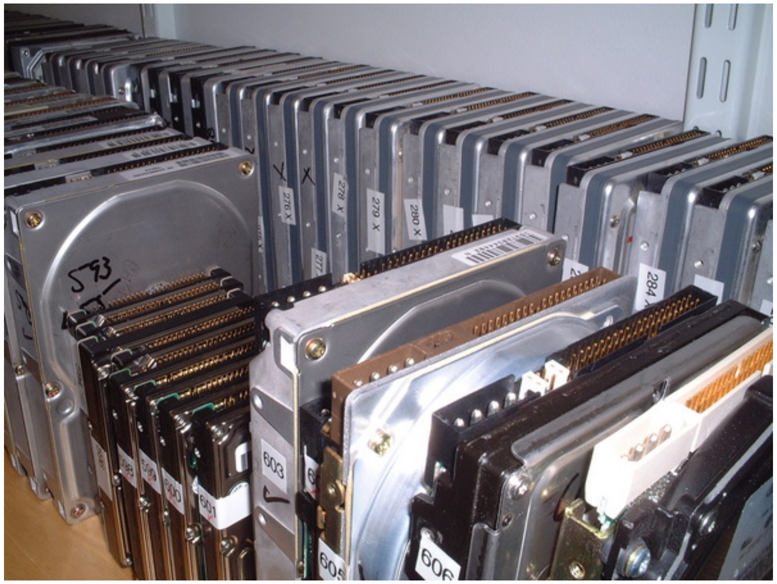
Complete Delete



Simson L. Garfinkel

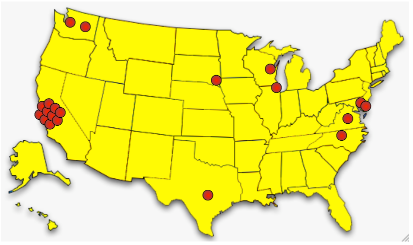
November 27, 2006
11:00am

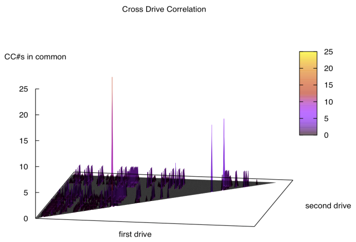
Postdoctoral Fellow,
Center for Research on Computation and Society
Harvard University

Associate Professor,
Naval Postgraduate School
Monterey, CA

This talk presents new tools and techniques for performing forensic analysis on a large number of disk drives.

The drives Project

The Traceback Study 

Cross Drive Forensics and AFF 

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?

Hard drives pose special problem for computer security

Do not forget data when power is removed.

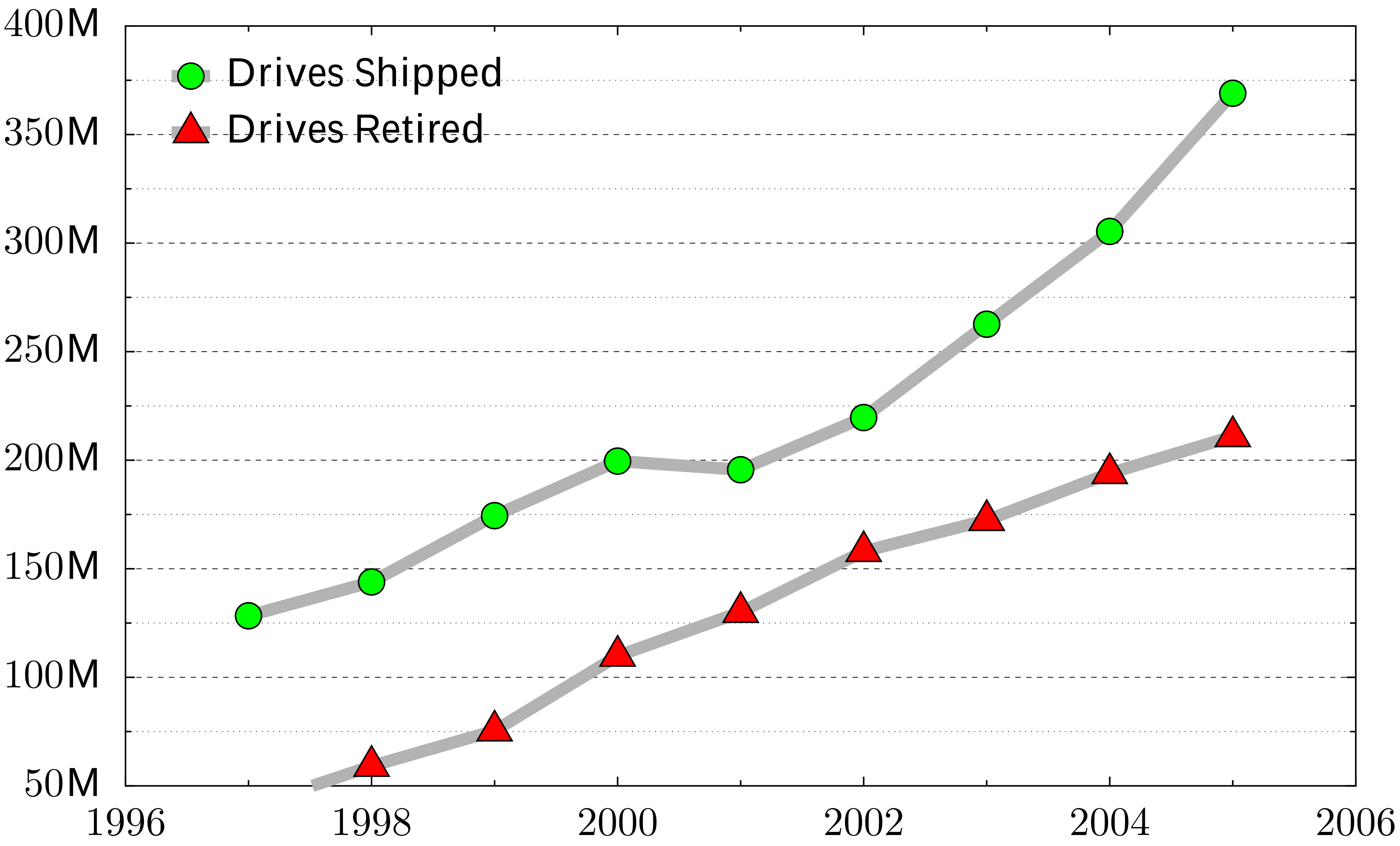
Contain data that is not immediately visible.

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

WESTERN DIGITAL
Caviar 22500
AT Compatible Intelligent Drive
Drive Parameters: 4960 cyl • 16 heads • 63 spt • 2559.8 MB
MDL: AC22500-71LA
CCC: A3 23 AUG 97
DCM: BMAAECLOBBRA
Made in: SINGAPORE
193
5VDC 0.54A
12VDC 0.27A
HEWLETT PACKARD
Replacement P/N: D2686-63001
Exchange P/N: D2686-69001
Manufactured to Hewlett-Packard Standards
2500 MB AT Drive AC22500-71LA
HP P/N: D2686-60101
A QA 15
S/N: WT349 102 9543

Scale of the problem: huge!



210 million drives will be retired this year.

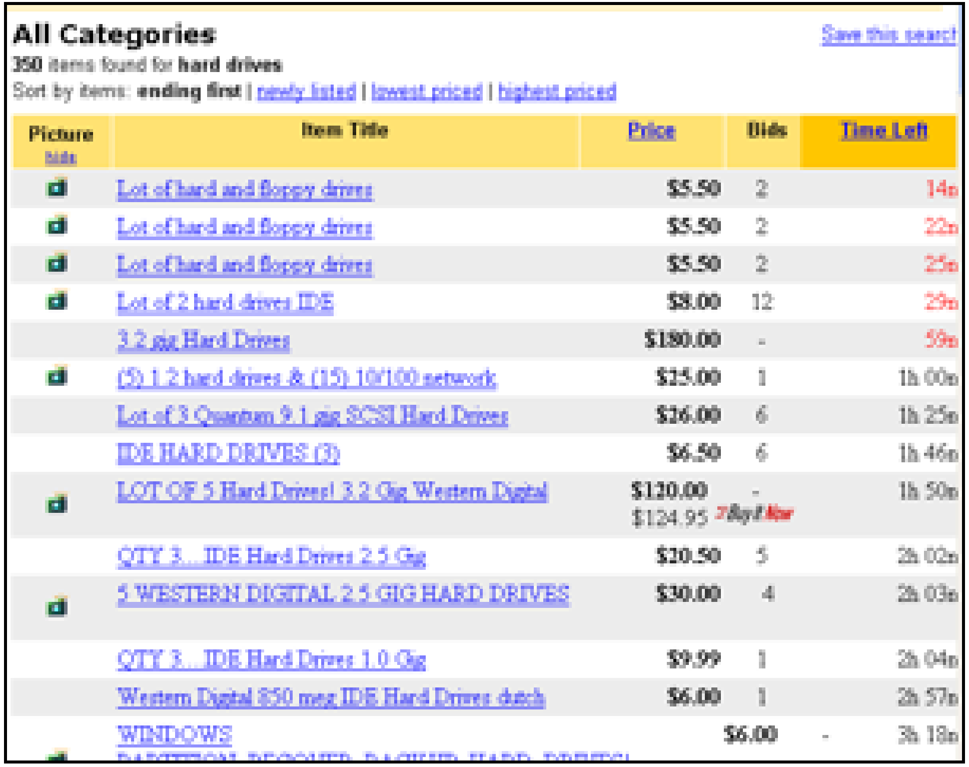
Physical destruction will remove the information...

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

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 hide	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	$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 Buy It Now	-	1h 50m
	QTY 3....IDE Hard Drives 2.5 Gig	$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 Gig	$9.99	1	2h 04m
	Western Digital 850 meg IDE Hard Drives dutch	$6.00	1	2h 57m
📷	WINDOWS PARTITION, RECOVER, BACKUP HARD DRIVES!	$6.00	-	3h 18m

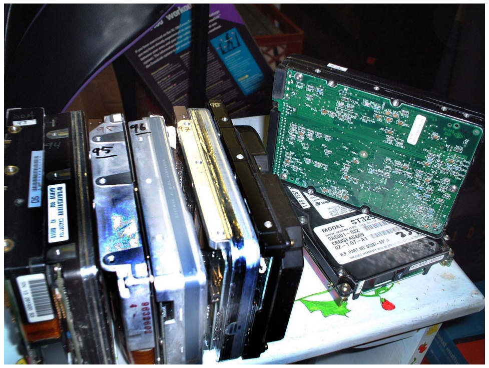
About 1000 used drives/day sold on eBay.

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.

I purchase hard drives on the secondary market.



2001: 100 drives

2003: 150 drives

2005: 500 drives

2006: 1200 drives

Drives arrive by UPS and USPS

Some drives are purchased in person

10GB drive: $19 “tested”

500 MB drive: $3 “as is”

Q: “How do you sanitize them?”

A: “We FDISK them!”

Weird Stuff, Sunnyvale California, January 1999

Drives “imaged” using FreeBSD and AImage

Images stored on DIY RAID.
(Moving to Amazon S3)

I am not considering exotic recovery techniques.

I assume that writing a sector destroys its previous contents.

Some people claim that secret government agencies with advanced technology can recover overwritten data.

This technology has never been publicly demonstrated.

Even without the Men In Black, a lot of data can be recovered!

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

Purchased for $5 from a Mass retail store on eBay

Copied the data off: 541MB

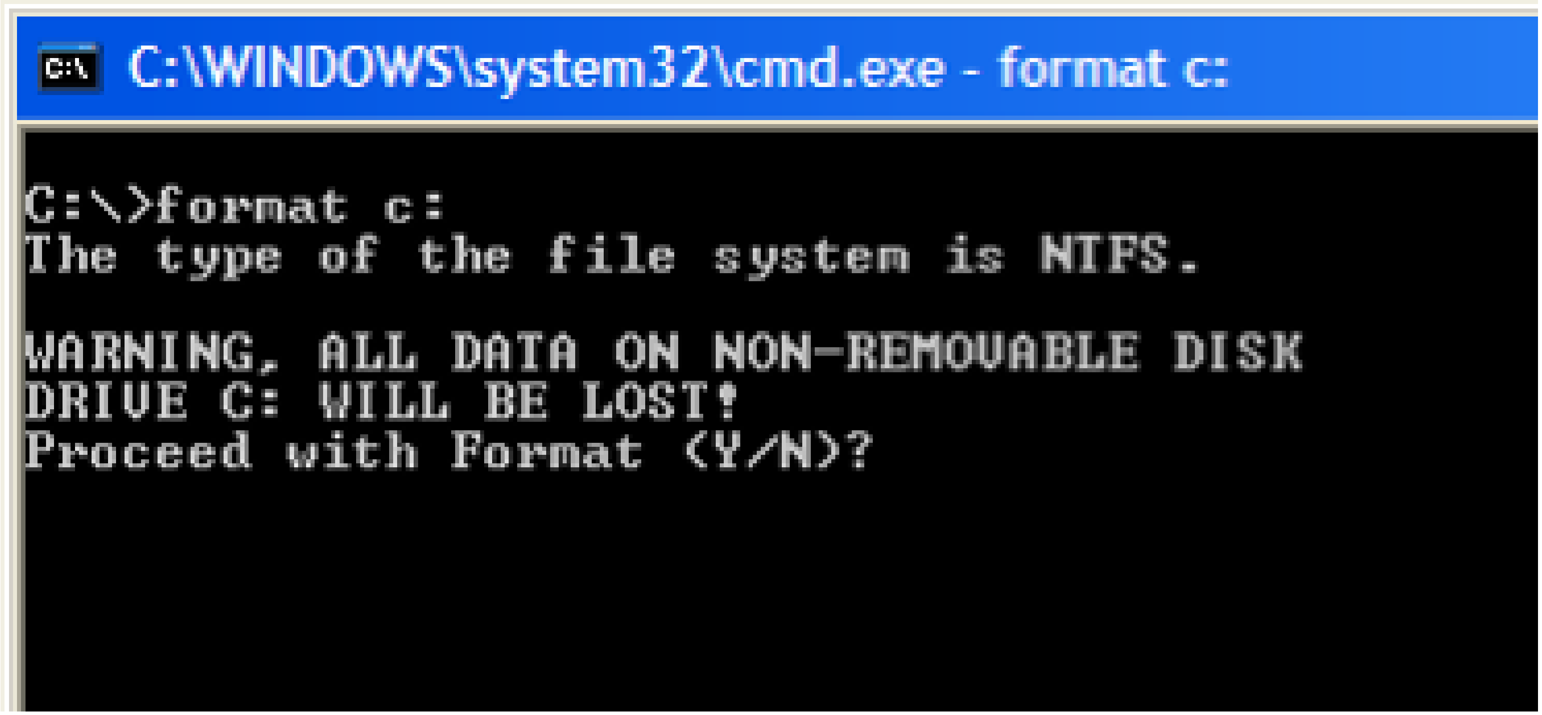
Initial analysis:

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 FORMATed...



```
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 and FDISK overwrite very few disk sectors.

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.**

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

```
% strings 70.img | more
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
```

% strings 70.img

```
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
```

% strings 70.img

```
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
- Financial records
- Medical information
- Trade secrets
- Highly personal information

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

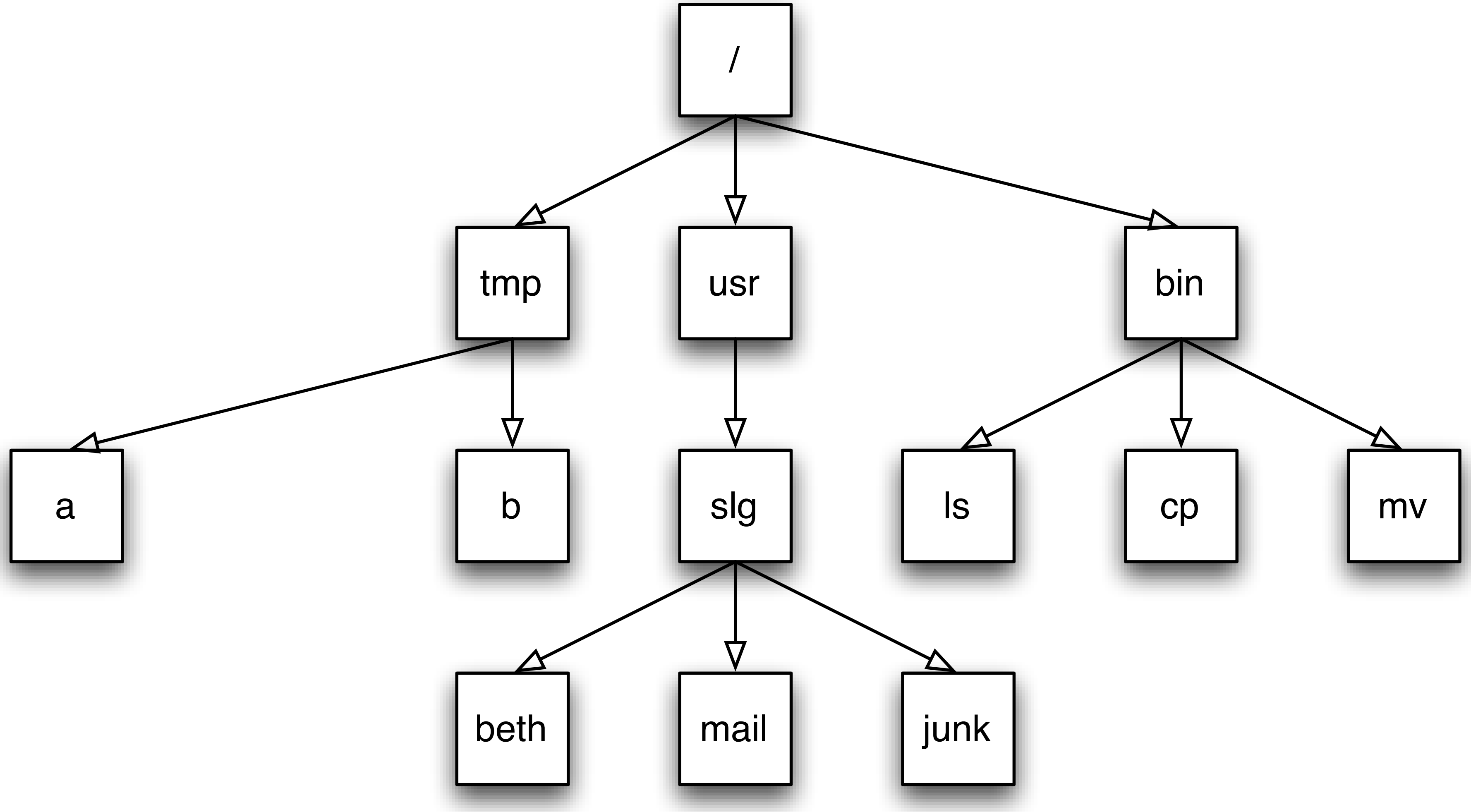
Why don’t we hear more stories?

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.

