Clean Delete

Simson L. Garfinkel

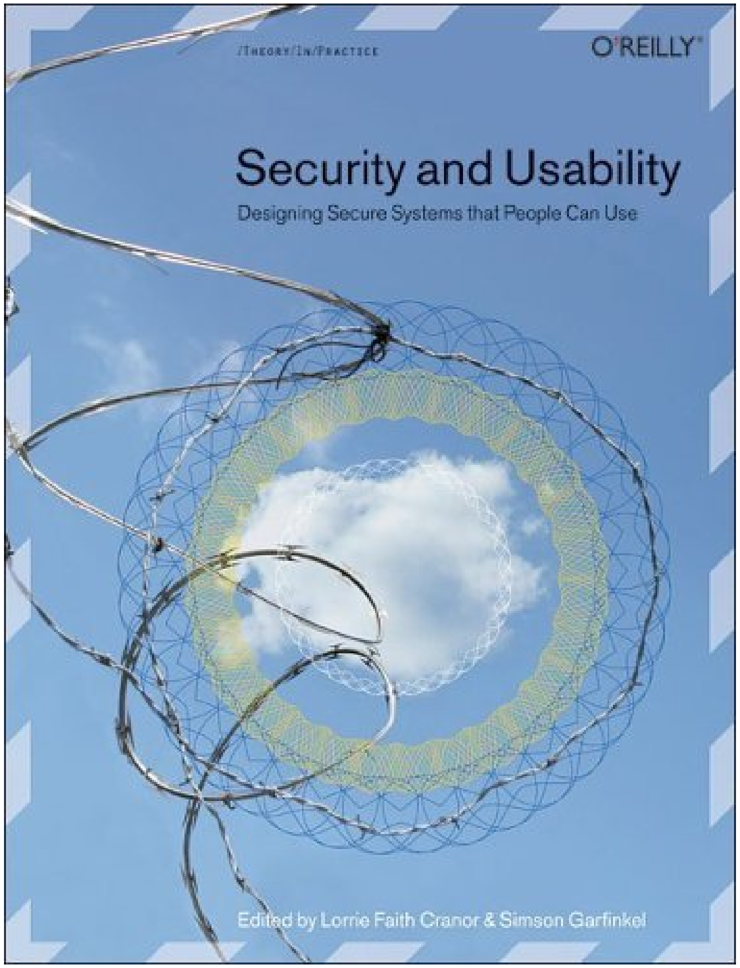
Center for Research on Computation and Society

Harvard University

April 3, 2006

Aligning Security & Usability.

There are two main approaches to this work:


/Theory/In/Practice
O'REILLY
Security and Usability
Designing Secure Systems that People Can Use
Edited by Lorrie Faith Cranor & Simson Garfinkel

✘ Work on authentication

✘ Work on new interfaces.

- biometrics
- better passwords
- anti-phishing

Cranor & Garfinkel, 2005

A different approach to Usability & Security:

✔ Revisit underlying models and mechanisms to make systems inherently more secure and usable.

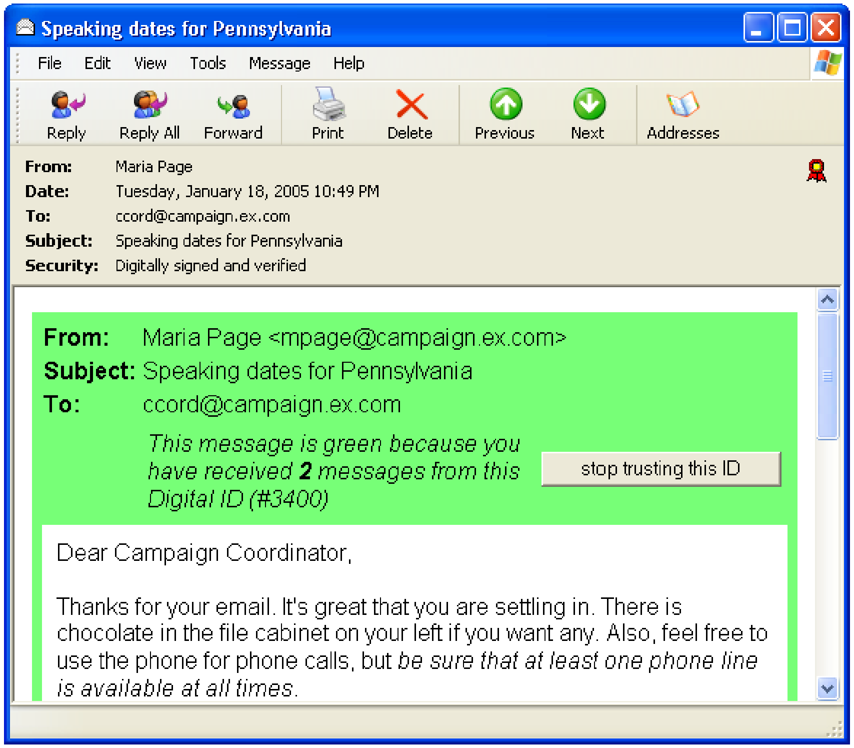
- Secure Messaging
- Data Sanitization

✔ Finding the best ideas and trying to put them all in one place.

