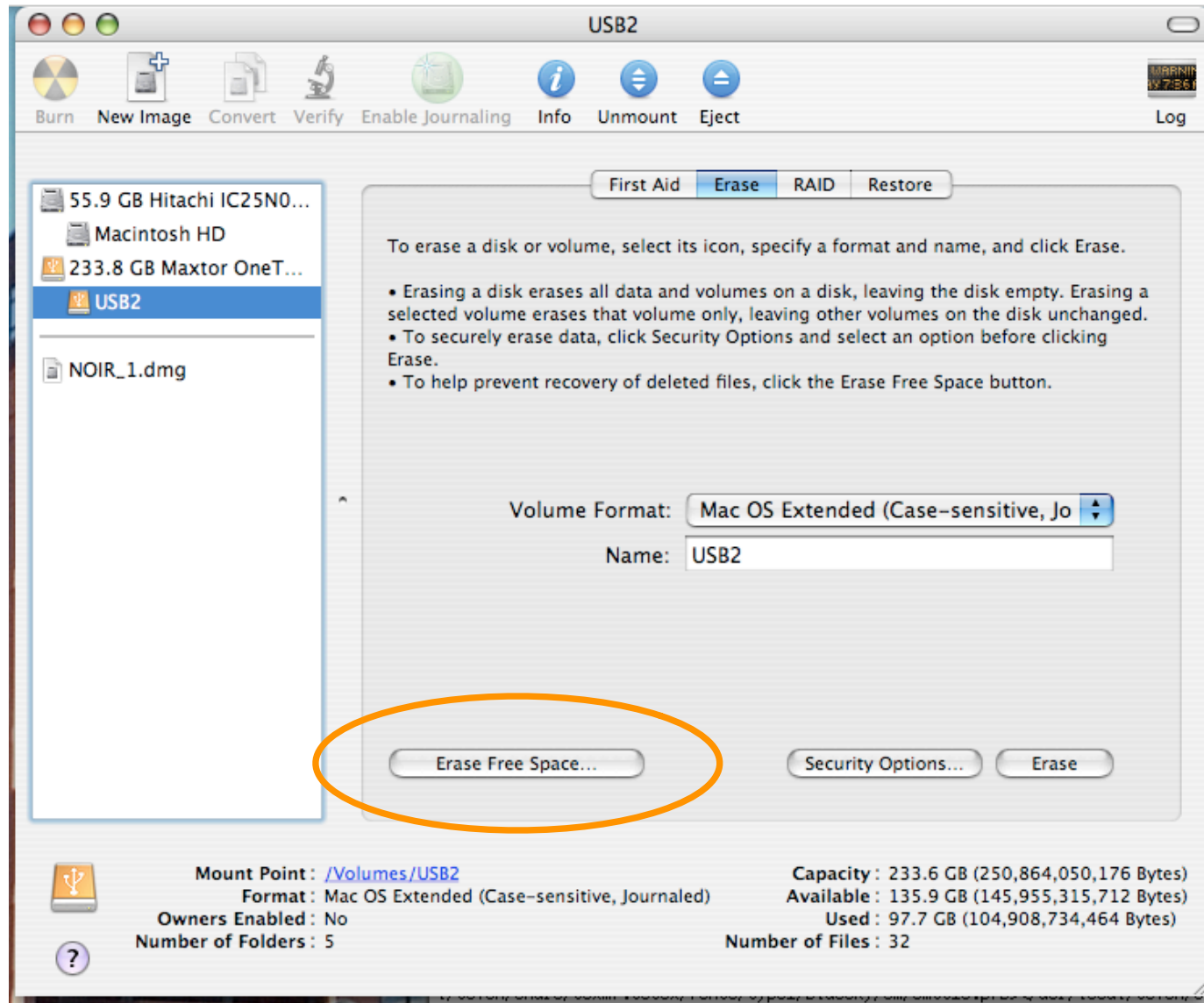


Unfortunately, “Secure Empty Trash” is incomplete.

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



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



Research Agenda: Forensic Tools

- Improved tools for acquisition and analysis tools.
- Improved feature extractors.
- Augmenting cross-drive analysis with machine learning & datamining.
- Support for languages other than English.

Download tools from <http://www.afflib.org/>

Research Agenda: Economics and Society

- Who is buying used hard drives and why?
- Time-based analysis.
- Compliance with FACT-A.

Research Agenda: Privacy-Enhancing Tools

- Analysis of “one big file” technique.
- Make FORMAT actually erase the disk.
- Make “Empty Trash” actually overwrite data.
- Add complete delete to web browsers, word processors, operating systems.
- Improved backup/synchronization.

Research Summary

- Patterns for aligning usability & security.
<http://www.simson.net/thesis/>
- AFFLIB: Formats and tools for disk forensics.
<http://www.afflib.org>
- 1000 Drive Corpus (→ 2500 drives).
- Societal impacts.
- S/MIME adoption.
- Privacy models for data provenance and DRM.

Questions?