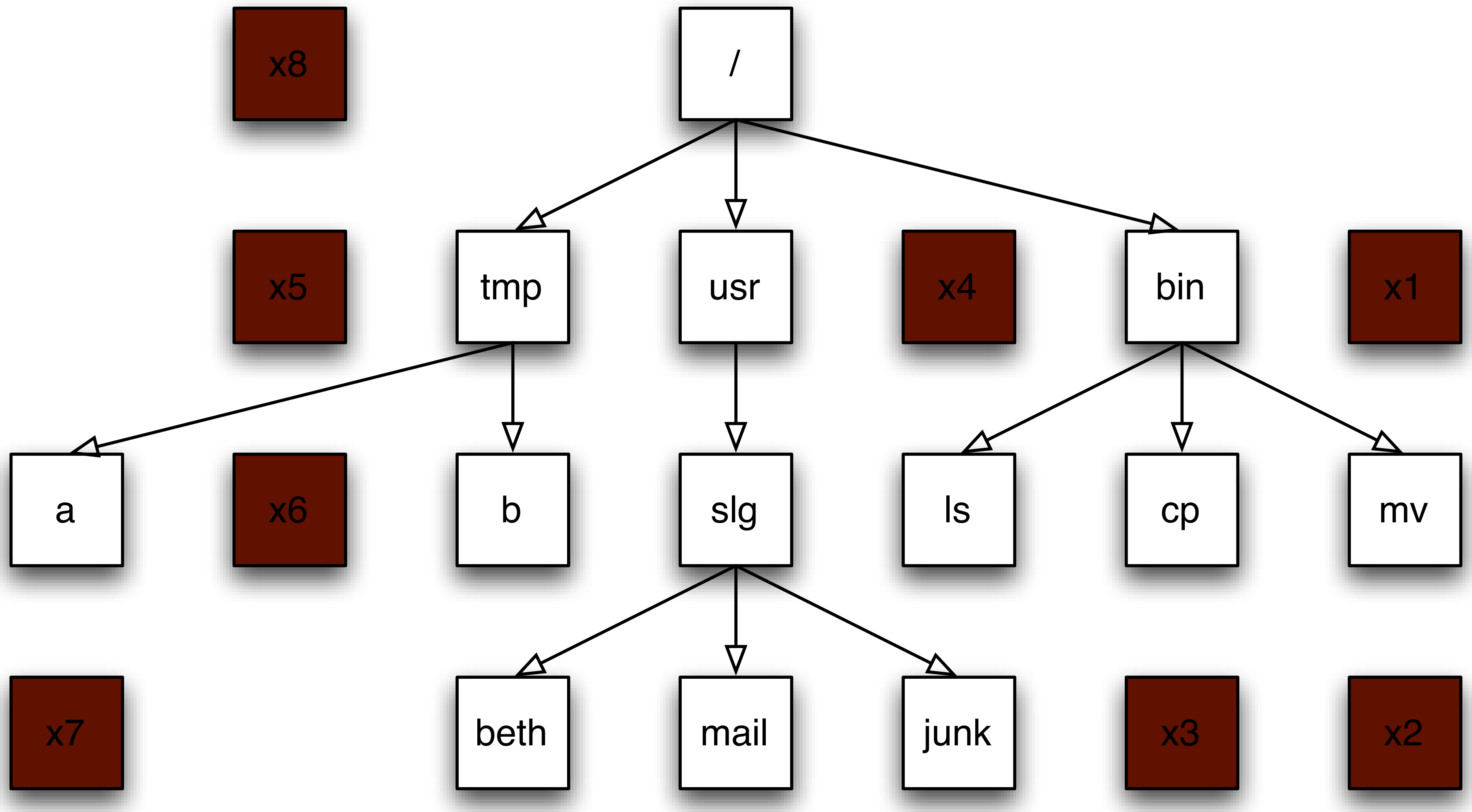
Hypothesis #3: Systems aren’t properly cleared, but few people notice the data.

Data on a hard drive is arranged in sectors.



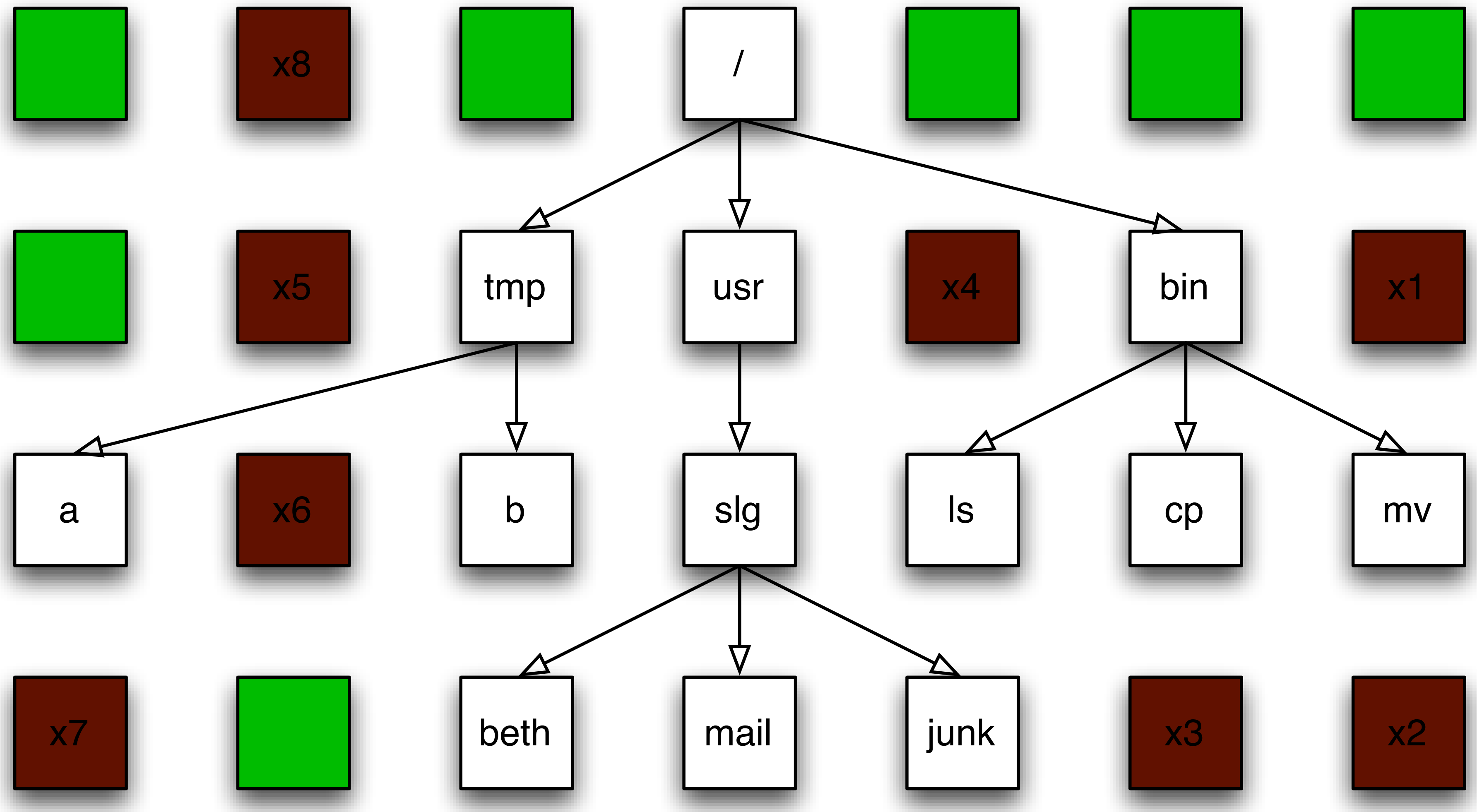
The white sectors indicate directories and files that are visible to the user.

Data on a hard drive is arranged in sectors.



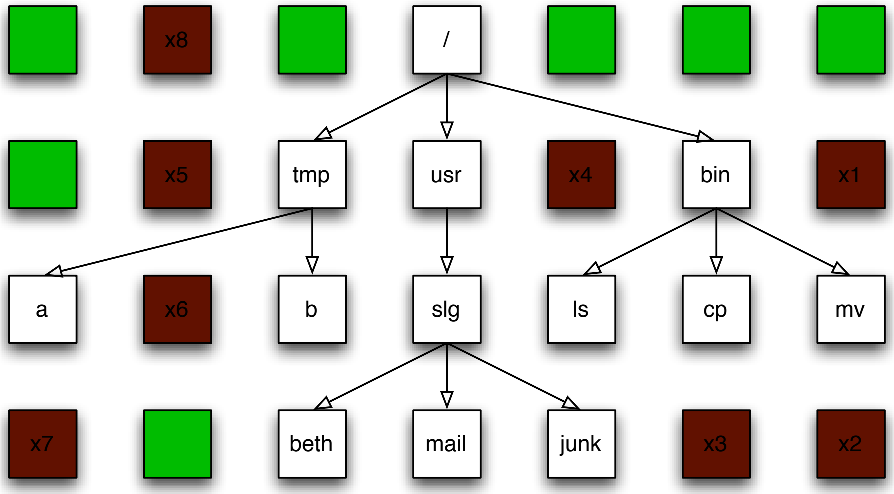
The brown sectors indicate files that were deleted.

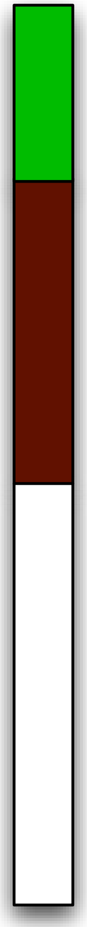
Data on a hard drive is arranged in sectors.



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

Stack the disk sectors:



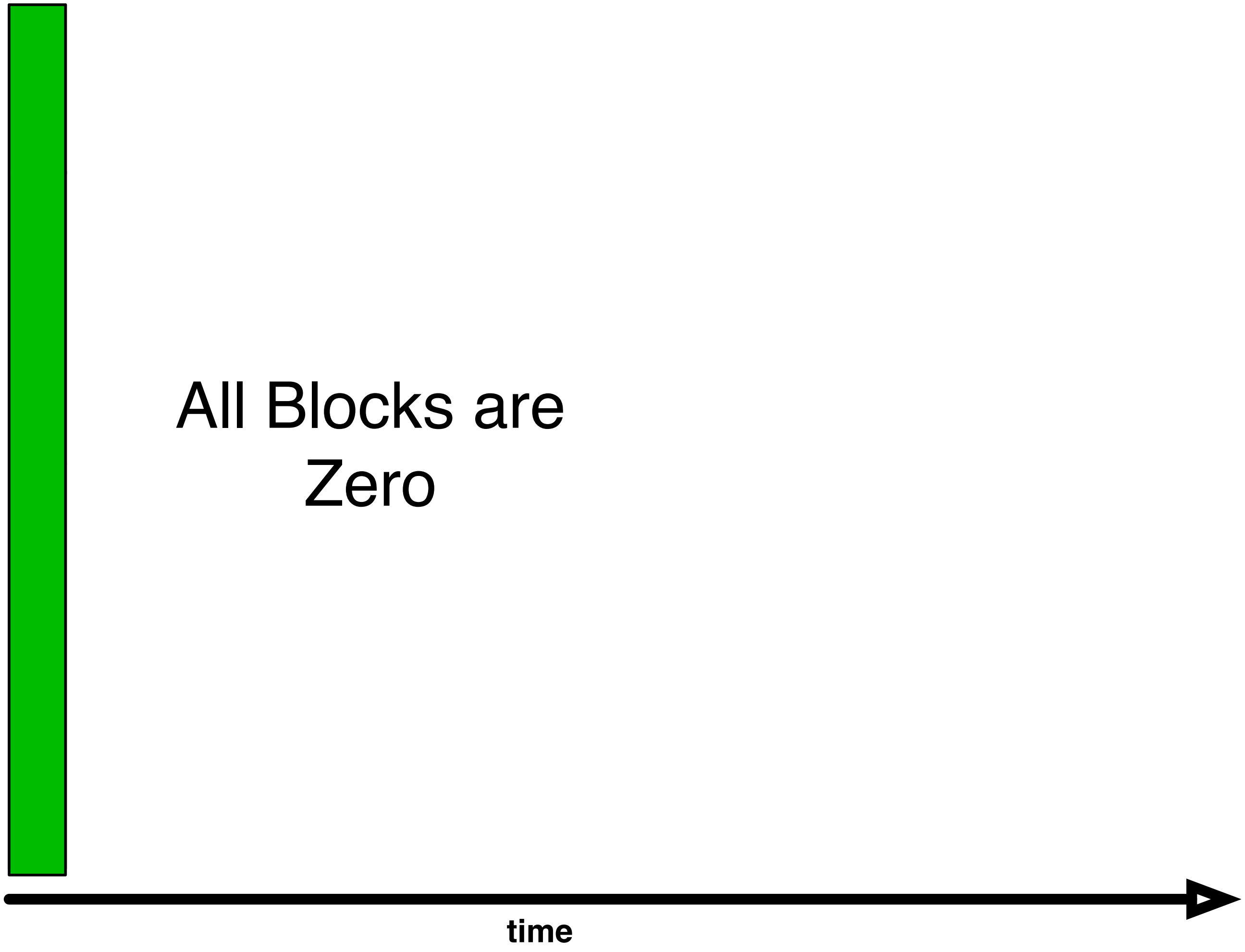


Zero Blocks

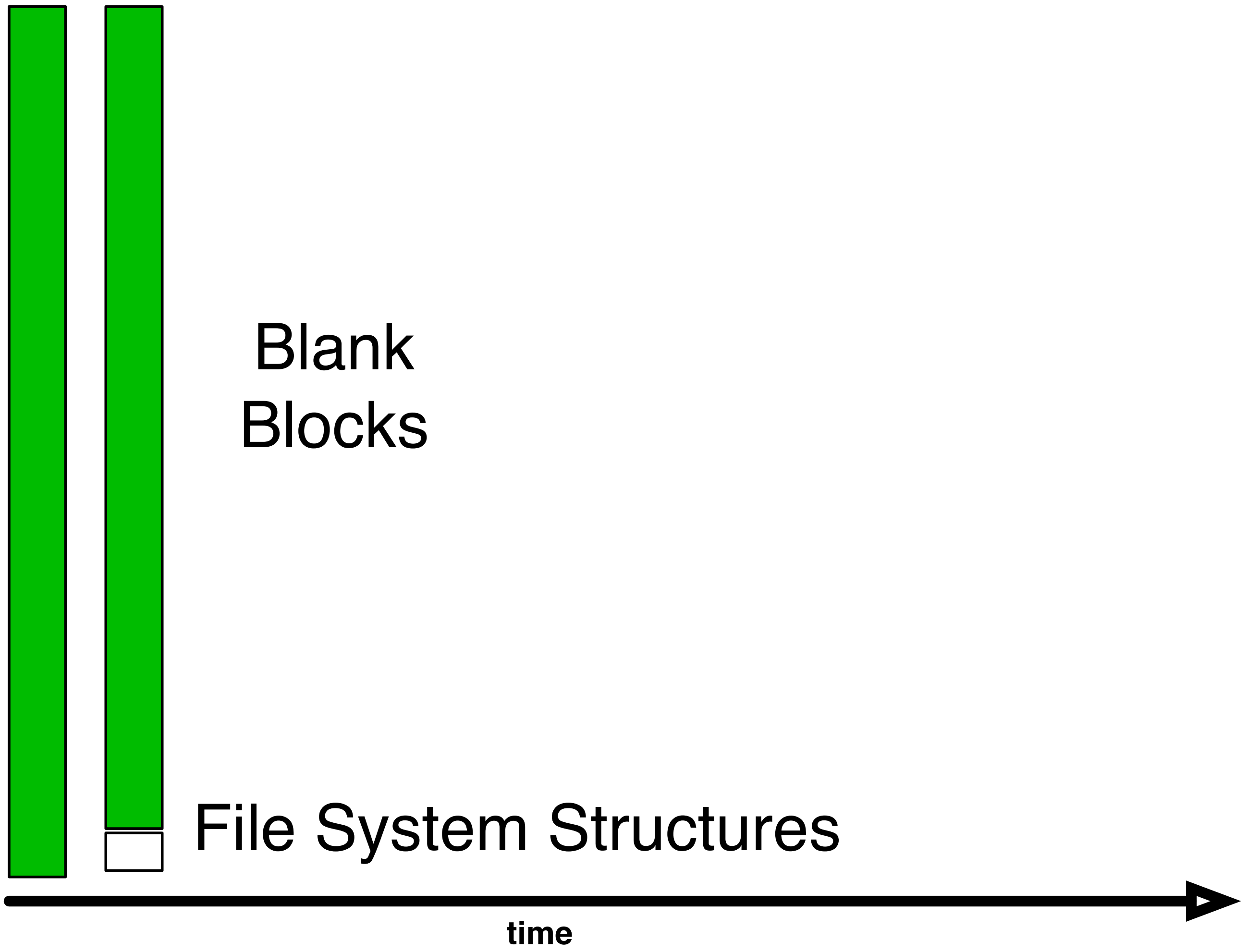
Deleted Files

Files

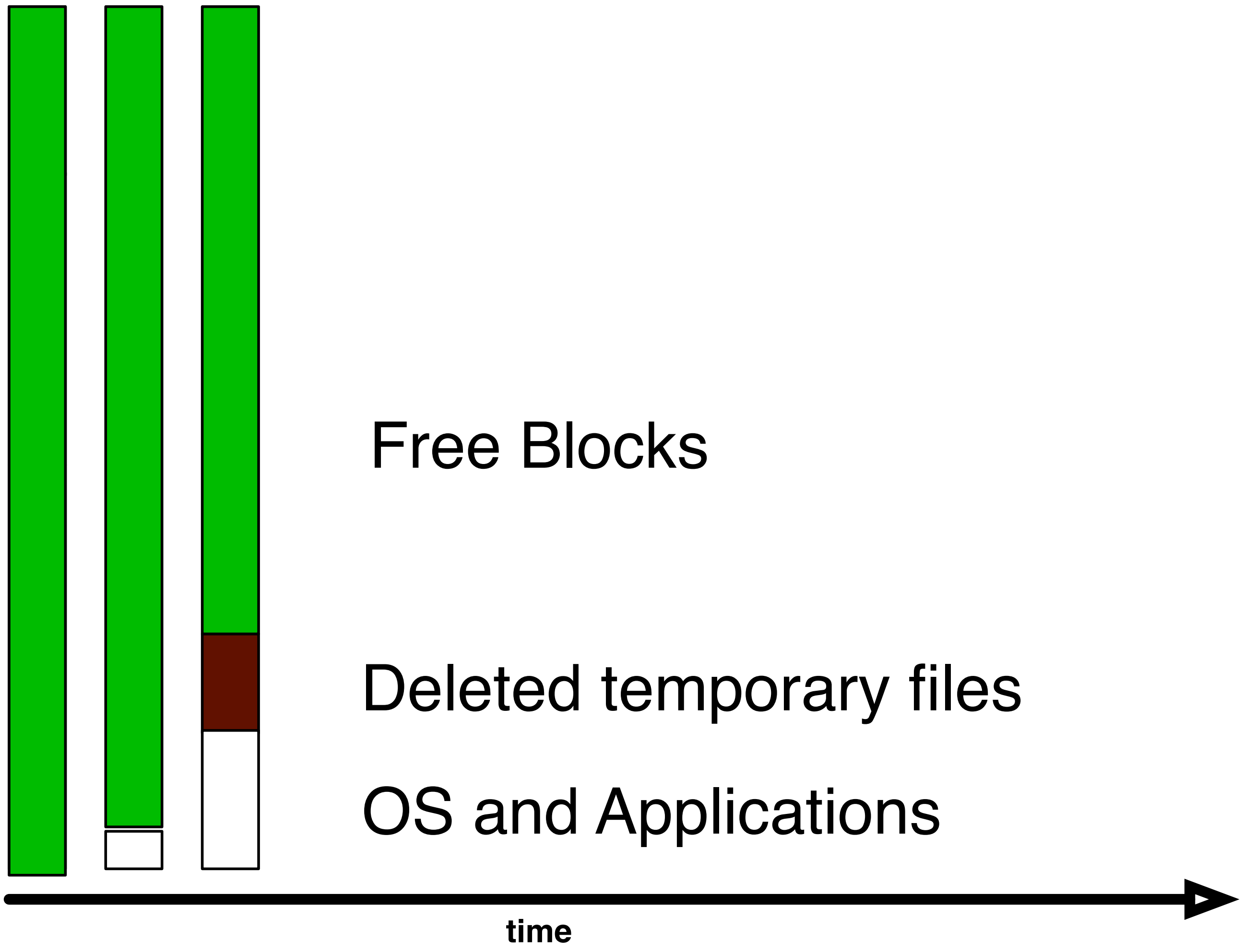
NO DATA: The disk is factory fresh.