✔ Convince vendors to incorporate these ideas into their products.

Johnny 2: Making Secure Email Easy with S/MIME and Key Continuity Management (KCM)

- Stream — a transparent PGP proxy.
- Survey of Amazon.com merchants receiving S/MIME-signed messages.
- Design of KCM plug-in for Outlook Express.
- User-test of KCM in a realistic attack scenario.



From: Maria Page
Date: Tuesday, January 18, 2005 10:49 PM
To: ccord@campaign.ex.com
Subject: Speaking dates for Pennsylvania
Security: Digitally signed and verified

From: Maria Page <mpage@campaign.ex.com>
Subject: Speaking dates for Pennsylvania
To: ccord@campaign.ex.com

This message is green because you have received ***2*** *messages from this Digital ID (#3400)*

Dear Campaign Coordinator,

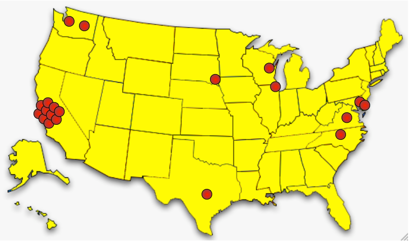
Thanks for your email. It's great that you are settling in. There is chocolate in the file cabinet on your left if you want any. Also, feel free to use the phone for phone calls, but *be sure that at least one phone line is available at all times.*

Conclusion: organizations should be sending S/MIME-signed mail.

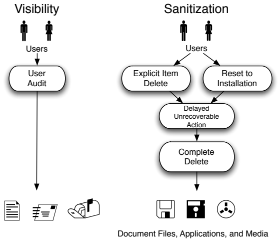
Clean Delete:
Hiding Data Is Not Good Enough!

The drives Project

The Traceback Study



Deletion Patterns



Visibility

Users

User Audit

Sanitization

Users

Explicit Item Delete

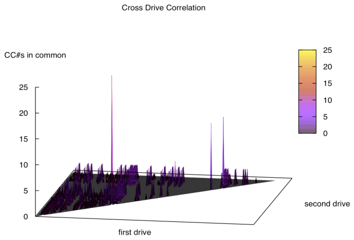
Reset to Installation

Delayed Unrecoverable Action

Complete Delete

Document Files, Applications, and Media

Cross Drive Analysis



Data Sanitization: What’s on this computer?

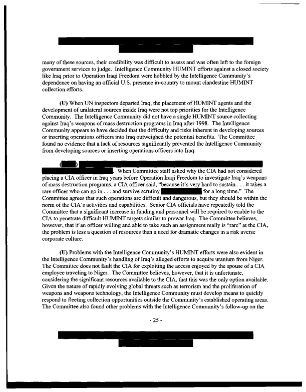
Purchased for $10 in 1998 from a retail computer store.

Hidden information is a widespread Usability/Security problem today.

Lawfirm Server

USB drive

Hard Drive



[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]