FORMATTED: The disk has an empty file system



AFTER OS INSTALL: Temp. files have been deleted



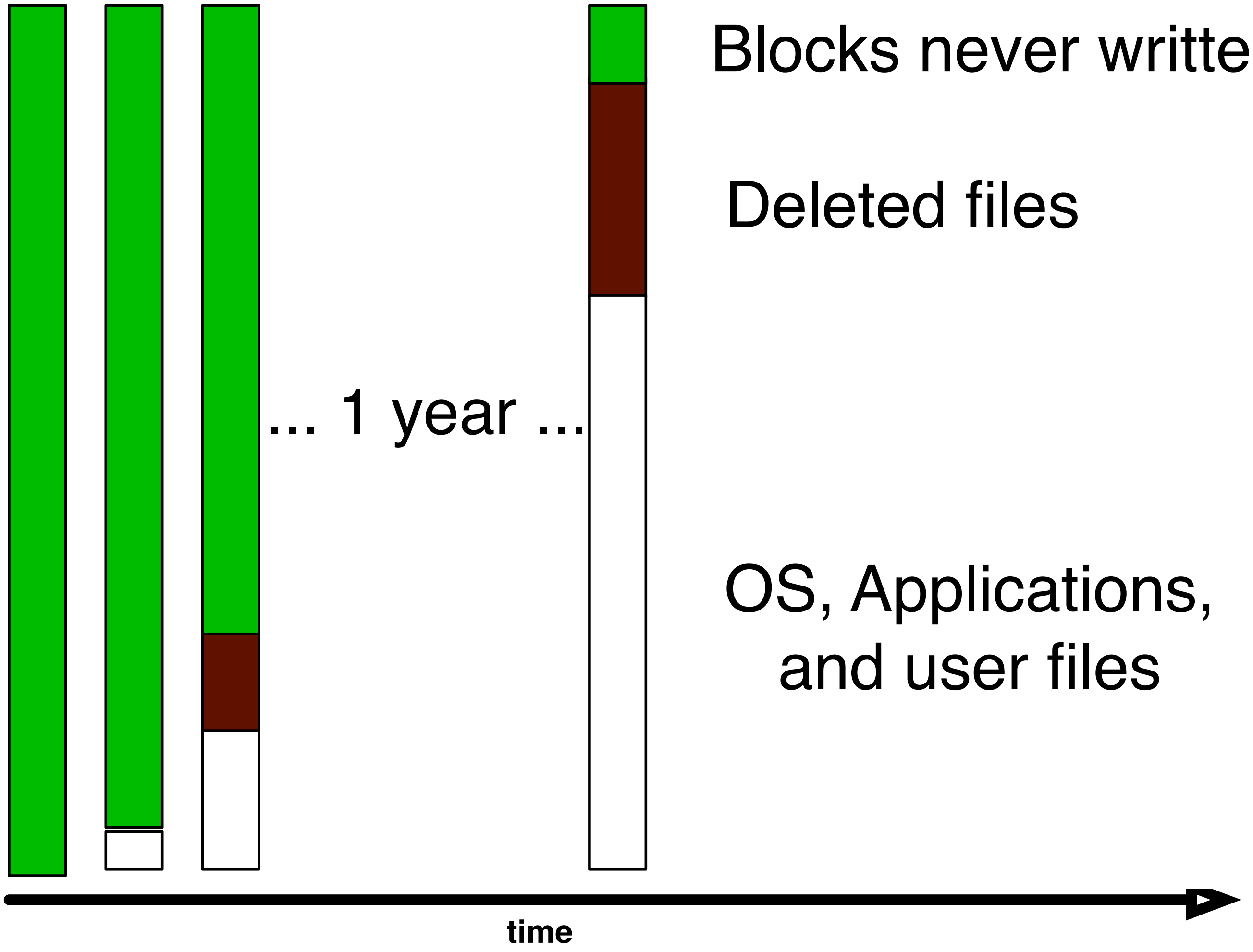
Free Blocks

Deleted temporary files

OS and Applications

time

AFTER A YEAR OF SERVICE

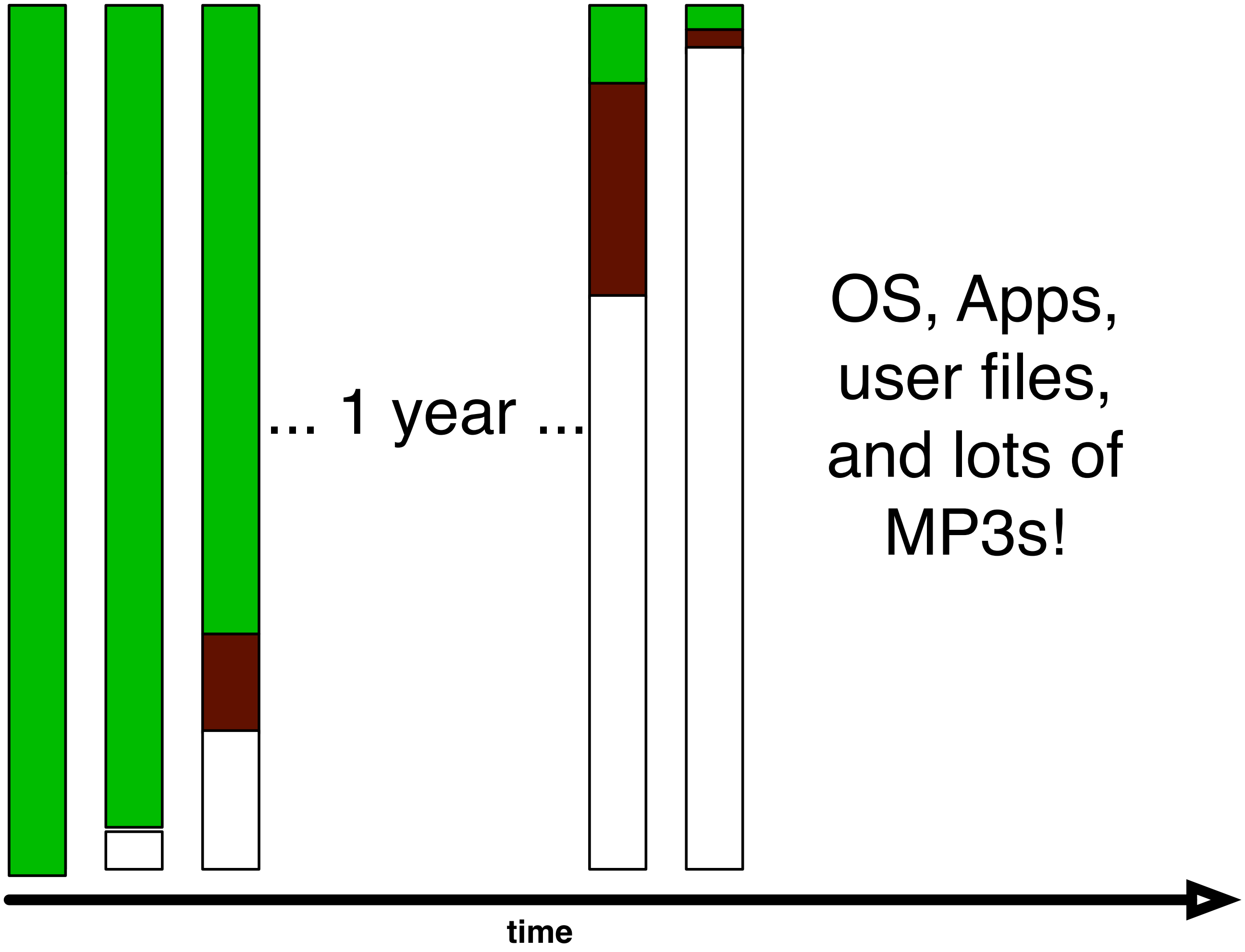


Blocks never written

Deleted files

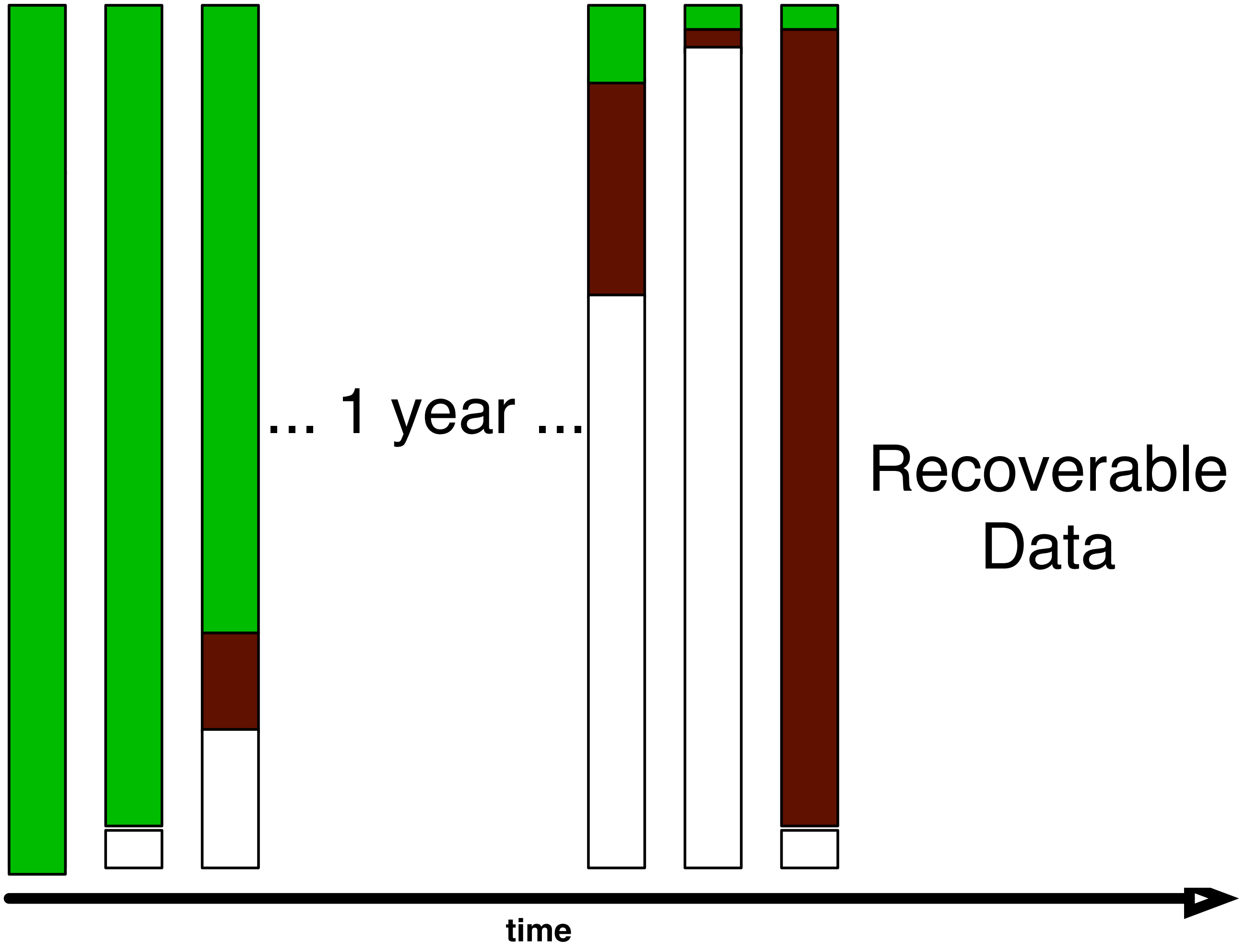
OS, Applications, and user files

DISK NEARLY FULL!

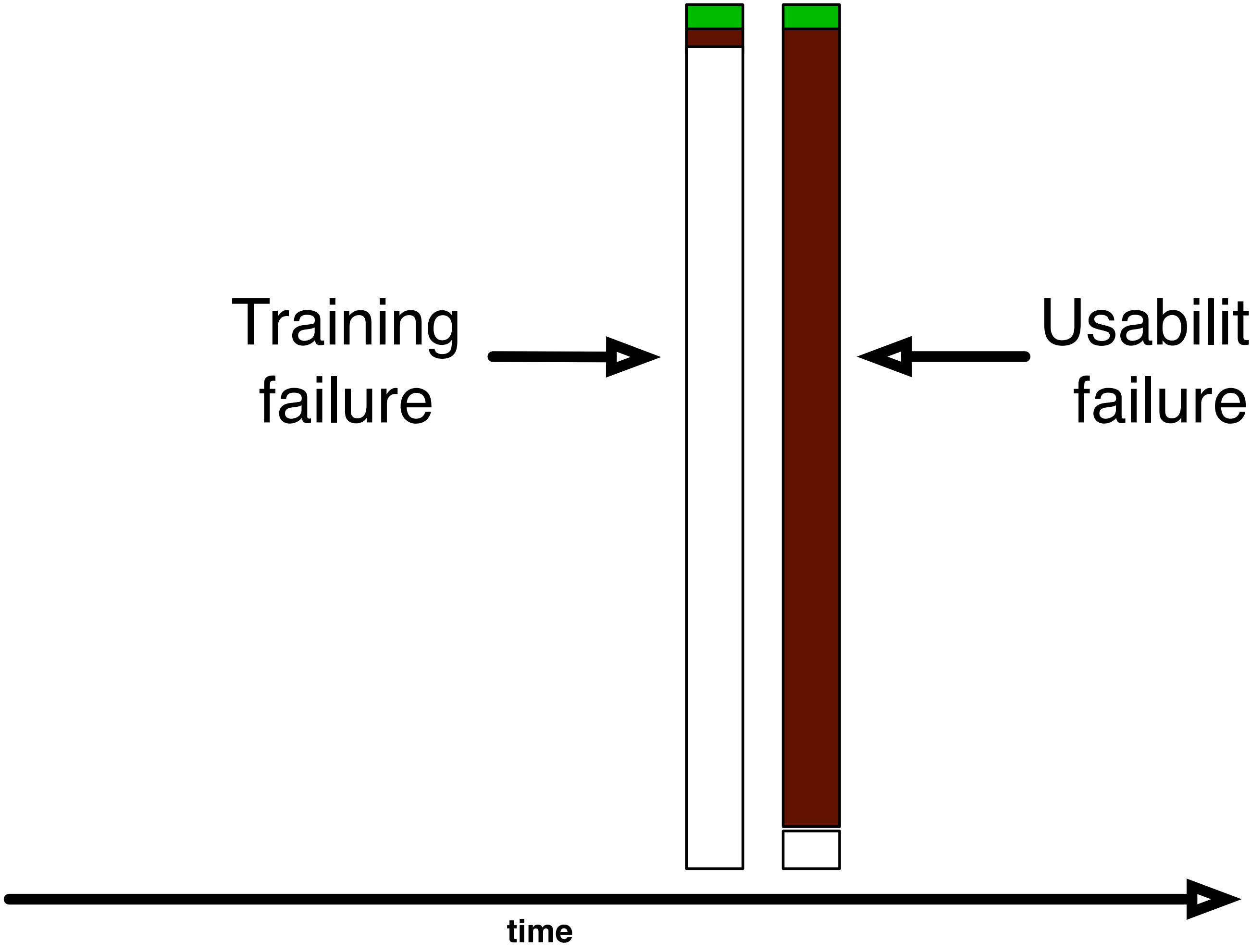


OS, Apps, user files, and lots of MP3s!

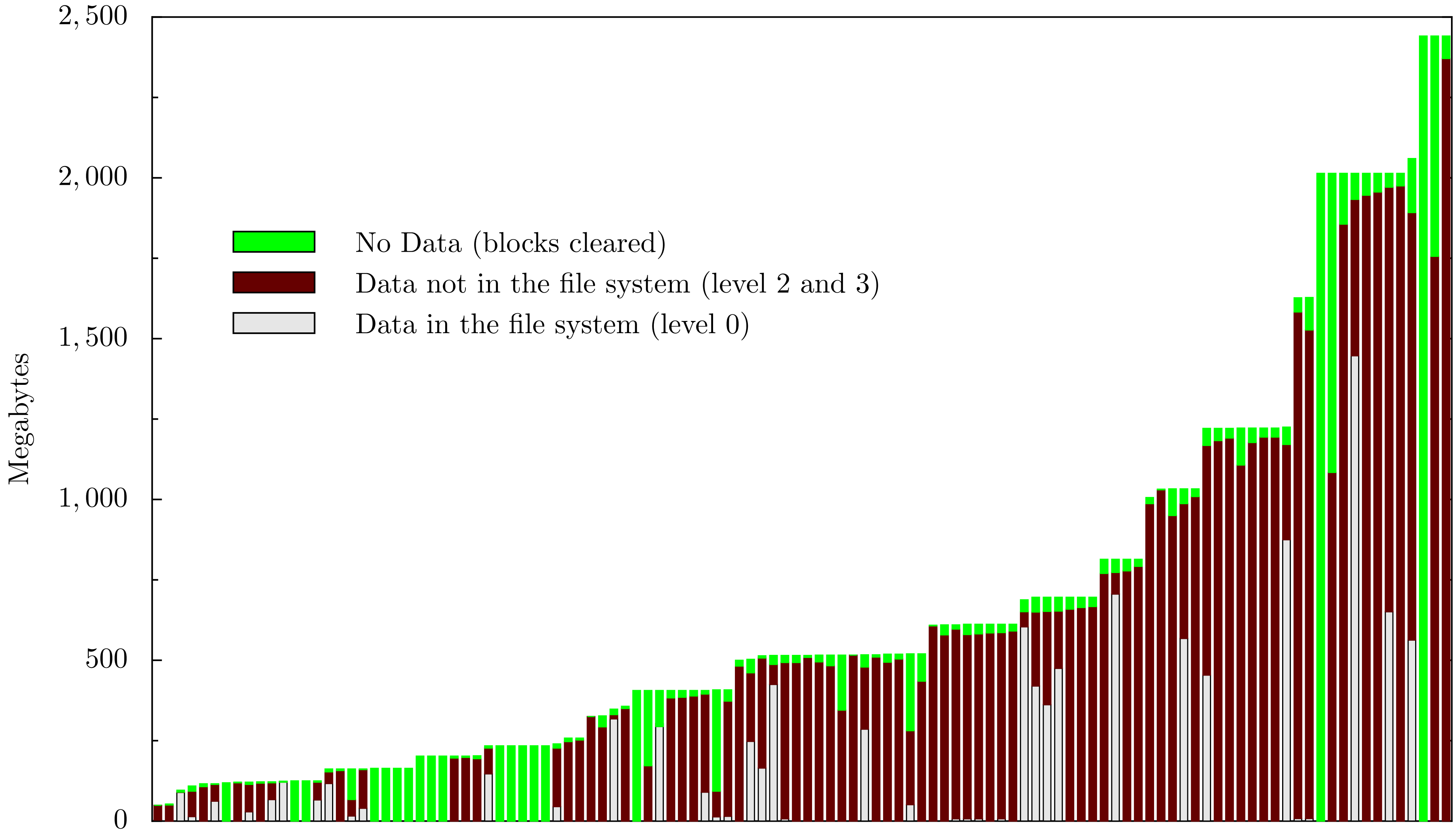
FORMAT C:\ (to sell the computer.)



We can use forensics to reconstruct motivations:



Drives 1–236 are dominated by failed sanitization attempts.



..but training failures are also important.

Overall numbers for the June 2005 report:

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.

There is no consistency on which organizations deliver cleared drives.

But what *really* happened?

?

I needed to contact the original drive owners.

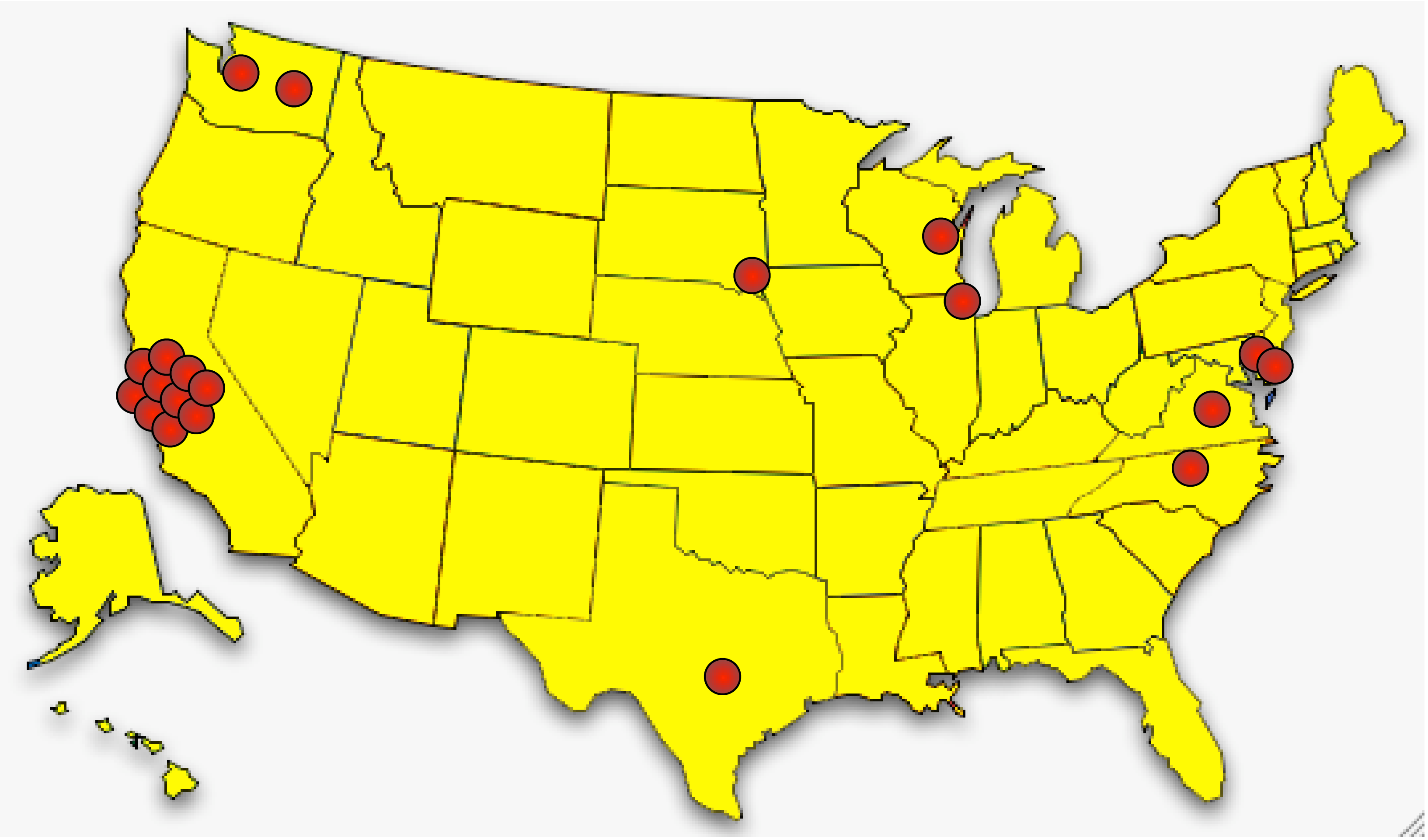
The *Remembrance of Data Passed Traceback Study.* [Garfinkel 05]

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

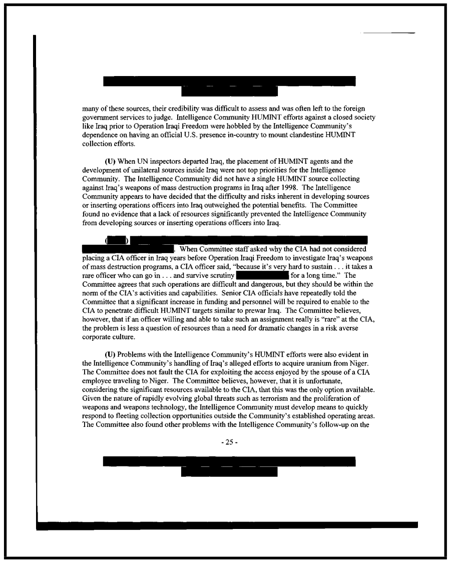
Unknown Reason: 7 cases

- ✘ 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.

“Deleted” data can be recovered in other areas

Document Files



[redacted]

many of these sources, their credibility was difficult to assess and was often left to the foreign government services to judge. Intelligence Community HUMINT efforts against a closed society like Iraq prior to Operation Iraqi Freedom were hobbled by the Intelligence Community's dependence on having an official U.S. presence in-country to mount clandestine HUMINT collection efforts.

(U) When UN inspectors departed Iraq, the placement of HUMINT agents and the development of unilateral sources inside Iraq were not top priorities for the Intelligence Community. The Intelligence Community did not have a single HUMINT source collecting against Iraq's weapons of mass destruction programs in Iraq after 1998. The Intelligence Community appears to have decided that the difficulty and risks inherent in developing sources or inserting operations officers into Iraq outweighed the potential benefits. The Committee found no evidence that a lack of resources significantly prevented the Intelligence Community from developing sources or inserting operations officers into Iraq.

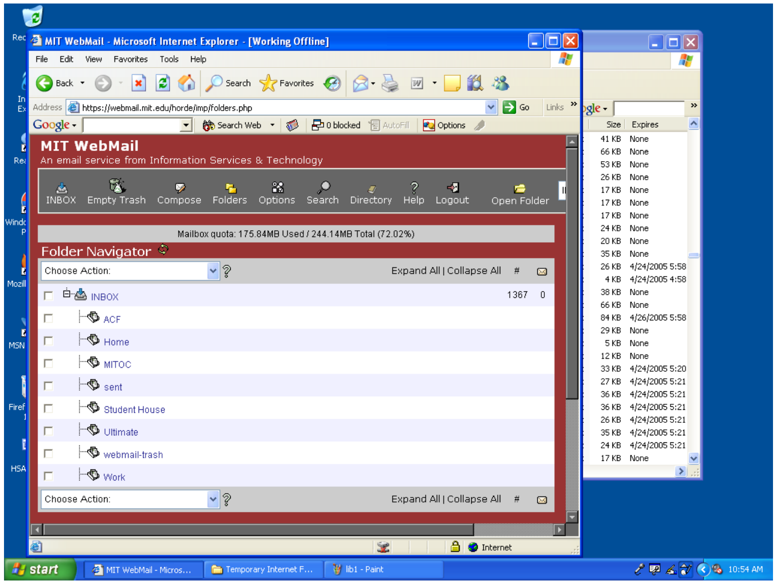
([redacted]) [redacted]. When Committee staff asked why the CIA had not considered placing a CIA officer in Iraq years before Operation Iraqi Freedom to investigate Iraq's weapons of mass destruction programs, a CIA officer said, “because it's very hard to sustain . . . it takes a rare officer who can go in . . . and survive scrutiny [redacted] for a long time.” The Committee agrees that such operations are difficult and dangerous, but they should be within the norm of the CIA's activities and capabilities. Senior CIA officials have repeatedly told the Committee that a significant increase in funding and personnel will be required to enable to the CIA to penetrate difficult HUMINT targets similar to prewar Iraq. The Committee believes, however, that if an officer willing and able to take such an assignment really is “rare” at the CIA, the problem is less a question of resources than a need for dramatic changes in a risk averse corporate culture.

(U) Problems with the Intelligence Community's HUMINT efforts were also evident in the Intelligence Community's handling of Iraq's alleged efforts to acquire uranium from Niger. The Committee does not fault the CIA for exploiting the access enjoyed by the spouse of a CIA employee traveling to Niger. The Committee believes, however, that it is unfortunate, considering the significant resources available to the CIA, that this was the only option available. Given the nature of rapidly evolving global threats such as terrorism and the proliferation of weapons and weapons technology, the Intelligence Community must develop means to quickly respond to fleeting collection opportunities outside the Community's established operating areas. The Committee also found other problems with the Intelligence Community's follow-up on the

- 25 -

[redacted]

Web Browsers



https://webmail.mit.edu/horde/imp/folders.php

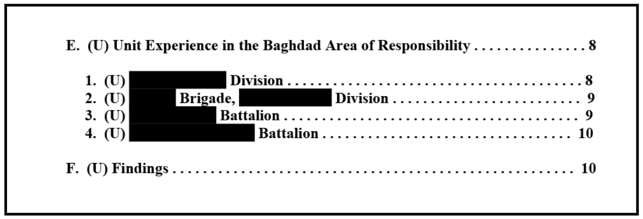
Mailbox quota: 175.84MB Used / 244.14MB Total (72.02%)

	#	
□ INBOX	1367	0
□ ACF		
□ Home		
□ MITOC		
□ sent		
□ Student House		
□ Ultimate		
□ webmail-trash		
□ Work		

Size	Expires
41 KB	None
66 KB	None
53 KB	None
26 KB	None
17 KB	None
17 KB	None
17 KB	None
24 KB	None
20 KB	None
35 KB	None
26 KB	4/24/2005 5:58
4 KB	4/24/2005 4:58
38 KB	None
66 KB	None
84 KB	4/26/2005 5:58
29 KB	None
5 KB	None
12 KB	None
33 KB	4/24/2005 5:20
27 KB	4/24/2005 5:21
36 KB	4/24/2005 5:21
36 KB	4/24/2005 5:21
26 KB	4/24/2005 5:21
35 KB	4/24/2005 5:21
24 KB	4/24/2005 5:21
17 KB	None

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

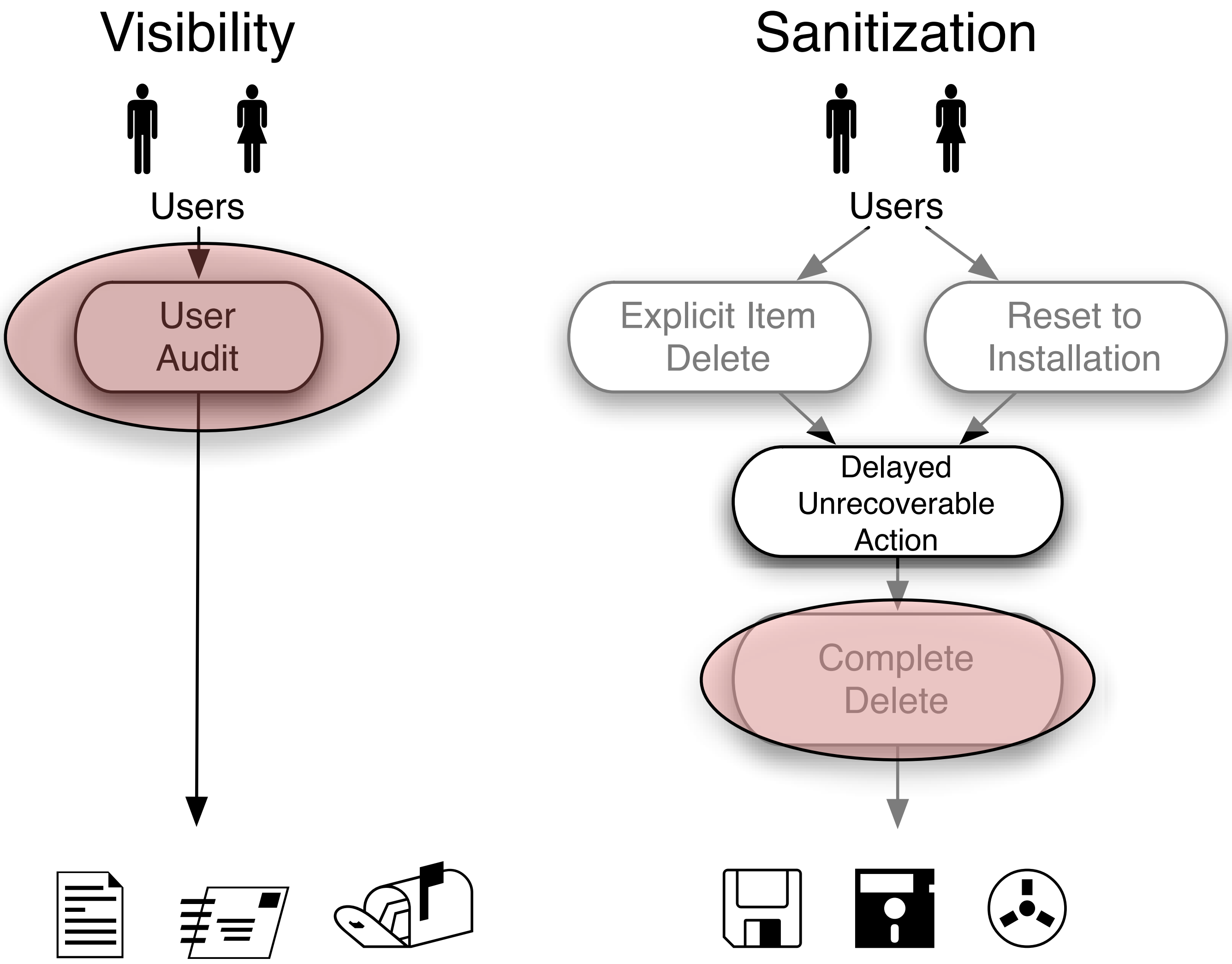


E. (U) Unit Experience in the Baghdad Area of Responsibility 8

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

The information leaked because two patterns were not implemented.



Visibility

Users

User Audit

Sanitization

Users

Explicit Item Delete

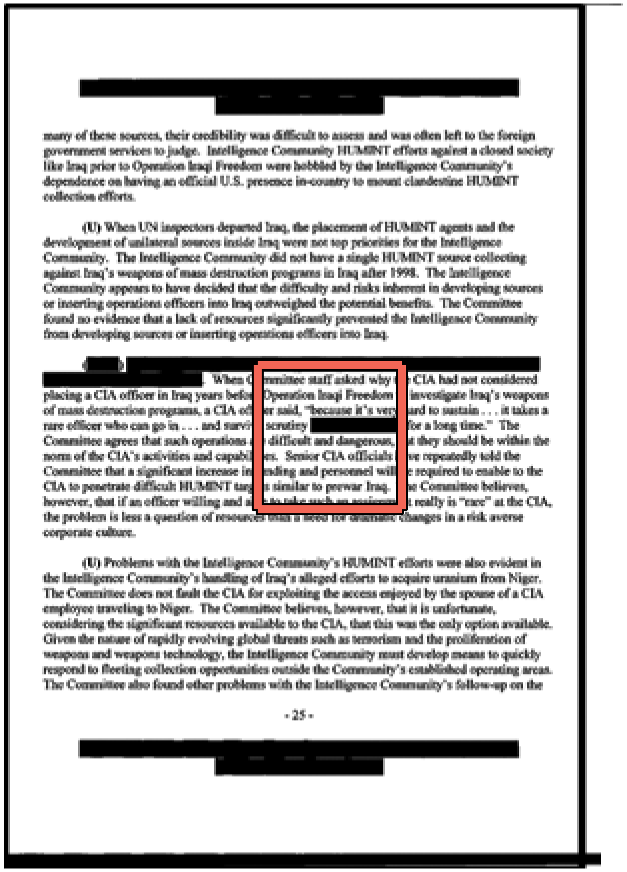
Reset to Installation

Delayed Unrecoverable Action

Complete Delete

Document Files, Applications, and Media

The Senate Foreign Intelligence Committee prevented leakage by *scanning* its redacted report on pre-war Iraq intelligence failures to create the PDF that it distributed.



[redacted]

many of these sources, their credibility was difficult to assess and was often left to the foreign government services to judge. Intelligence Community HUMINT efforts against a closed society like Iraq prior to Operation Iraqi Freedom were hobbled by the Intelligence Community's dependence on having an official U.S. presence in-country to mount clandestine HUMINT collection efforts.

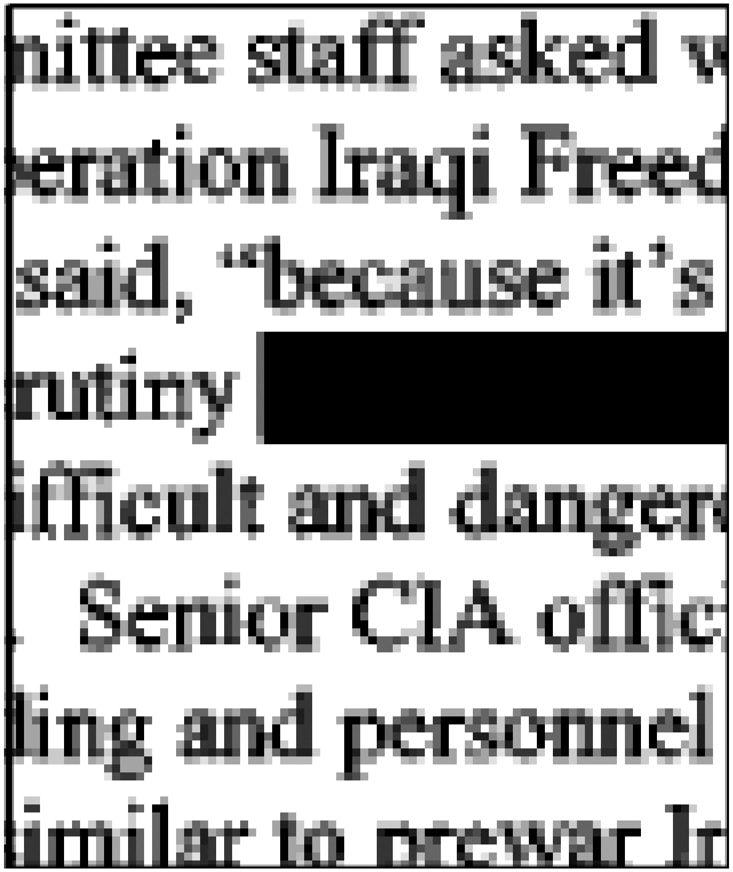
(U) When UN inspectors departed Iraq, the placement of HUMINT agents and the development of unilateral sources inside Iraq were not top priorities for the Intelligence Community. The Intelligence Community did not have a single HUMINT source collecting against Iraq's weapons of mass destruction programs in Iraq after 1998. The Intelligence Community appears to have decided that the difficulty and risks inherent in developing sources or inserting operations officers into Iraq outweighed the potential benefits. The Committee found no evidence that a lack of resources significantly prevented the Intelligence Community from developing sources or inserting operations officers into Iraq.

[redacted] When Committee staff asked why the CIA had not considered placing a CIA officer in Iraq years before Operation Iraqi Freedom to investigate Iraq's weapons of mass destruction programs, a CIA officer said, "because it's very hard to sustain . . . it takes a rare officer who can go in . . . and survive scrutiny [redacted] for a long time." The Committee agrees that such operations are difficult and dangerous, but they should be within the norm of the CIA's activities and capabilities. Senior CIA officials have repeatedly told the Committee that a significant increase in funding and personnel will be required to enable to the CIA to penetrate difficult HUMINT targets similar to prewar Iraq. The Committee believes, however, that if an officer willing and able to take such an assignment really is "rare" at the CIA, the problem is less a question of resources than a need for dramatic changes in a risk averse corporate culture.

(U) Problems with the Intelligence Community's HUMINT efforts were also evident in the Intelligence Community's handling of Iraq's alleged efforts to acquire uranium from Niger. The Committee does not fault the CIA for exploiting the access enjoyed by the spouse of a CIA employee traveling to Niger. The Committee believes, however, that it is unfortunate, considering the significant resources available to the CIA, that this was the only option available. Given the nature of rapidly evolving global threats such as terrorism and the proliferation of weapons and weapons technology, the Intelligence Community must develop means to quickly respond to fleeting collection opportunities outside the Community's established operating areas. The Committee also found other problems with the Intelligence Community's follow-up on the

- 25 -

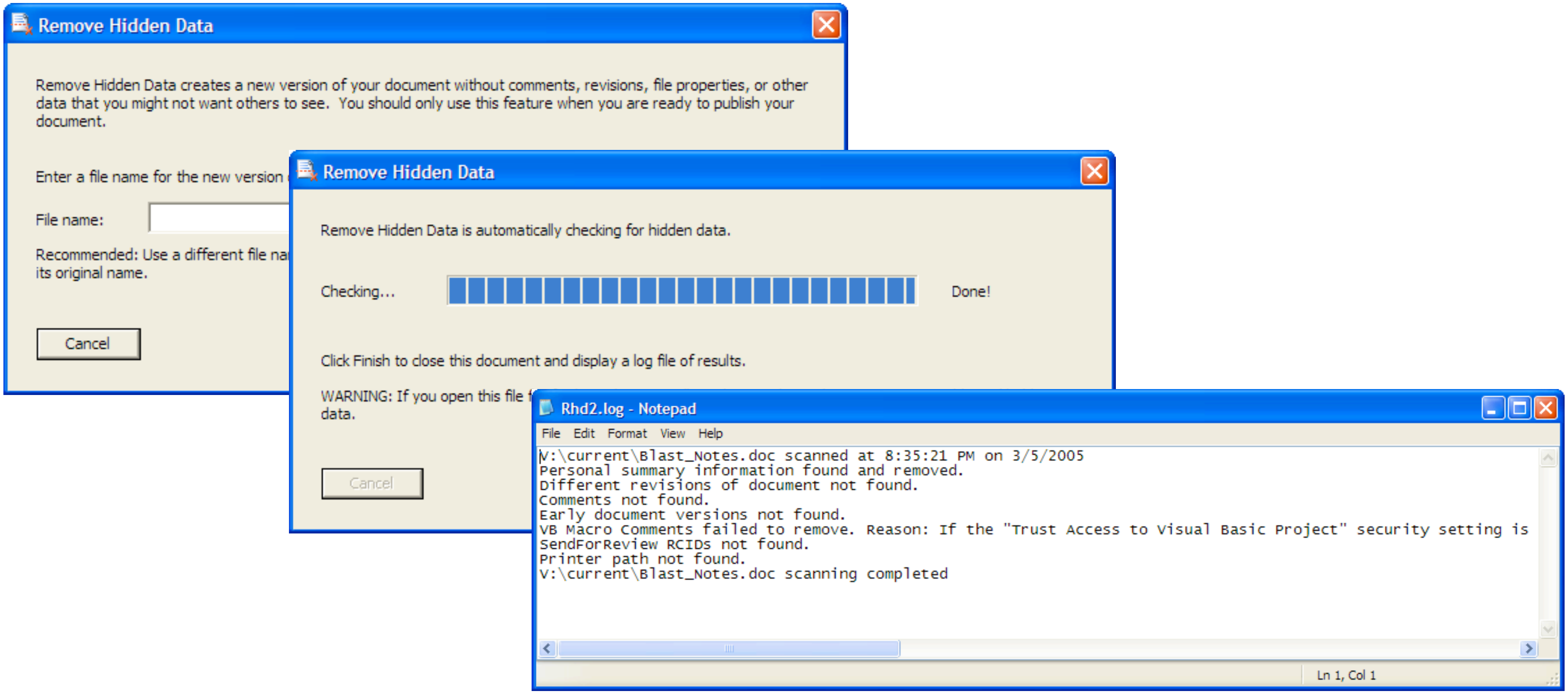
[redacted]



nittee staff asked w
eration Iraqi Freed
said, "because it's
rutiny [redacted]
ifficult and danger
Senior CIA offic
ling and personnel
imilar to prewar Ir

This violates Section 503 (but they don't care).