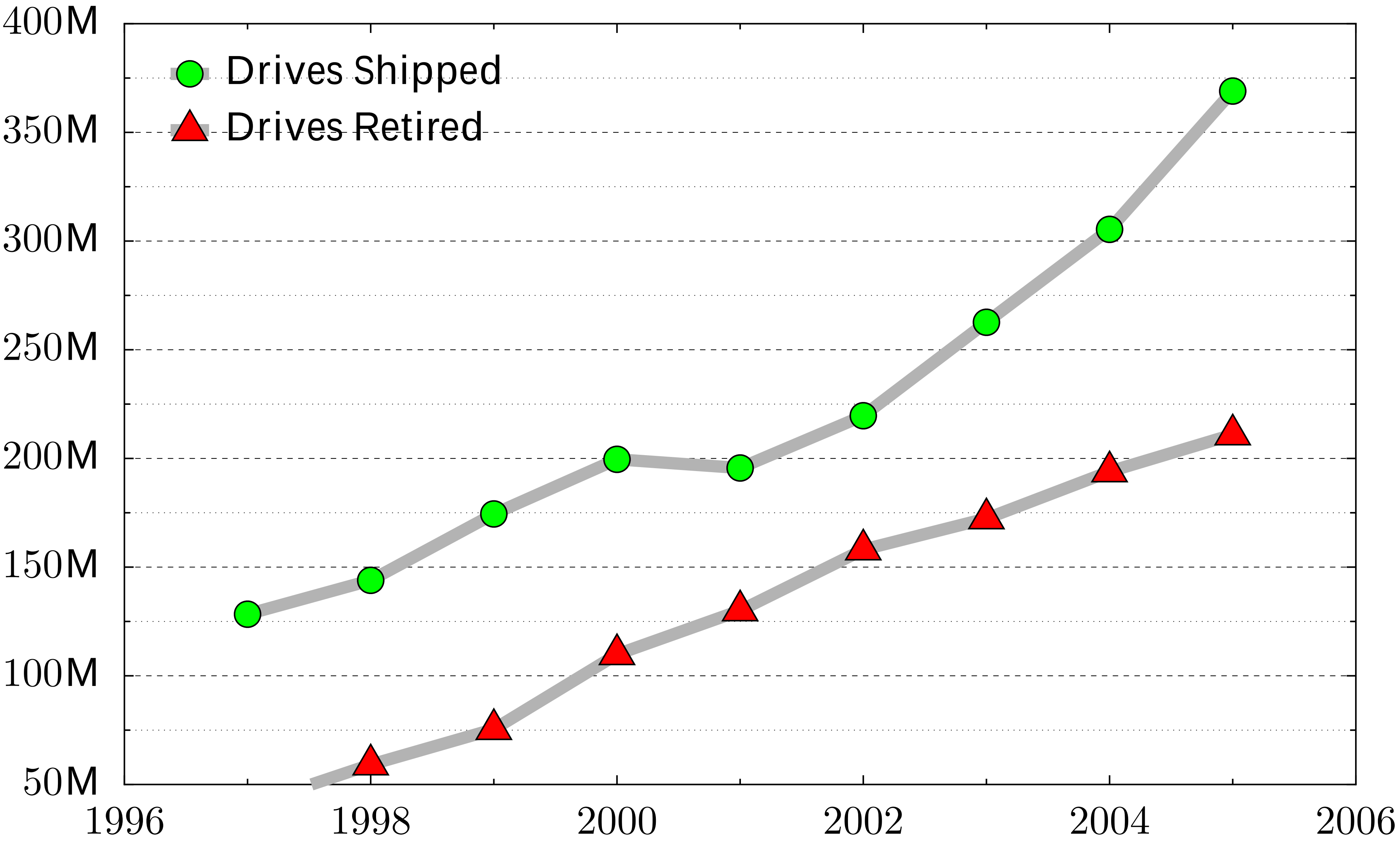
PDF file

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.

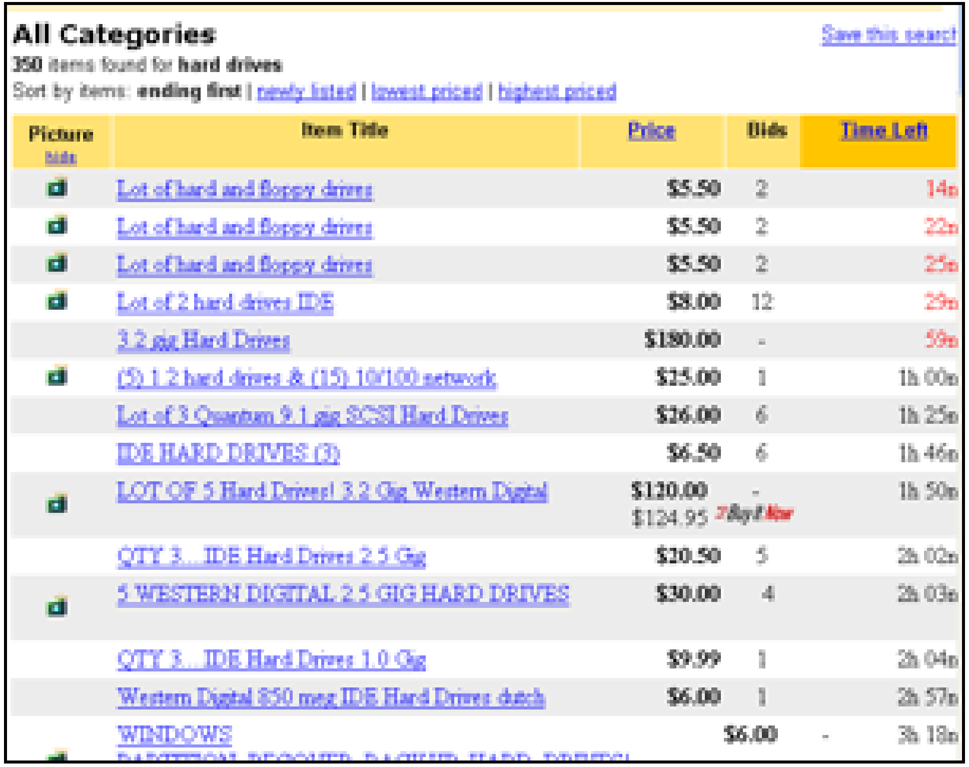
This is a huge problem:
210 million drives were retired in 2005!