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



Remove Hidden Data creates a new version of your document without comments, revisions, file properties, or other data that you might not want others to see. You should only use this feature when you are ready to publish your document.

Enter a file name for the new version

File name:

Recommended: Use a different file na its original name.

Remove Hidden Data is automatically checking for hidden data.

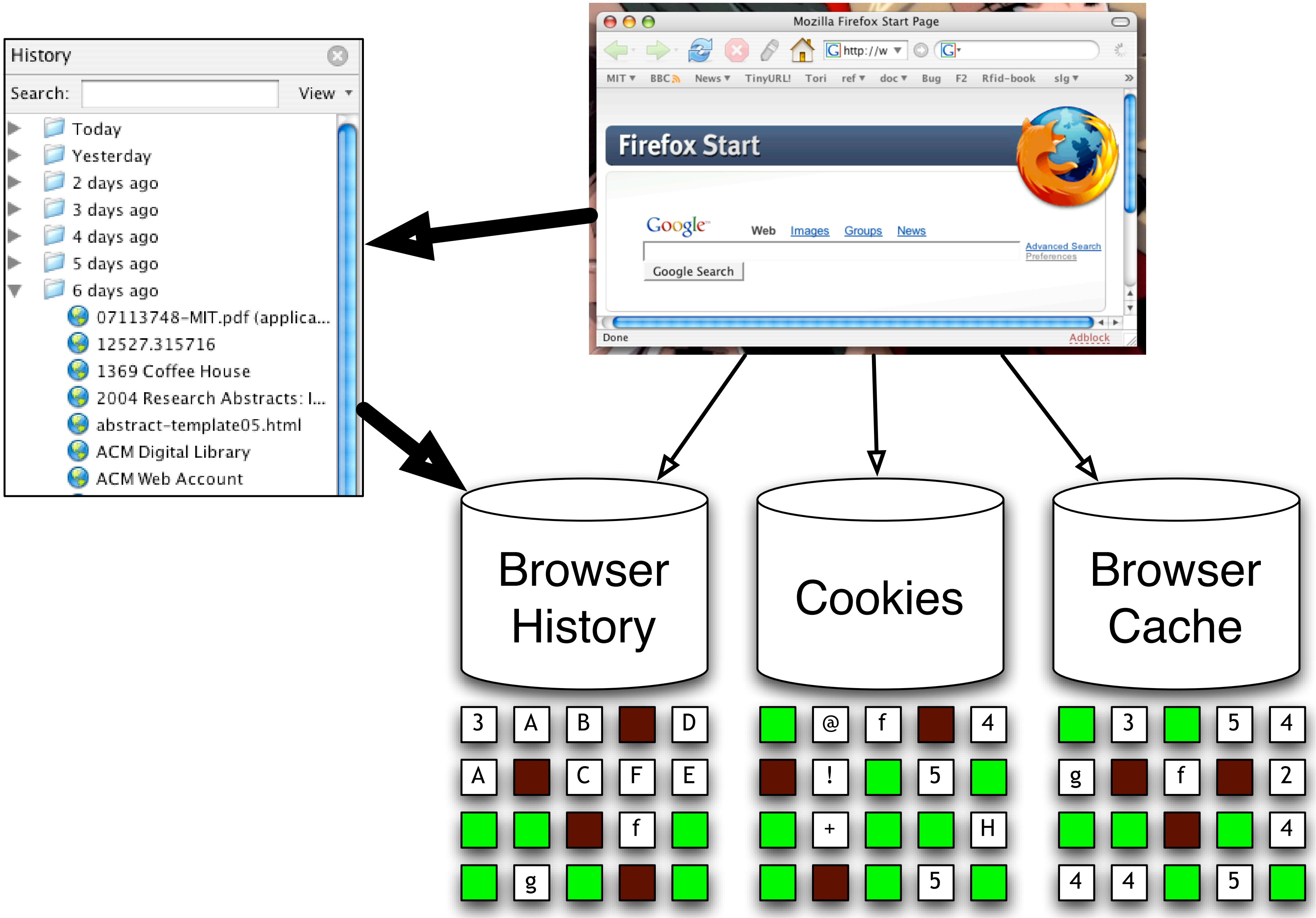
Click Finish to close this document and display a log file of results.

WARNING: If you open this file data.

```
V:\current\Blast_Notes.doc scanned at 8:35:21 PM on 3/5/2005
Personal summary information found and removed.
Different revisions of document not found.
Comments not found.
Early document versions not found.
VB Macro Comments failed to remove. Reason: If the "Trust Access to Visual Basic Project" security setting is
SendForReview RCIDs not found.
Printer path not found.
V:\current\Blast_Notes.doc scanning completed
```

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.



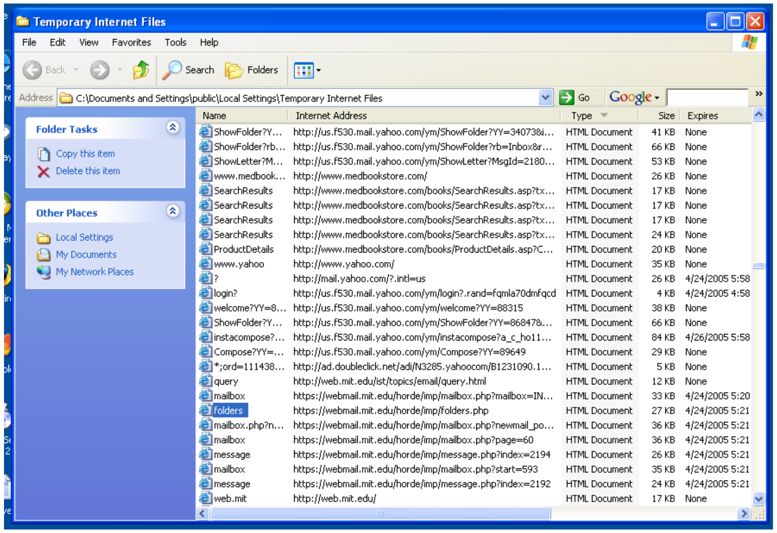
3	A	B	🟫	D
A	🟫	C	F	E
🟩	🟩	🟫	f	🟩
🟩	g	🟩	🟫	🟩

🟩	@	f	🟫	4
🟫	!	🟩	5	🟩
🟩	+	🟩	🟩	H
🟩	🟫	🟩	5	🟩

🟩	3	🟩	5	4
g	🟫	f	🟫	2
🟩	🟩	🟫	🟩	4
4	4	🟩	5	🟩

Two key problems: ① Deleted files; ② The cache

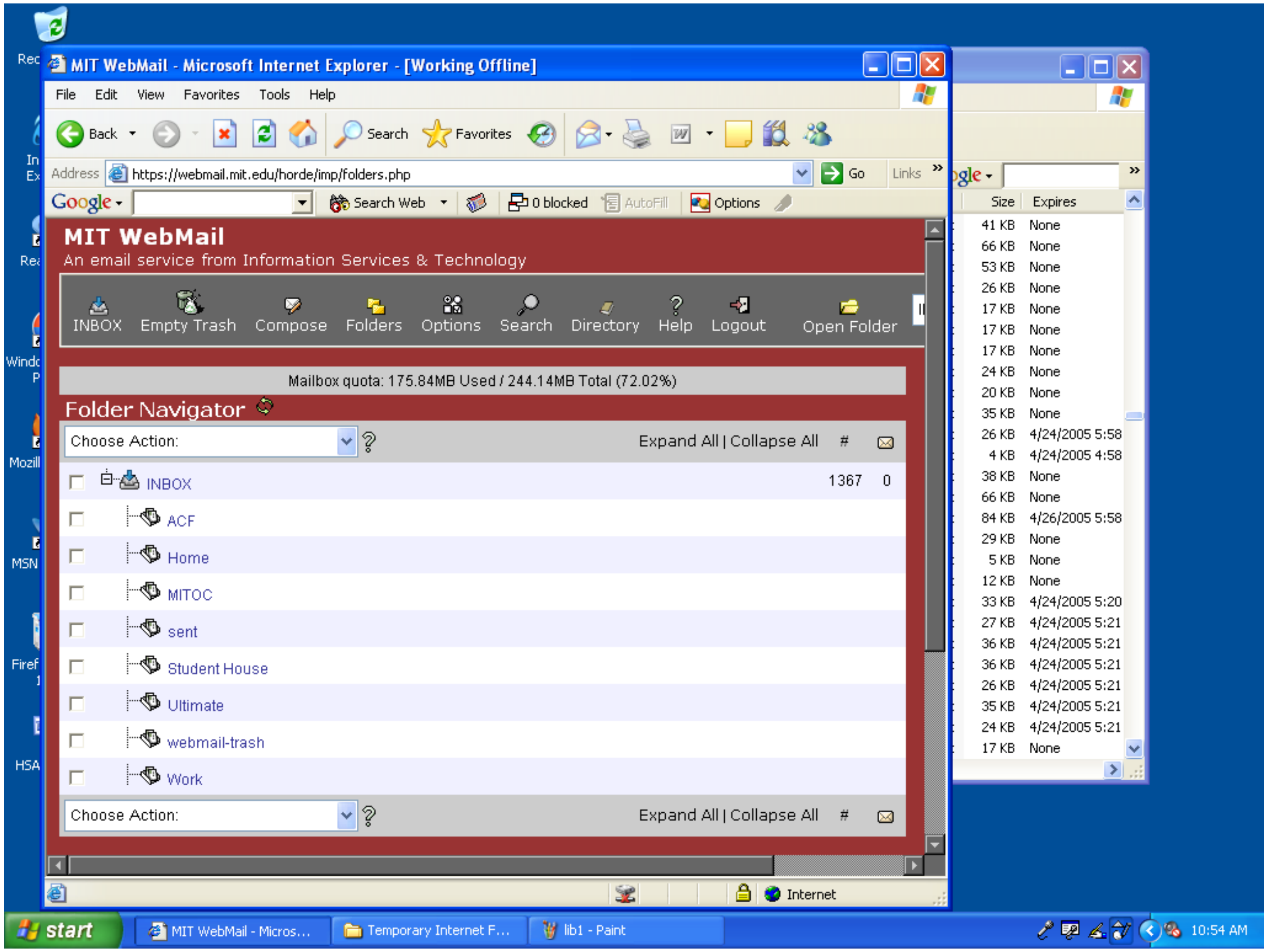
In fact, a lot of information is left behind in web browsers.



Name	Internet Address	Type	Size	Expires
ShowFolder?Y...	http://us.f530.mail.yahoo.com/ym/ShowFolder?YY=34073&i...	HTML Document	41 KB	None
ShowFolder?rb...	http://us.f530.mail.yahoo.com/ym/ShowFolder?rb=Inbox&r...	HTML Document	66 KB	None
ShowLetter?M...	http://us.f530.mail.yahoo.com/ym/ShowLetter?MsgId=2180...	HTML Document	53 KB	None
www.medbook...	http://www.medbookstore.com/	HTML Document	26 KB	None
SearchResults	http://www.medbookstore.com/books/SearchResults.asp?tx...	HTML Document	17 KB	None
SearchResults	http://www.medbookstore.com/books/SearchResults.asp?tx...	HTML Document	17 KB	None
SearchResults	http://www.medbookstore.com/books/SearchResults.asp?tx...	HTML Document	17 KB	None
SearchResults	http://www.medbookstore.com/books/SearchResults.asp?tx...	HTML Document	24 KB	None
ProductDetails	http://www.medbookstore.com/books/ProductDetails.asp?C...	HTML Document	20 KB	None
www.yahoo	http://www.yahoo.com/	HTML Document	35 KB	None
?	http://mail.yahoo.com/?.intl=us	HTML Document	26 KB	4/24/2005 5:58
login?	http://us.f530.mail.yahoo.com/ym/login?.rand=fqmla70dmfqcd	HTML Document	4 KB	4/24/2005 4:58
welcome?YY=8...	http://us.f530.mail.yahoo.com/ym/welcome?YY=88315	HTML Document	38 KB	None
ShowFolder?Y...	http://us.f530.mail.yahoo.com/ym/ShowFolder?YY=86847&...	HTML Document	66 KB	None
instacompose?...	http://us.f530.mail.yahoo.com/ym/instacompose?a_c_ho11...	HTML Document	84 KB	4/26/2005 5:58
Compose?YY=...	http://us.f530.mail.yahoo.com/ym/Compose?YY=89649	HTML Document	29 KB	None
*;ord=111438...	http://ad.doubleclick.net/adi/N3285.yahoocom/B1231090.1...	HTML Document	5 KB	None
query	http://web.mit.edu/ist/topics/email/query.html	HTML Document	12 KB	None
mailbox	https://webmail.mit.edu/horde/imp/mailbox.php?mailbox=IN...	HTML Document	33 KB	4/24/2005 5:20
folders	https://webmail.mit.edu/horde/imp/folders.php	HTML Document	27 KB	4/24/2005 5:21
mailbox.php?n...	https://webmail.mit.edu/horde/imp/mailbox.php?newmail_po...	HTML Document	36 KB	4/24/2005 5:21
mailbox	https://webmail.mit.edu/horde/imp/mailbox.php?page=60	HTML Document	36 KB	4/24/2005 5:21
message	https://webmail.mit.edu/horde/imp/message.php?index=2194	HTML Document	26 KB	4/24/2005 5:21
mailbox	https://webmail.mit.edu/horde/imp/mailbox.php?start=593	HTML Document	35 KB	4/24/2005 5:21
message	https://webmail.mit.edu/horde/imp/message.php?index=2192	HTML Document	24 KB	4/24/2005 5:21
web.mit	http://web.mit.edu/	HTML Document	17 KB	None

MIT Humanities Library, April 25, 2005

4 out of 4 computers had personal email in their browser caches.


MIT WebMail - Microsoft Internet Explorer - [Working Offline]
File Edit View Favorites Tools Help
Back Search Favorites
Address
Go Links
Google Search Web 0 blocked AutoFill Options
MIT WebMail
An email service from Information Services & Technology
INBOX Empty Trash Compose Folders Options Search Directory Help Logout Open Folder
Folder Navigator
Choose Action:
Expand All | Collapse All
Internet
start
MIT WebMail - Micros...
Temporary Internet F...
lib1 - Paint
10:54 AM

Address: https://webmail.mit.edu/horde/imp/folders.php

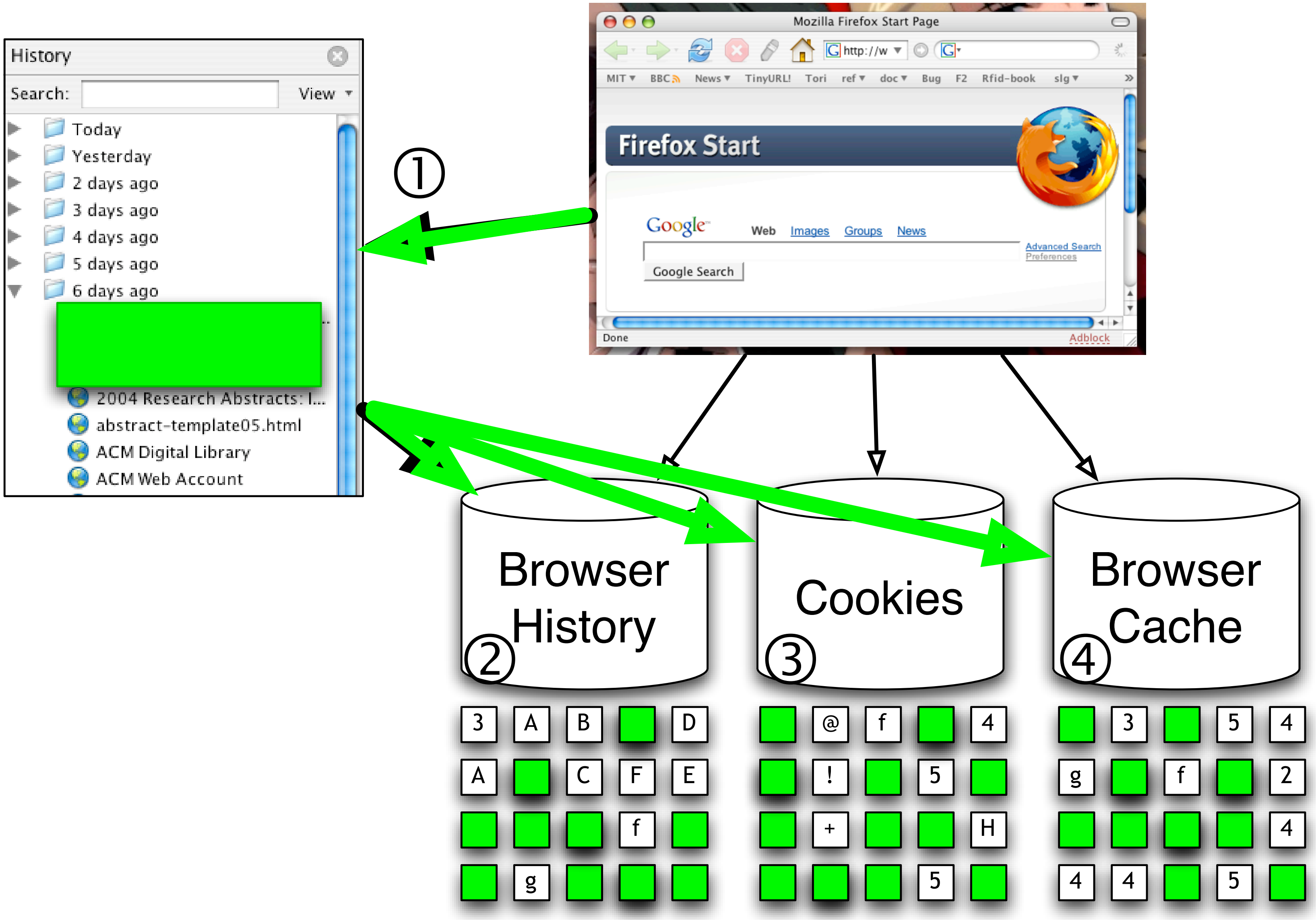
Mailbox quota: 175.84MB Used / 244.14MB Total (72.02%)

	#	✉
□ INBOX	1367	0
□ ACF		
□ Home		
□ MITOC		
□ sent		
□ Student House		
□ Ultimate		
□ webmail-trash		
□ Work		

Size	Expires
41 KB	None
66 KB	None
53 KB	None
26 KB	None
17 KB	None
17 KB	None
17 KB	None
24 KB	None
20 KB	None
35 KB	None
26 KB	4/24/2005 5:58
4 KB	4/24/2005 4:58
38 KB	None
66 KB	None
84 KB	4/26/2005 5:58
29 KB	None
5 KB	None
12 KB	None
33 KB	4/24/2005 5:20
27 KB	4/24/2005 5:21
36 KB	4/24/2005 5:21
36 KB	4/24/2005 5:21
26 KB	4/24/2005 5:21
35 KB	4/24/2005 5:21
24 KB	4/24/2005 5:21
17 KB	None

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

The solution is to integrate the history, cache and cookies



3	A	B	🟩	D
A	🟩	C	F	E
🟩	🟩	🟩	f	🟩
🟩	g	🟩	🟩	🟩

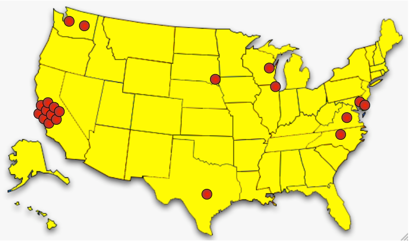
🟩	@	f	🟩	4
🟩	!	🟩	5	🟩
🟩	+	🟩	🟩	H
🟩	🟩	🟩	5	🟩

🟩	3	🟩	5	4
g	🟩	f	🟩	2
🟩	🟩	🟩	🟩	4
4	4	🟩	5	🟩

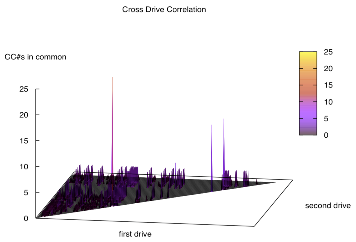
This talk presents new tools and techniques for performing forensic analysis on a large number of disk drives.

The drives Project

The Traceback Study



Cross Drive Forensics and AFF



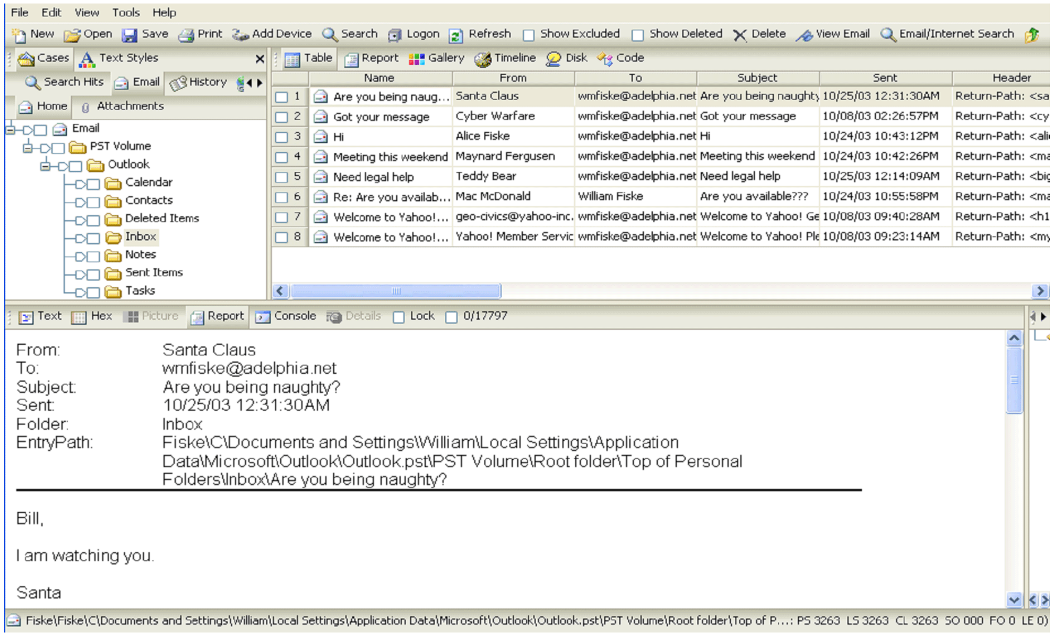
Today’s forensic tools are designed for one drive at a time.

Primary Goals: Search and Recovery.

Interactive user interface.

Usage scenarios:

- Recovery of “deleted” files.
- Child porn scanning.
- Trial preparation.



	Name	From	To	Subject	Sent	Header
□ 1	Are you being naug…	Santa Claus	wmfiske@adelphia.net	Are you being naughty	10/25/03 12:31:30AM	Return-Path: <sa
□ 2	Got your message	Cyber Warfare	wmfiske@adelphia.net	Got your message	10/08/03 02:26:57PM	Return-Path: <cy
□ 3	Hi	Alice Fiske	wmfiske@adelphia.net	Hi	10/24/03 10:43:12PM	Return-Path: <ali
□ 4	Meeting this weekend	Maynard Fergusen	wmfiske@adelphia.net	Meeting this weekend	10/24/03 10:42:26PM	Return-Path: <ma
□ 5	Need legal help	Teddy Bear	wmfiske@adelphia.net	Need legal help	10/25/03 12:14:09AM	Return-Path: <big
□ 6	Re: Are you availab…	Mac McDonald	William Fiske	Are you available???	10/24/03 10:55:58PM	Return-Path: <ma
□ 7	Welcome to Yahoo!…	geo-civics@yahoo-inc.	wmfiske@adelphia.net	Welcome to Yahoo! Ge	10/08/03 09:40:28AM	Return-Path: <h1
□ 8	Welcome to Yahoo!…	Yahoo! Member Servic	wmfiske@adelphia.net	Welcome to Yahoo! Ple	10/08/03 09:23:14AM	Return-Path: <my

From: Santa Claus
To: wmfiske@adelphia.net
Subject: Are you being naughty?
Sent: 10/25/03 12:31:30AM
Folder: Inbox
EntryPath: Fiske\C\Documents and Settings\William\Local Settings\Application Data\Microsoft\Outlook\Outlook.pst\PST Volume\Root folder\Top of Personal Folders\Inbox\Are you being naughty?

Bill,

I am watching you.

Santa

Fiske\Fiske\C\Documents and Settings\William\Local Settings\Application Data\Microsoft\Outlook\Outlook.pst\PST Volume\Root folder\Top of P…: PS 3263 LS 3263 CL 3263 SO 000 FO 0 LE 0)

Today’s tools choke when confronted with hundreds or thousands of drives.

Which drives were used by my target?

Do any drives belong to the target’s associates?

Who is talking to who?

Where should I start?

Police departments and intelligence agencies have thousands of drives...

Additional problems with today's tools

- Improper prioritization

  Letting priority be determined by the statute of limitations.

- Lost opportunities for data correlation

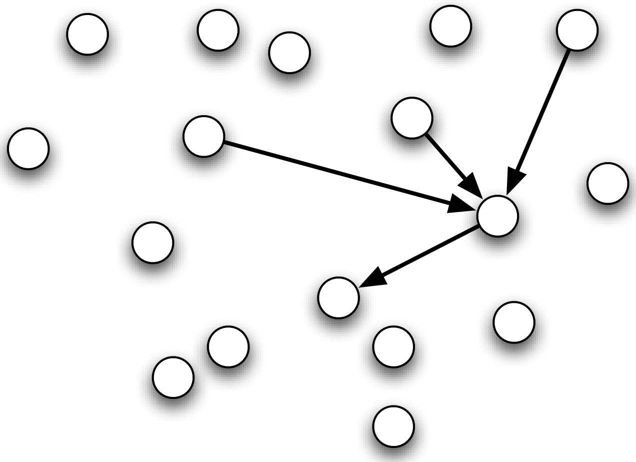
  Was a message on hard drive X sent to hard drive Y?

- Emphasis on document recovery rather than in furthering the investigation.

Correlating data *between* drives is an untapped opportunity.

How large is my target’s reach?

Who is in the organization?



Captured drives are an ideal social network analysis.

Forensic Feature Extraction and Cross-Drive Analysis

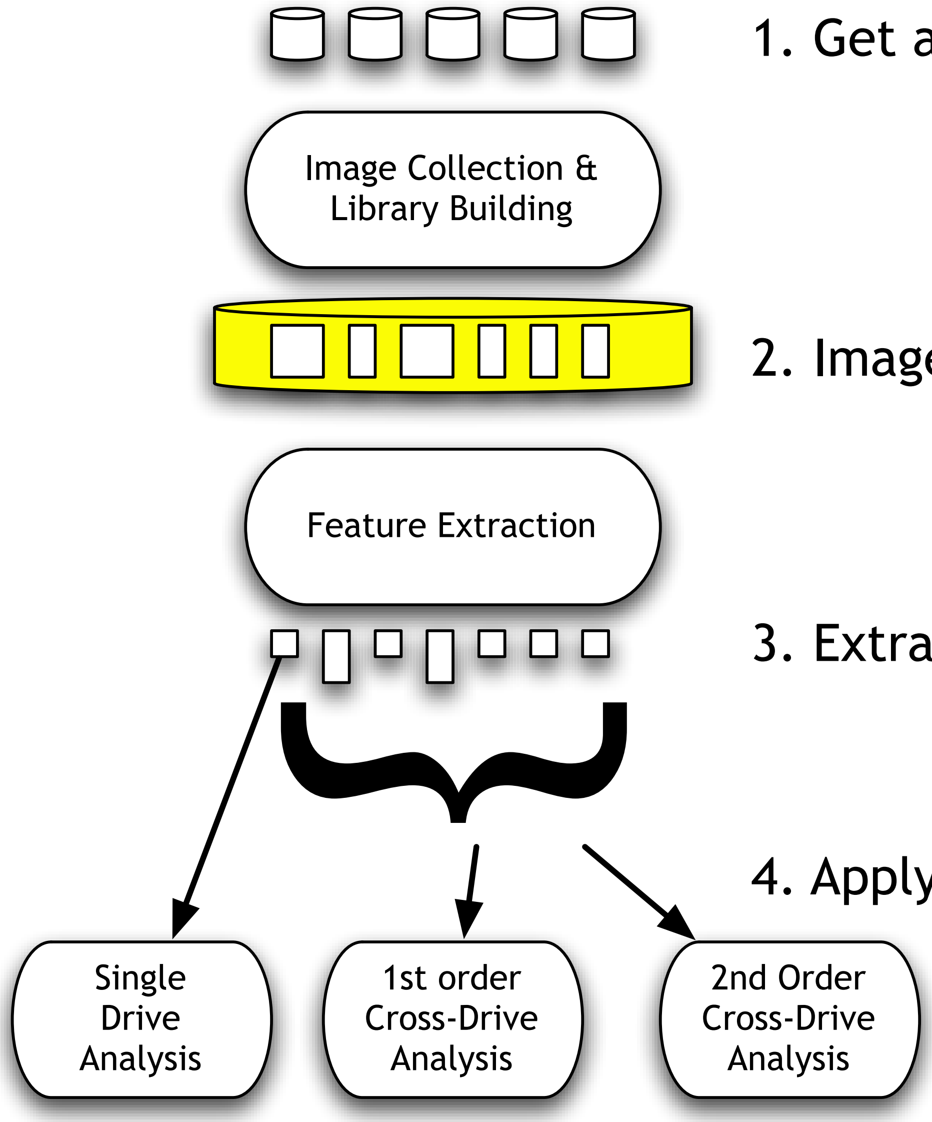


Image Collection & Library Building

Feature Extraction

Single Drive Analysis

1st order Cross-Drive Analysis

2nd Order Cross-Drive Analysis

1. Get a lot of drives

2. Image to a big disk

3. Extract the Features

4. Apply statistics and correlation

Uses of Cross-Drive Analysis

1. Automatic identification of hot drives
2. Improvements to single-drive systems
3. Identification of social network membership
4. Unsupervised social network discovery

Related Work:

- Garfinkel & Shelat, 158 drives, 2002
- AFF [Garfinkel, Malan, et al; 2006]

Feature extractors find *pseudo-unique* features

Pseudo-Unique characteristics:

- Long enough so collisions by chance are unlikely.
- Recognizable with regular expressions.
- Persistent over time.
- Correlated with specific documents, people or organizations.

Typical Features:

- email addresses
- Message-IDs
- Subject: lines
- Cookies
- US Social Security Numbers
- Credit card numbers
- Hash codes of drive sectors

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 XXXXXXXXXXXXXXXX)

- 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 Drive #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
```

Even with the tests, there are occasional false positives.

CCN scan of Drive #115: (772MB)

Test	# pass
pattern	9196
known prefixes	898
CCV1	29
patterns	27
histogram	13