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

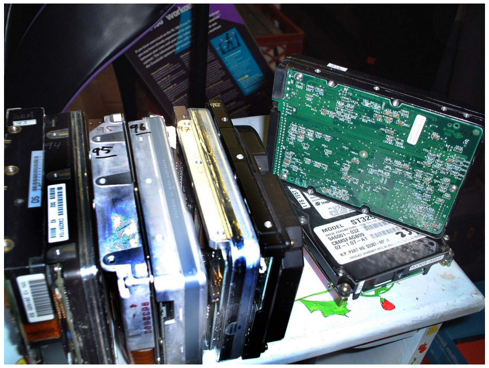
350 items found for **hard drives**

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.

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.

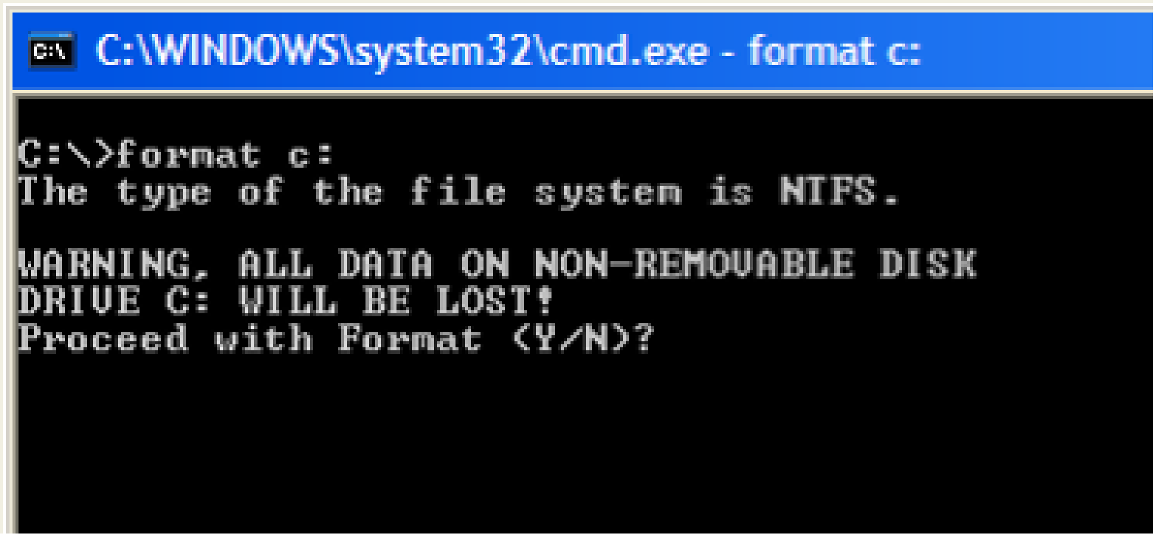
1. Physical security.
2. Logical access controls. (operating system)
3. Cryptography (disk & link)

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.

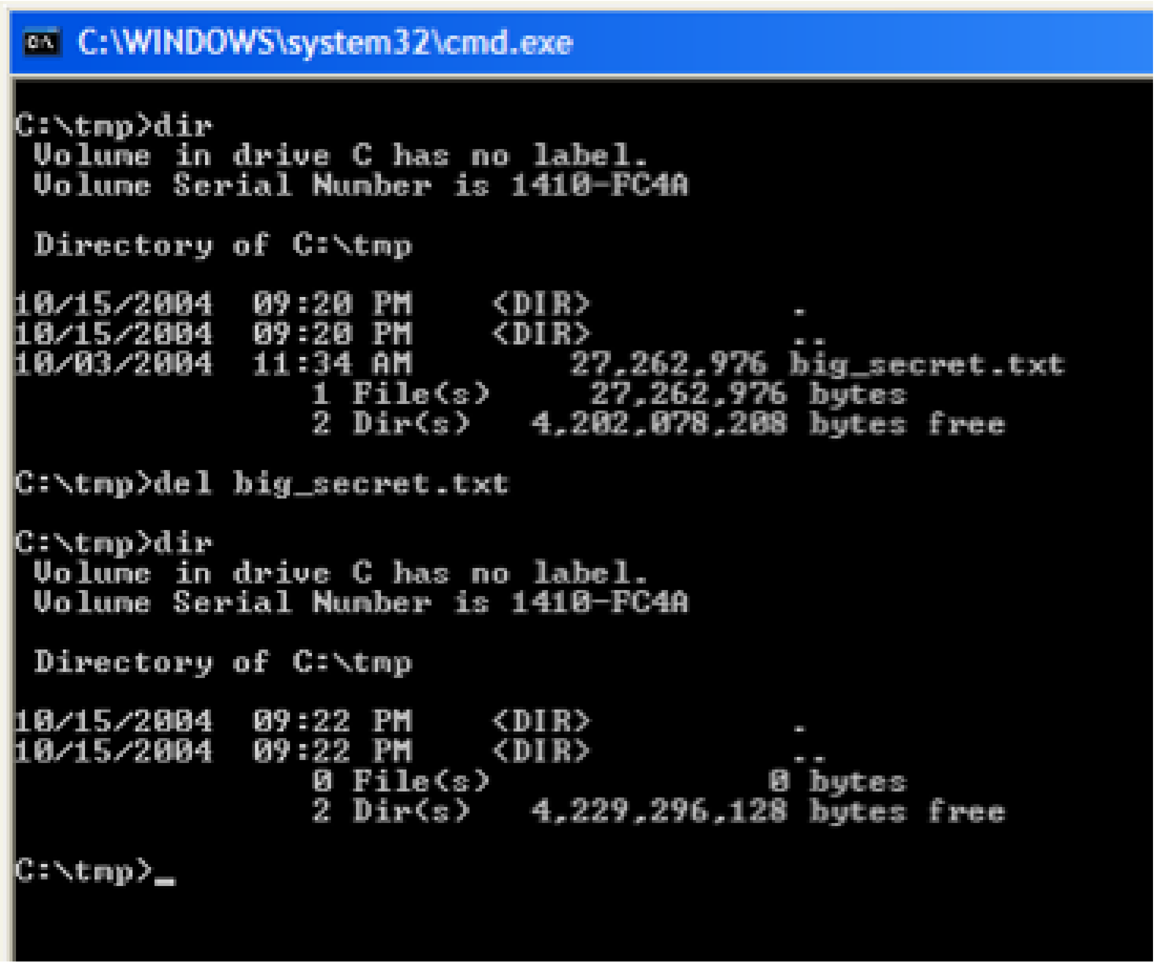


```
C:\>format c:
The type of the file system is NTFS.

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

 Directory of C:\tmp

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

C:\tmp>del big_secret.txt

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

 Directory of C:\tmp

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

C:\tmp>_
```