```
.................@:|44444486666108|:<@<74444:@@@<<44   pos=82473275
............#"&’&&’|445447667667667|..050014&’4"1"&’.  pos=86493675
......221267241667&|454676676654450|&566746566726322.  pos=86507818
3..30210212676677..|30232676630232|.1.........001.01   pos=86516059
"&#&&’&41&&’645445&|454454672676632|.3............0..  pos=86523223
..........".#""#"&’|445467667227023|..............366  pos=87540819
D#9?.32400.,,+14%?B|499745255278101|*02)46+;<17756669  pos=118912826
.GGJJB...>.JJGG...G|3534554333511116|...............6  pos=197711868
%.....}}}}}}.......|44444322233345|.....}}}}}}......   pos=228610295
%6"!) .&*%,,%-0)07.|373484553420378|<67<038+.5(+0+.3.  pos=638491849
%6"!) .&*%,,%-0)07.|373484553420378|<67<038+.5(+0+.3.  pos=645913801
```

CDA Prototype System

1000 drives purchased on secondary market (1998–2006)

750 images

1.5TB data compressed.

Many different organizations.

Single-drive feature application: drive attribution.

Drive #51: Top email addresses (sanitized)

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

Most common email address is (usually) drive's primary user.

Attribution histogram works even with lightly-used drives.

Extracted Email Addresses	Count on Drive #80	Total drives with address
premium-server@thawte.com	117	278
server-certs@thawte.com	104	278
CPS-requests@verisign.com	61	286
personal-premium@thawte.com	44	253
personal-basic@thawte.com	42	250
personal-freemail@thawte.com	40	250
info@netscape.com	36	58
ANGIE@ALPHA.com	32	1
BARRY@BETA.com	23	1
CHARLES@GAMMA.com	21	1
DAVE.HALL@DELTA.com	21	1
DAPHNE@UNIFORM.com	20	1
ELLY@LIMA.com	18	1
FRANK@ECHO.com	16	1
HUGH@LIMA.com	16	1
IGGY@LIMA.com	16	1
GRETTA@XYZZY.com	15	1
VISTA@SNARF.com	15	1

Email addresses found on $\approx> 20$ drives are not pseudo-unique

First Order Cross-Drive Analysis: $O(n)$ operations on feature files

Applications:

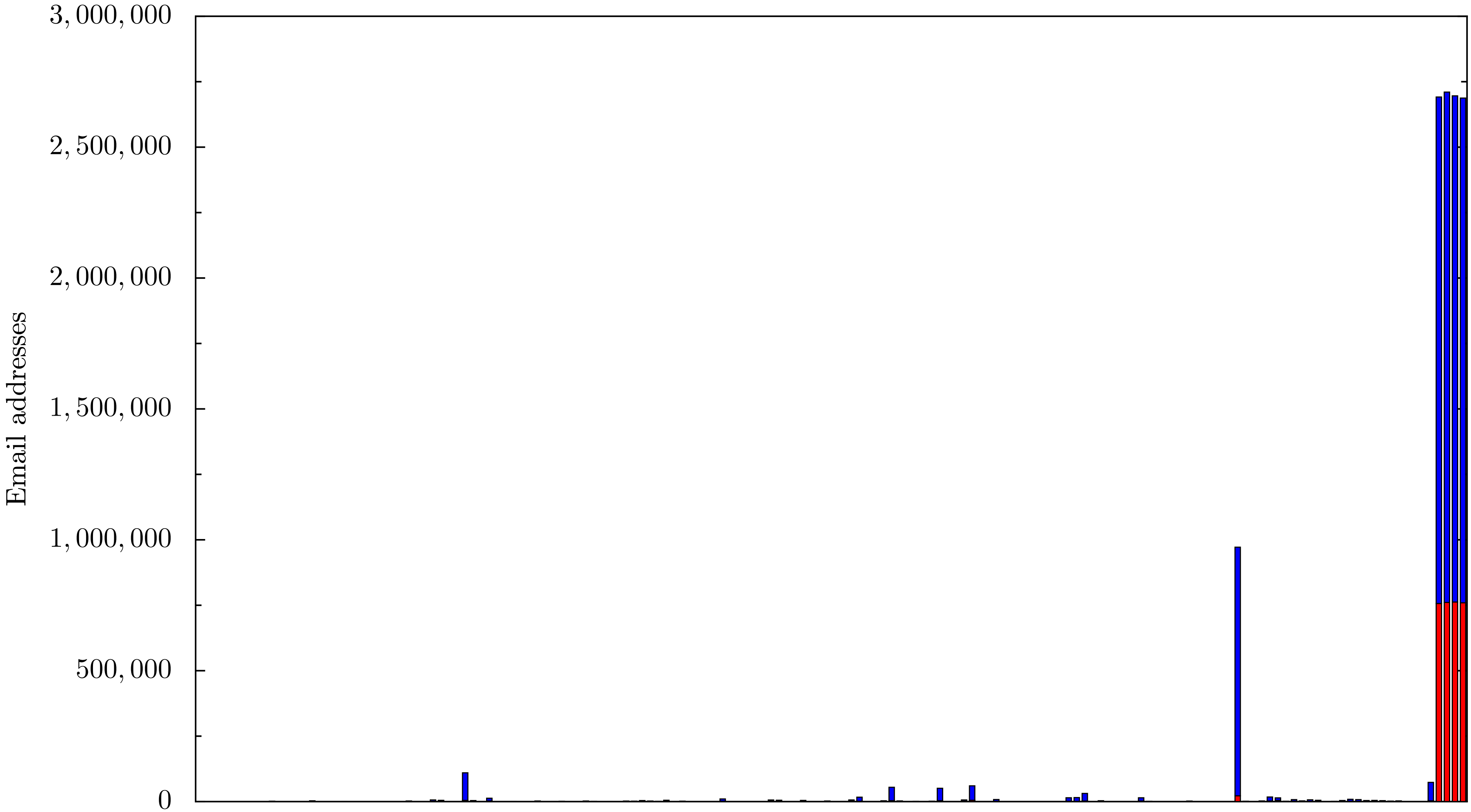
- Automatically building stop lists
- Hot drive identification

Automatic "stop lists:"
features on many drives are not pseudo-unique.

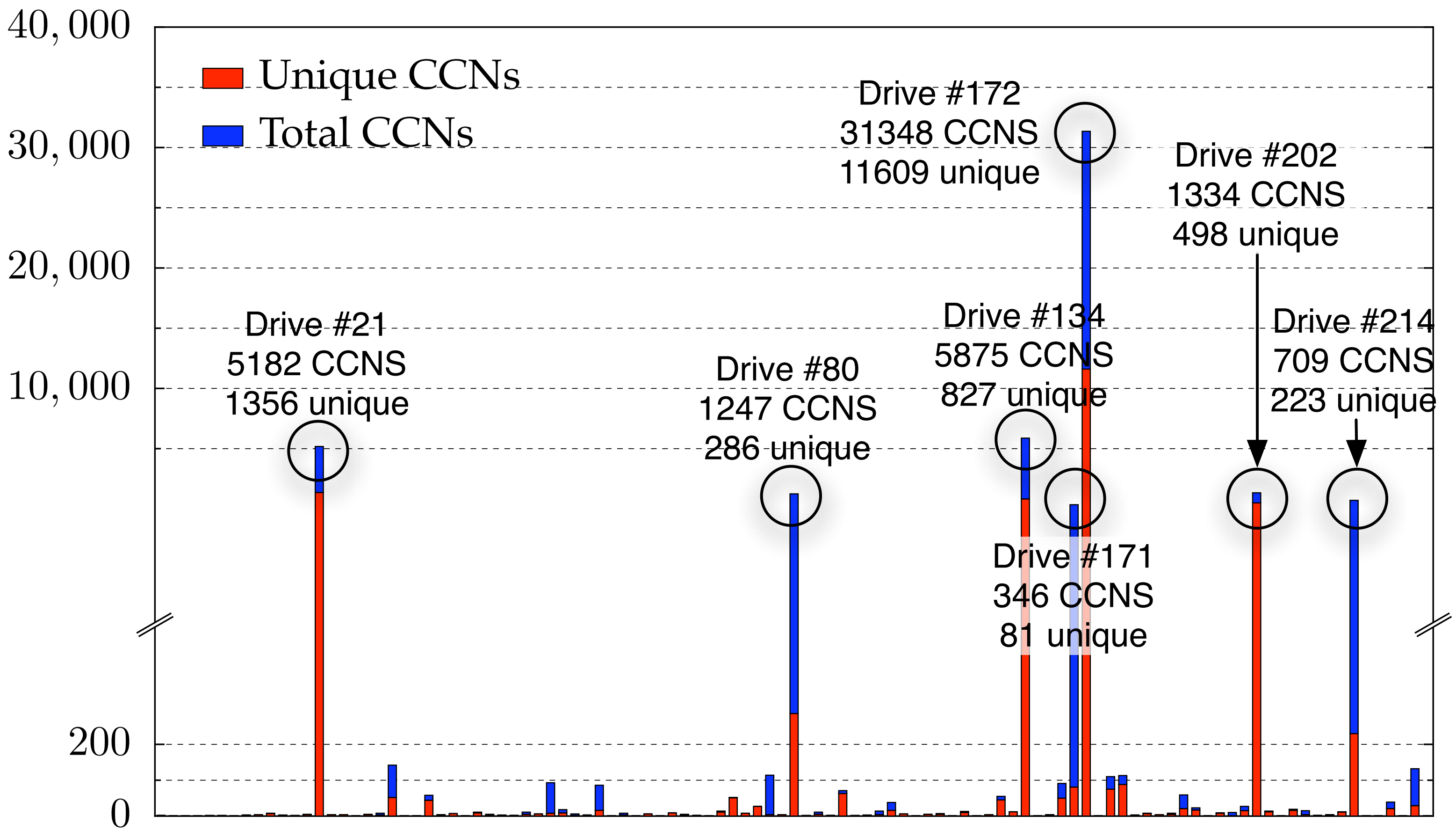
Extracted Email Address	Drives with address	Total count in corpus
CPS-requests@verisign.com	286	64424
server-certs@thawte.com	278	32873
premium-server@thawte.com	278	31141
Mouse.Exe@Mouse.Com	262	493
LMouse.Exe@LMouse.Com	262	493
personal-premium@thawte.com	253	14660
personal-freemail@thawte.com	250	14843
personal-basic@thawte.com	250	14290
inet@microsoft.com	244	31456
mazrob@panix.com(*)	221	3265
java-security@java.sun.com	200	1200
java-io@java.sun.com	198	413
someone@microsoft.com	195	6193
bugs@java.sun.com	192	351
ca@digsigtrust.com	173	36800
name@company.com	169	1763

***`mazrob@panix.com` appears in `clickerx.wav` (Utopia Sound Scheme)**

A graph of # email addresses on each drive automatically identified drives used by bulk e-mailers.

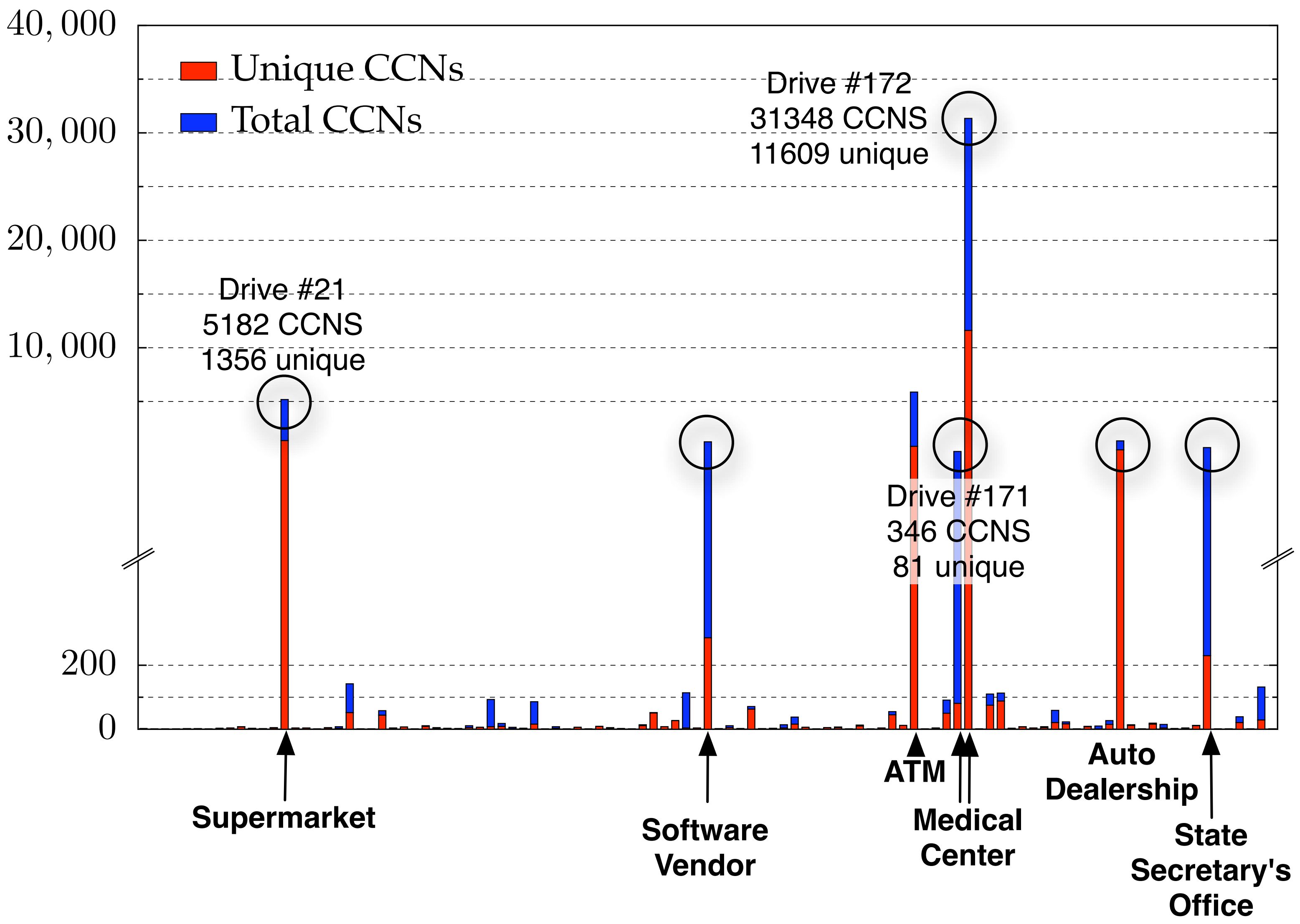


Hot drive identification:
Drives with high response warrant further attention.



Only 7 drives had more than 300 credit card numbers.

Hot drive identification:
Drives with high response warrant further attention.



These drives represent significant privacy violations.

First order analysis of # SSNs

	Unique	Total
Drive	SSNs	SSNs
Drive #959	260	447
Drive #974	178	674
Drive #696	33	872
Drive #969	33	33
Drive #690	8	14
Drive #680	2	4

Drive #959 contained consumer credit applications.

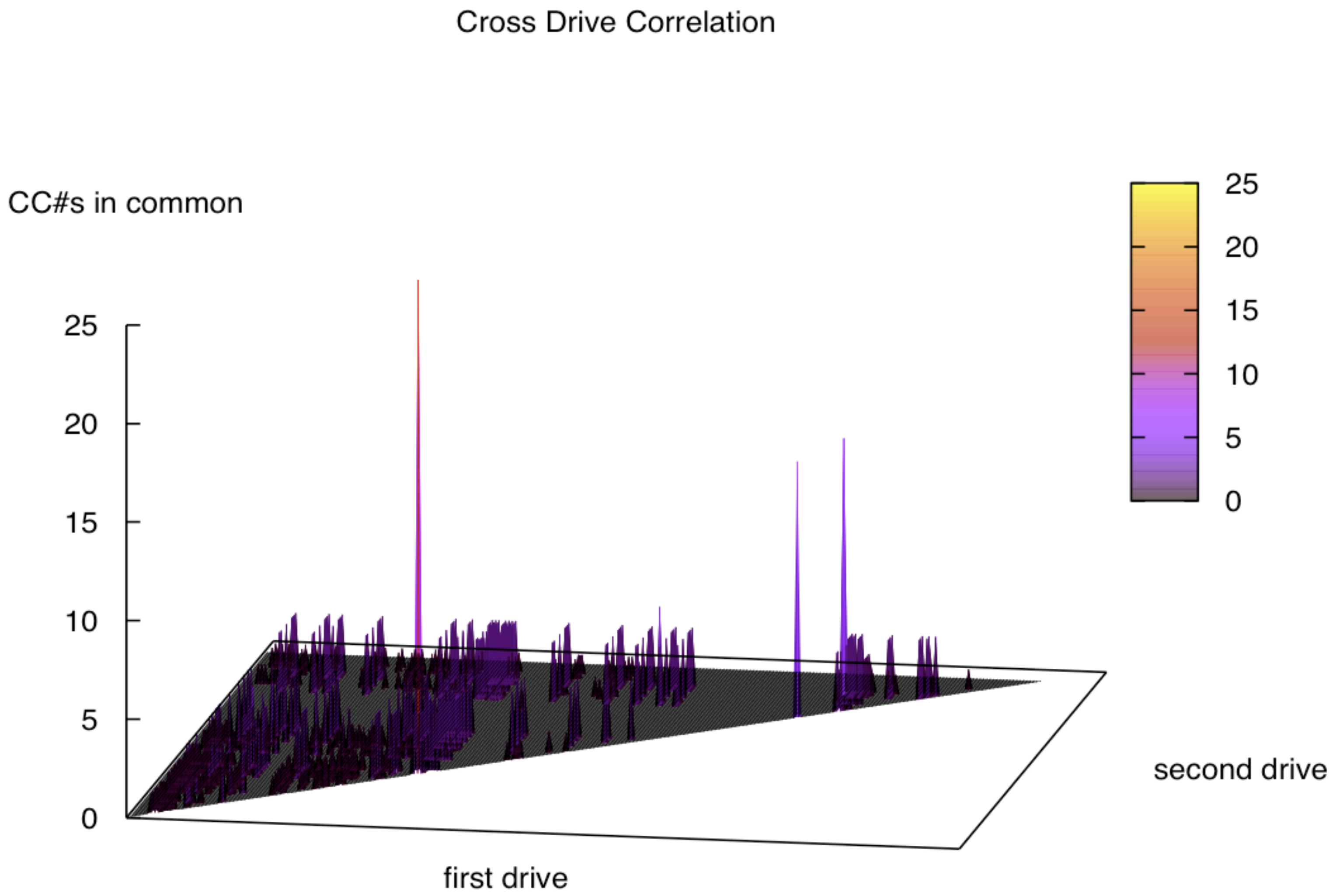
Second-order analysis uses the *multi-drive correlation*

$$
\begin{aligned}
D &= \# \text{ of drives} \\
F &= \# \text{ of extracted features} \\
d_0 \ldots d_D &= \text{Drives in corpus} \\
f_0 \ldots f_F &= \text{Extracted features} \\
FP(f_n, d_n) &= \begin{cases} 0 & f_n \text{ not present on } d_n \\ 1 & f_n \text{ present on } d_n \end{cases}
\end{aligned}
$$

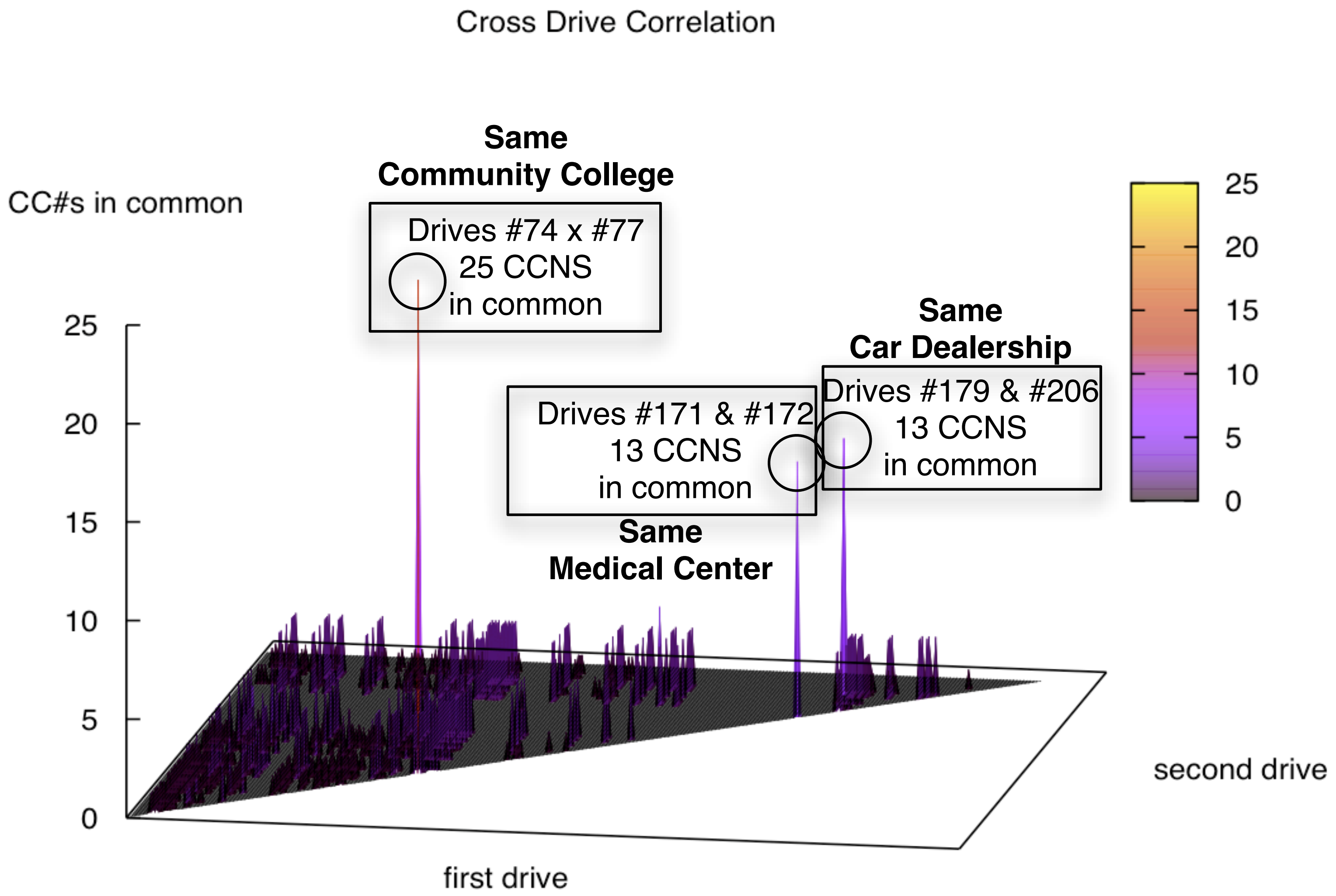
Scoring Function:

$$S_1(d_1, d_2) = \sum_{n=0}^{F} FP(f_n, d_1) \times FP(f_n, d_2)$$

Graph of scoring function:



Graph of scoring function:



Same Community College

Drives #74 x #77
25 CCNS
in common

Same Car Dealership

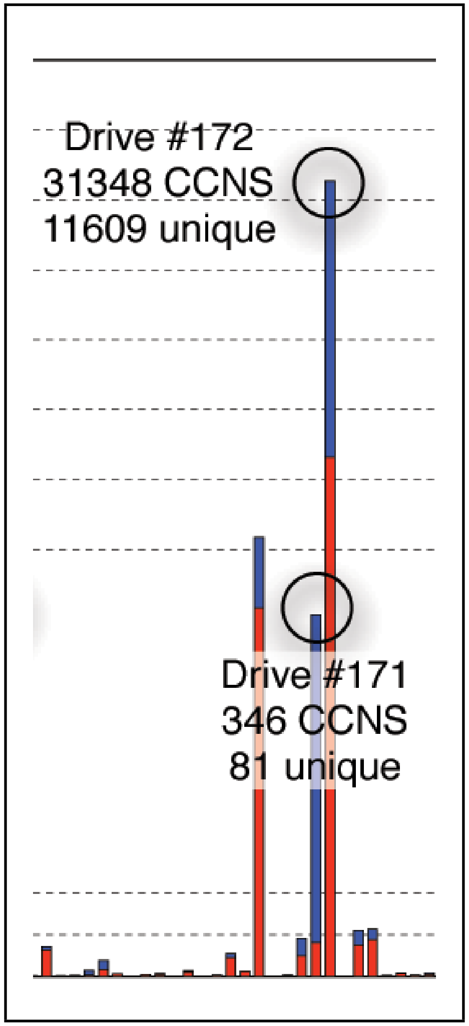
Drives #179 & #206
13 CCNS
in common

Drives #171 & #172
13 CCNS
in common

Same Medical Center

The three correlated drives have an extrinsic relationship. (180 drive corpus)

The correlation between Drives #171 and #172 tells a story...



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.

Other CCN correlations

#74, #77	Same college in Pacific Northwest. Correlated on CCN “false positive.”
#339 – #356	All used by same New York travel agency
#716, #718	Both from Union City, CA dealer
#814, #820	Both from same Stamford, CT dealer

In two cases, cross-drive correlation discovered drive cataloging errors!

SSN correlation: identical documents on different drives

SSN_1	#342, #343, #356	“Thanks, Laurie” memo
SSN_2	#350, #355	“great grandchildren” memo

But ignore these numbers:

666-66-6666	#313, #427, #429, #430, #612, #627, #744, #770, #808
123-45-6789	#328, #343, #345, #350, #351, #700
555-55-5555	#612, #690

Possible reasons for the same SSN found on two drives

- Two copies of the same document
- Two documents about the same person
- Accidental mismatch

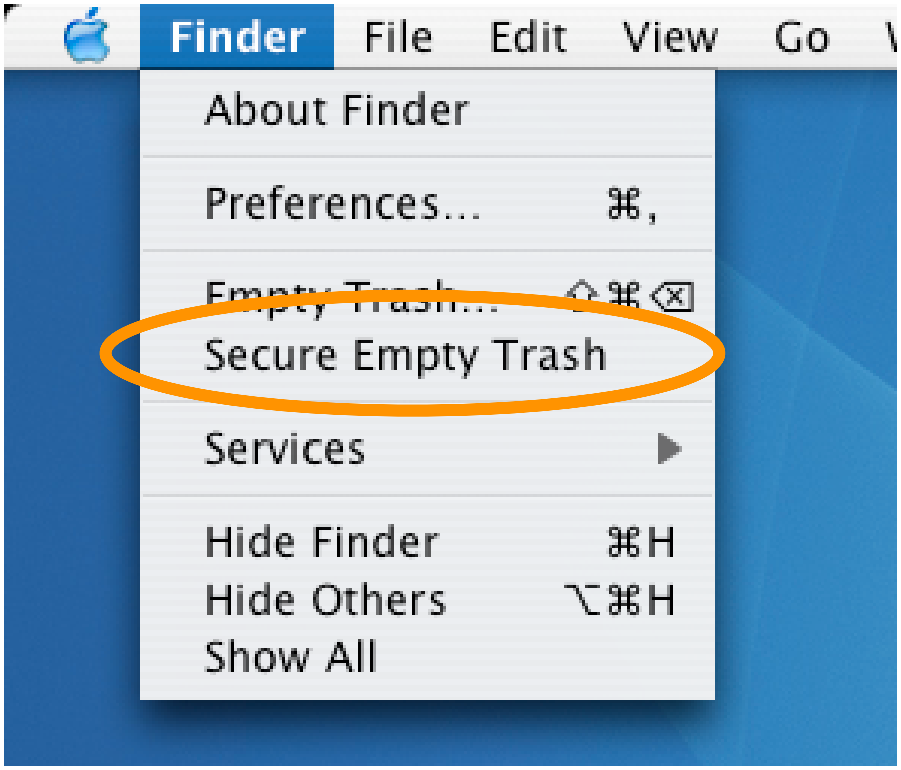
Chance of a false match is 1 in 10^9.

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.”

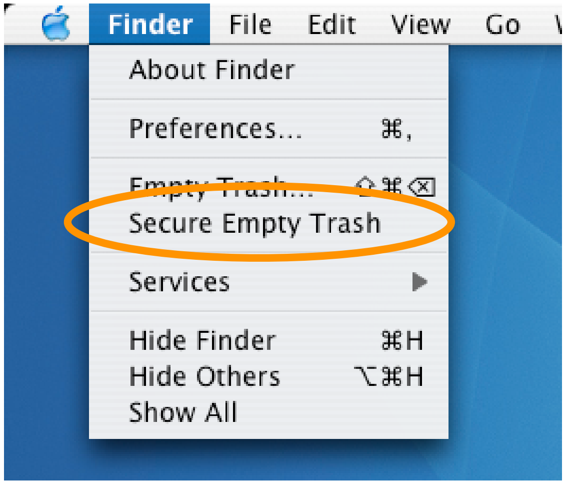
Testimony: `http://tinyurl.com/cd2my`

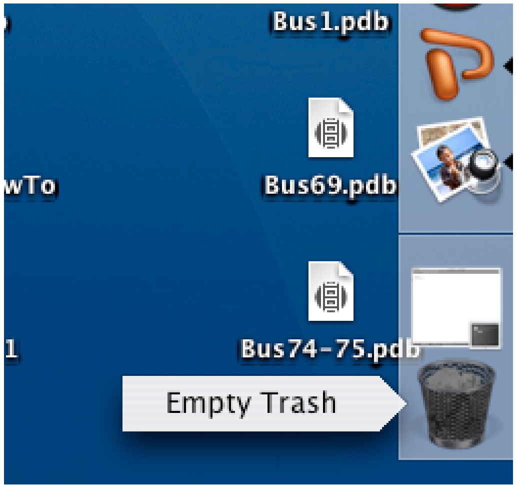
Technical reactions to this research: “Secure Empty Trash” in MacOS 10.3.



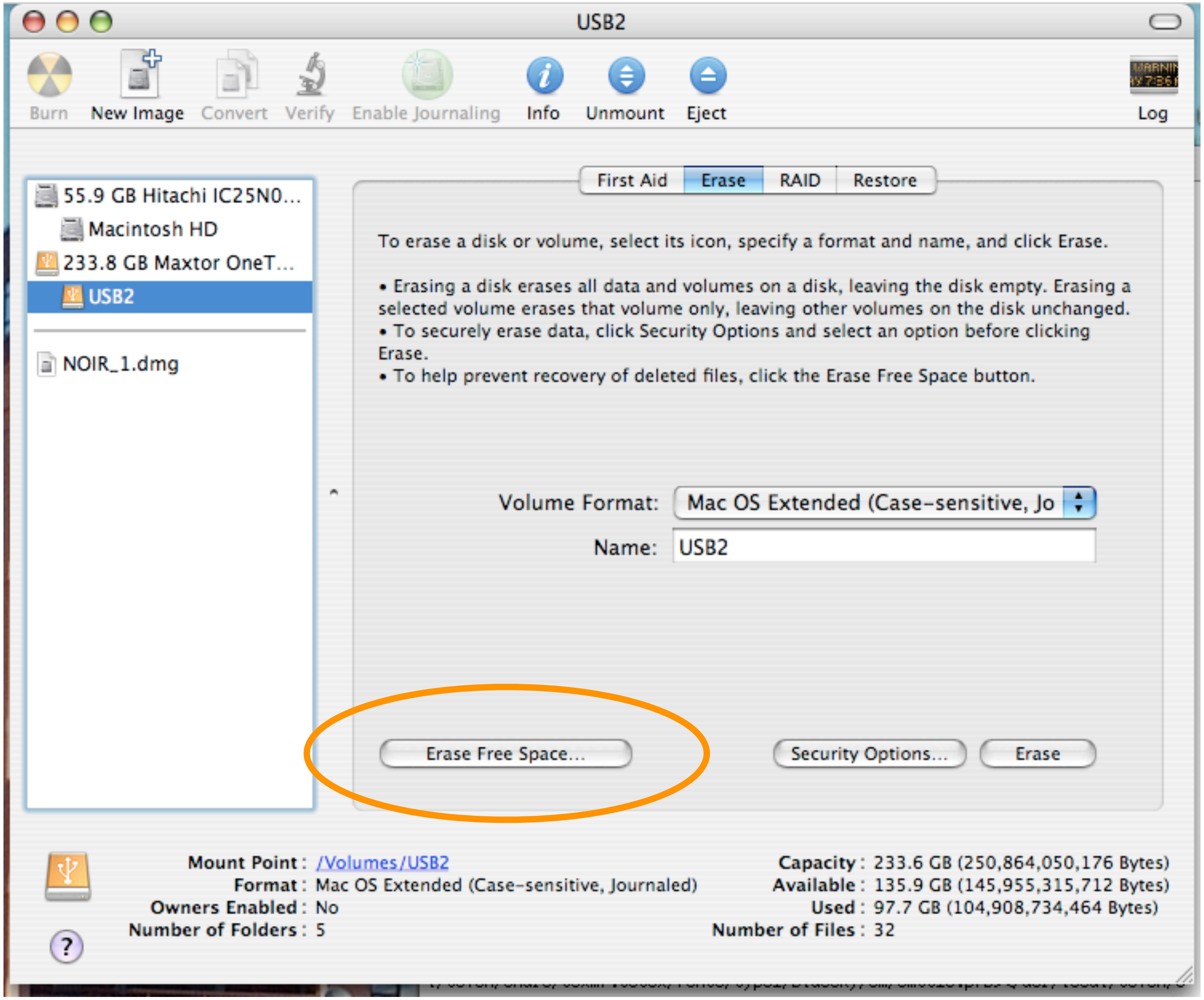
Unfortunately, “Secure Empty Trash” is incomplete.

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





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



To erase a disk or volume, select its icon, specify a format and name, and click Erase.

- Erasing a disk erases all data and volumes on a disk, leaving the disk empty. Erasing a selected volume erases that volume only, leaving other volumes on the disk unchanged.
- To securely erase data, click Security Options and select an option before clicking Erase.
- To help prevent recovery of deleted files, click the Erase Free Space button.

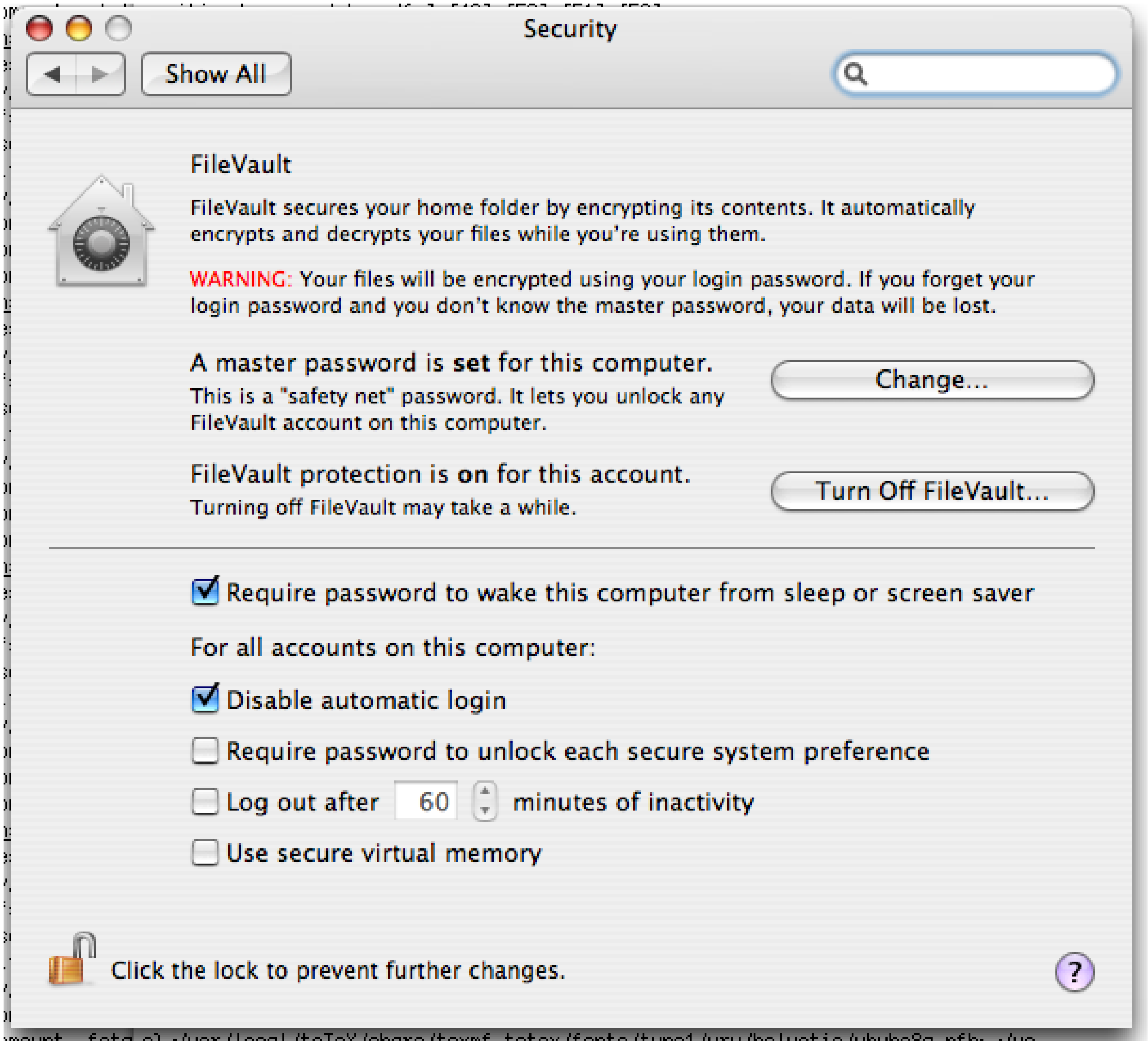
Volume Format: Mac OS Extended (Case-sensitive, Jo

Name: USB2

Mount Point : /Volumes/USB2
Format : Mac OS Extended (Case-sensitive, Journaled)
Owners Enabled : No
Number of Folders : 5

Capacity : 233.6 GB (250,864,050,176 Bytes)
Available : 135.9 GB (145,955,315,712 Bytes)
Used : 97.7 GB (104,908,734,464 Bytes)
Number of Files : 32

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



FileVault

FileVault secures your home folder by encrypting its contents. It automatically encrypts and decrypts your files while you're using them.

WARNING: Your files will be encrypted using your login password. If you forget your login password and you don't know the master password, your data will be lost.

A master password is **set** for this computer.

This is a "safety net" password. It lets you unlock any FileVault account on this computer.

FileVault protection is **on** for this account.

Turning off FileVault may take a while.

☑ Require password to wake this computer from sleep or screen saver

For all accounts on this computer:

☑ Disable automatic login

☐ Require password to unlock each secure system preference

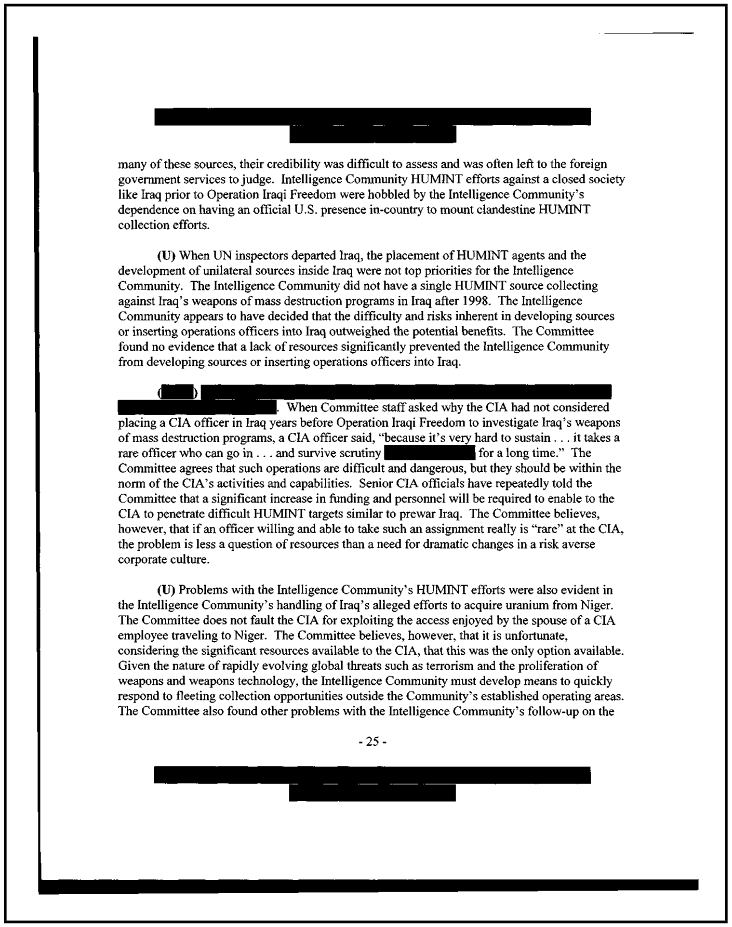
☐ Log out after 60 minutes of inactivity

☐ Use secure virtual memory

Click the lock to prevent further changes.

Current 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.



[redacted]

many of these sources, their credibility was difficult to assess and was often left to the foreign government services to judge. Intelligence Community HUMINT efforts against a closed society like Iraq prior to Operation Iraqi Freedom were hobbled by the Intelligence Community’s dependence on having an official U.S. presence in-country to mount clandestine HUMINT collection efforts.

(U) When UN inspectors departed Iraq, the placement of HUMINT agents and the development of unilateral sources inside Iraq were not top priorities for the Intelligence Community. The Intelligence Community did not have a single HUMINT source collecting against Iraq’s weapons of mass destruction programs in Iraq after 1998. The Intelligence Community appears to have decided that the difficulty and risks inherent in developing sources or inserting operations officers into Iraq outweighed the potential benefits. The Committee found no evidence that a lack of resources significantly prevented the Intelligence Community from developing sources or inserting operations officers into Iraq.

([redacted]) [redacted]. When Committee staff asked why the CIA had not considered placing a CIA officer in Iraq years before Operation Iraqi Freedom to investigate Iraq’s weapons of mass destruction programs, a CIA officer said, “because it’s very hard to sustain . . . it takes a rare officer who can go in . . . and survive scrutiny [redacted] for a long time.” The Committee agrees that such operations are difficult and dangerous, but they should be within the norm of the CIA’s activities and capabilities. Senior CIA officials have repeatedly told the Committee that a significant increase in funding and personnel will be required to enable to the CIA to penetrate difficult HUMINT targets similar to prewar Iraq. The Committee believes, however, that if an officer willing and able to take such an assignment really is “rare” at the CIA, the problem is less a question of resources than a need for dramatic changes in a risk averse corporate culture.

(U) Problems with the Intelligence Community’s HUMINT efforts were also evident in the Intelligence Community’s handling of Iraq’s alleged efforts to acquire uranium from Niger. The Committee does not fault the CIA for exploiting the access enjoyed by the spouse of a CIA employee traveling to Niger. The Committee believes, however, that it is unfortunate, considering the significant resources available to the CIA, that this was the only option available. Given the nature of rapidly evolving global threats such as terrorism and the proliferation of weapons and weapons technology, the Intelligence Community must develop means to quickly respond to fleeting collection opportunities outside the Community’s established operating areas. The Committee also found other problems with the Intelligence Community’s follow-up on the

- 25 -

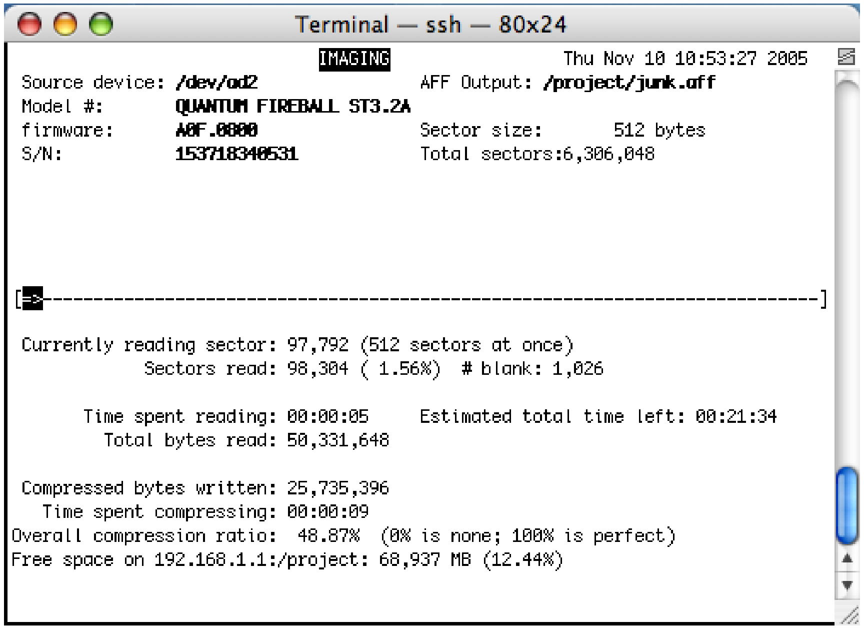
[redacted]

Current 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.

Current Work: AFF Toolkit

- Improved imaging, storage and backup.
- Web-based database of hash codes.

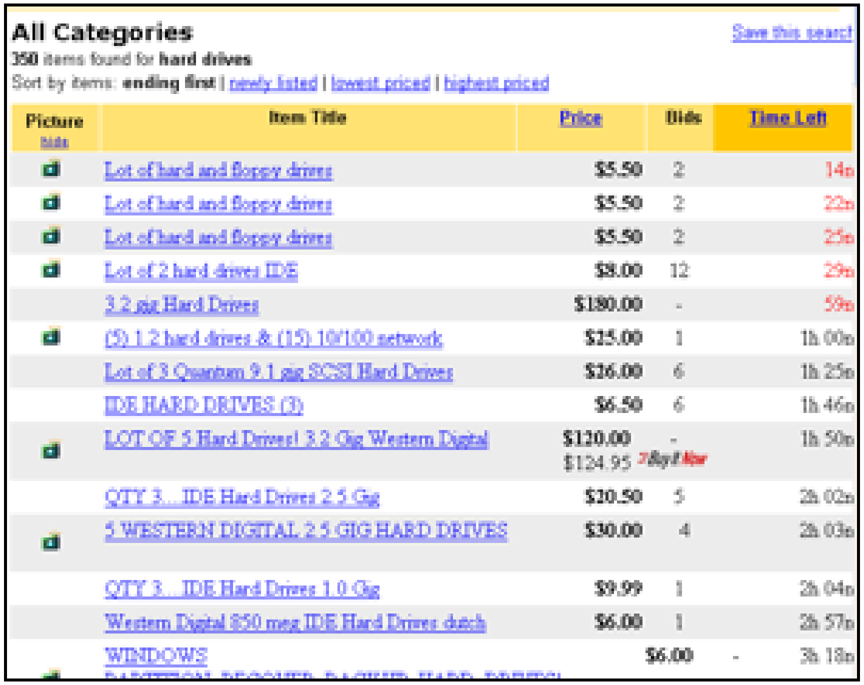


Terminal — ssh — 80x24