DEL simply removes the file’s name from the directory.

Drives arrive by UPS and USPS

Drives are imaged with `aimage` and stored in AFF format.

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

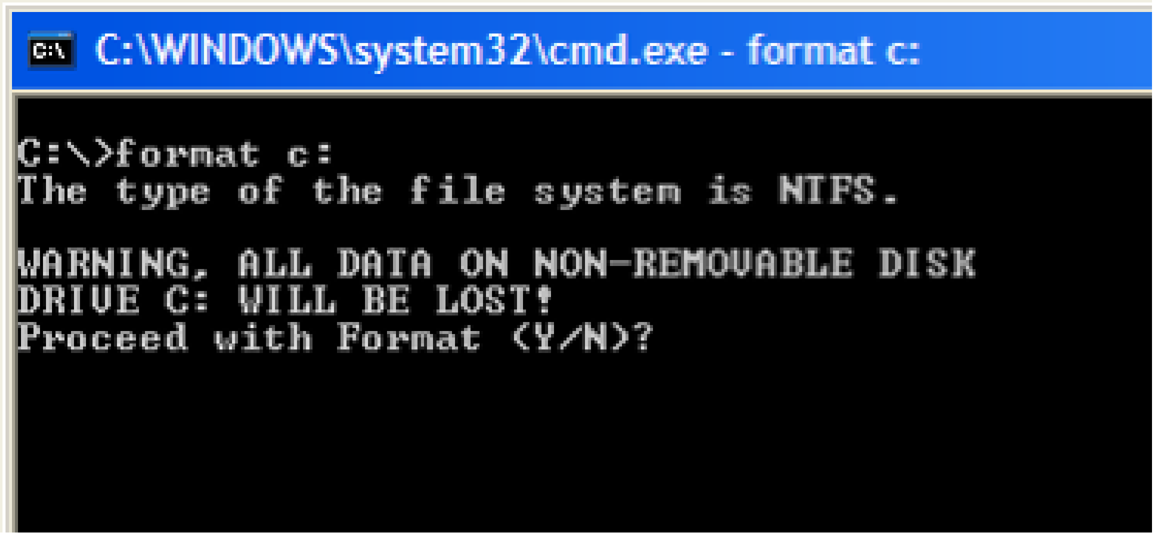
Initial analysis:

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:\>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)?
```

```
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
```

Windows FORMAT didn’t erase the disk...
FORMAT just wrote a new root directory.

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

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

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

70.img con't...

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

70.img con’t...

MAB-DEDUCTIBLE

MAB-MOOP

MAB-MOOP-DED

METHIMAZOLE

INSULIN (HUMAN)

COUMARIN ANTICOAGULANTS

CARBAMATE DERIVATIVES

AMANTADINE

MANNITOL

MAPROTILINE

CARBAMAZEPINE

CHLORPHENESIN CARBAMATE

ETHINAMATE

FORMALDEHYDE

MAFENIDE ACETATE

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:

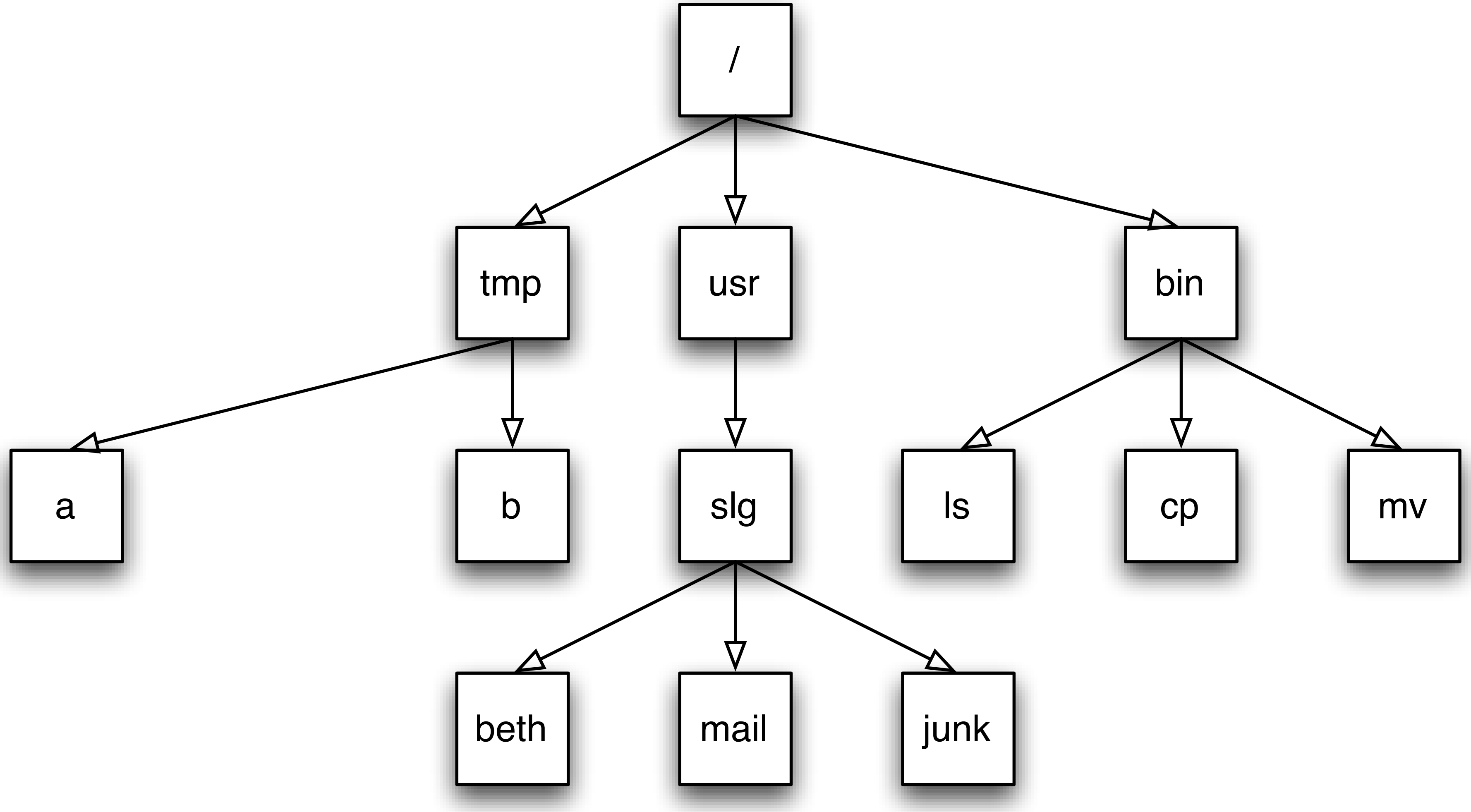
- Effective audit of information present on drives.
- Make DEL and FORMAT actually remove data. [Bauer & Priyantha 01]
- Provide alternative strategies for data recovery.

Education Problem:

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

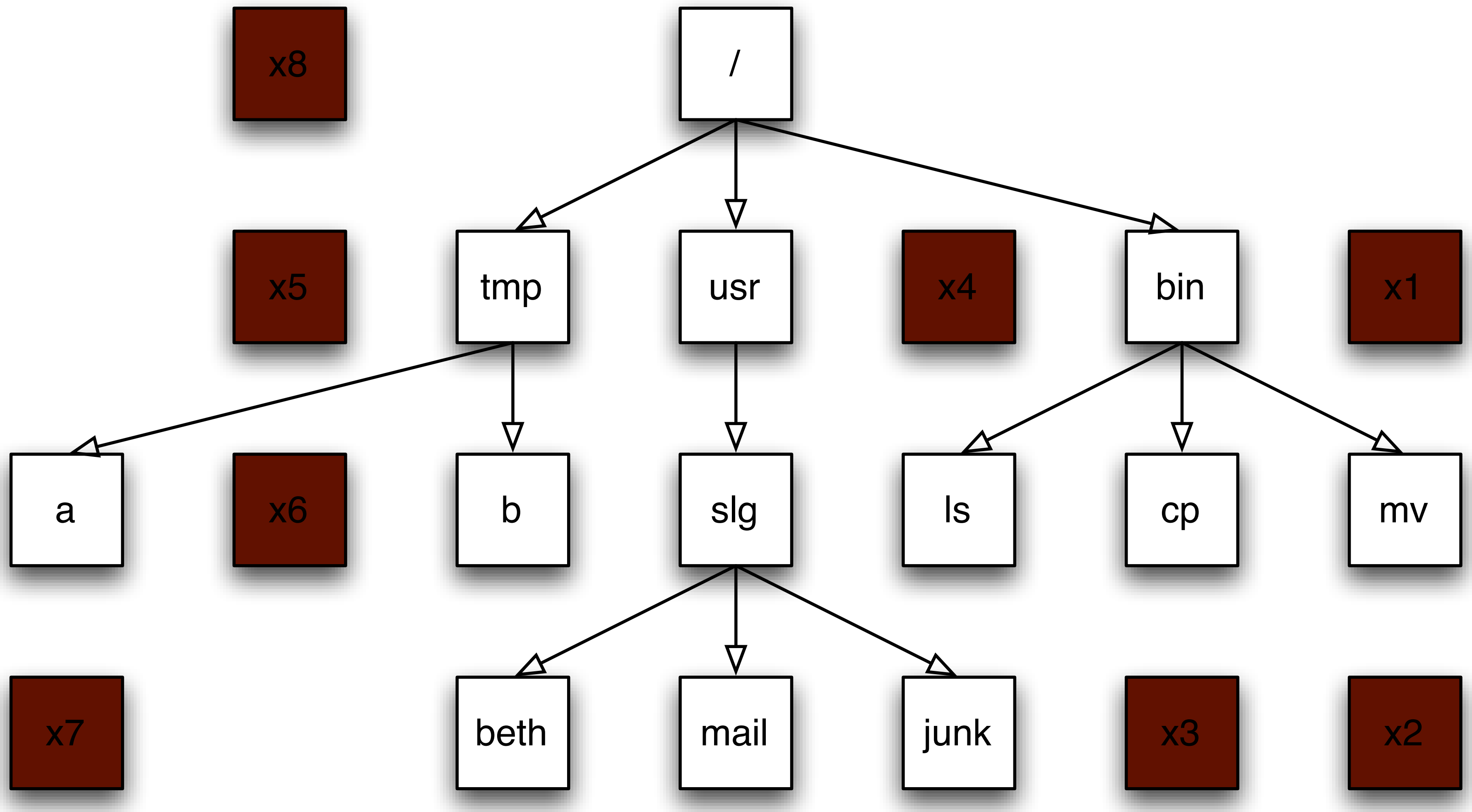
To find that cause,