```
                             IMAGING                Thu Nov 10 10:53:27 2005
 Source device: /dev/ad2                AFF Output: /project/junk.aff
 Model #:       QUANTUM FIREBALL ST3.2A
 firmware:      A0F.0000                Sector size:       512 bytes
 S/N:           153718340531            Total sectors:6,306,048




[=>---------------------------------------------------------------------------]

 Currently reading sector: 97,792 (512 sectors at once)
             Sectors read: 98,304 ( 1.56%)  # blank: 1,026

       Time spent reading: 00:00:05     Estimated total time left: 00:21:34
         Total bytes read: 50,331,648

 Compressed bytes written: 25,735,396
   Time spent compressing: 00:00:09
Overall compression ratio:  48.87%  (0% is none; 100% is perfect)
Free space on 192.168.1.1:/project: 68,937 MB (12.44%)
```

Current Work: Economics and Society

- Who is buying used hard drives and why?
- Compliance with FACT-A
- Increasing adoption of S/MIME-signed mail



All Categories — Save this search

350 items found for **hard drives**

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

Picture hide	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	$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 Buy It Now	-	1h 50m
	QTY 3....IDE Hard Drives 2.5 Gig	$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 Gig	$9.99	1	2h 04m
	Western Digital 850 meg IDE Hard Drives dutch	$6.00	1	2h 57m
📷	WINDOWS PARTITION, RECOVER, BACKUP, HARD DRIVES!	$6.00	-	3h 18m

Summary

A lot of information is left on used drives.

Working with these drives gives insights for improving forensic practice.

Cross drive forensics and AFF are two tangible benefits to date.

Questions?

References

[Garfinkel & Shelat 03] Garfinkel, S. and Shelat, A., "Remembrance of Data Passed: A Study of Disk Sanitization Practices," *IEEE Security and Privacy*, January/February 2003. `http://www.simson.net/clips/academic/2003.IEEE.DiskDriveForensics.pdf`

[Markoff 97] John Markoff, "Patient Files Turn Up in Used Computer," *The New York Times*, April 1997.

[Villano 02] Matt Villano, "Hard-Drive Magic: Making Data Disappear Forever," *The New York TImes*, May 2002.