I looked *on the drives* and *contacted the data subjects*.

Data on a hard drive is arranged in sectors.



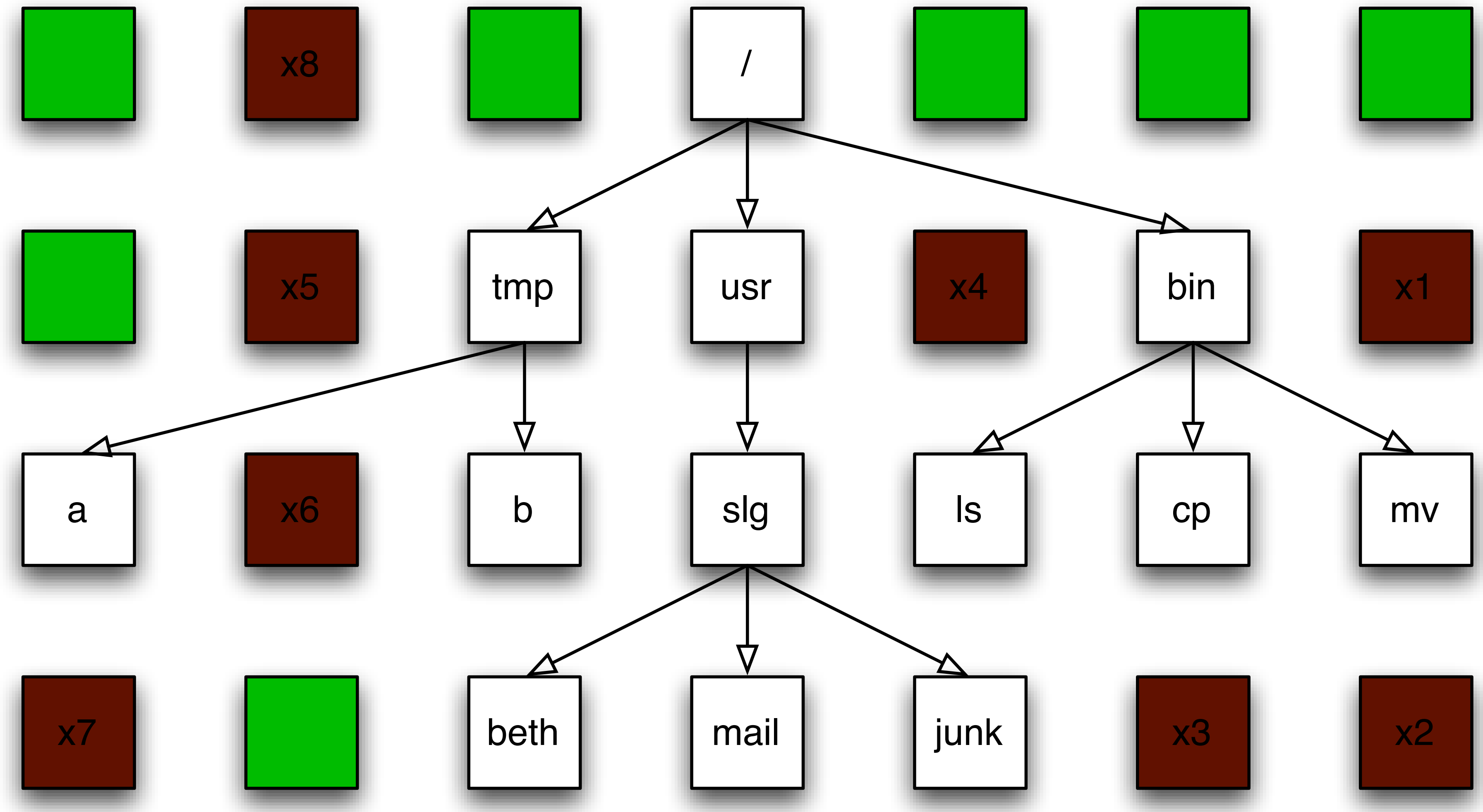
The white sectors indicate directories and files that are visible to the user.

Data on a hard drive is arranged in sectors.



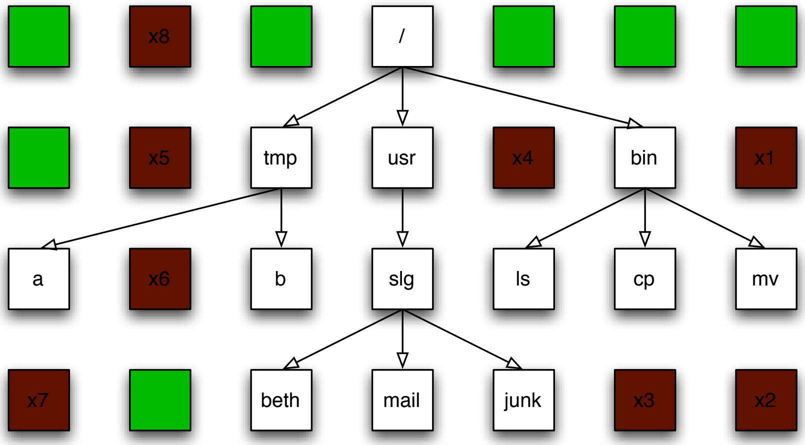
The brown sectors indicate files that were deleted.

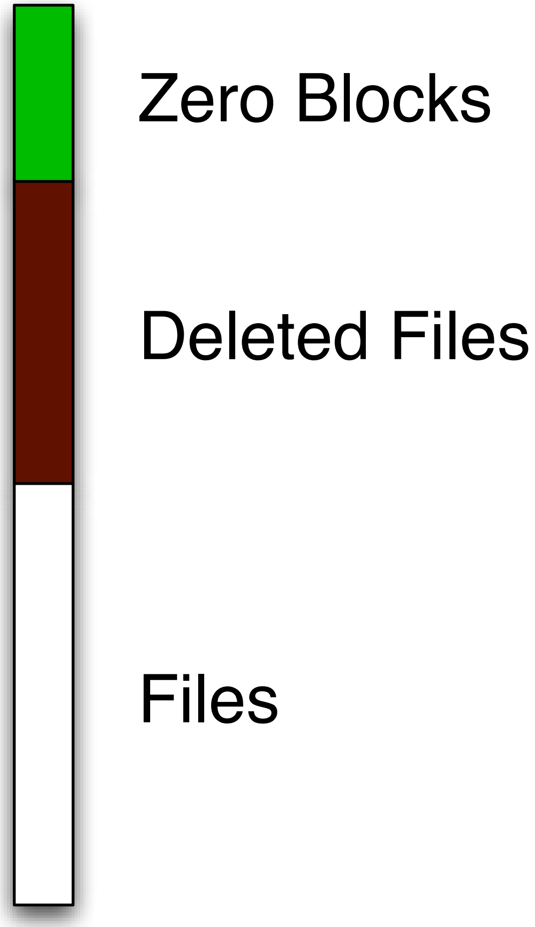
Data on a hard drive is arranged in sectors.



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

Stack the disk sectors:

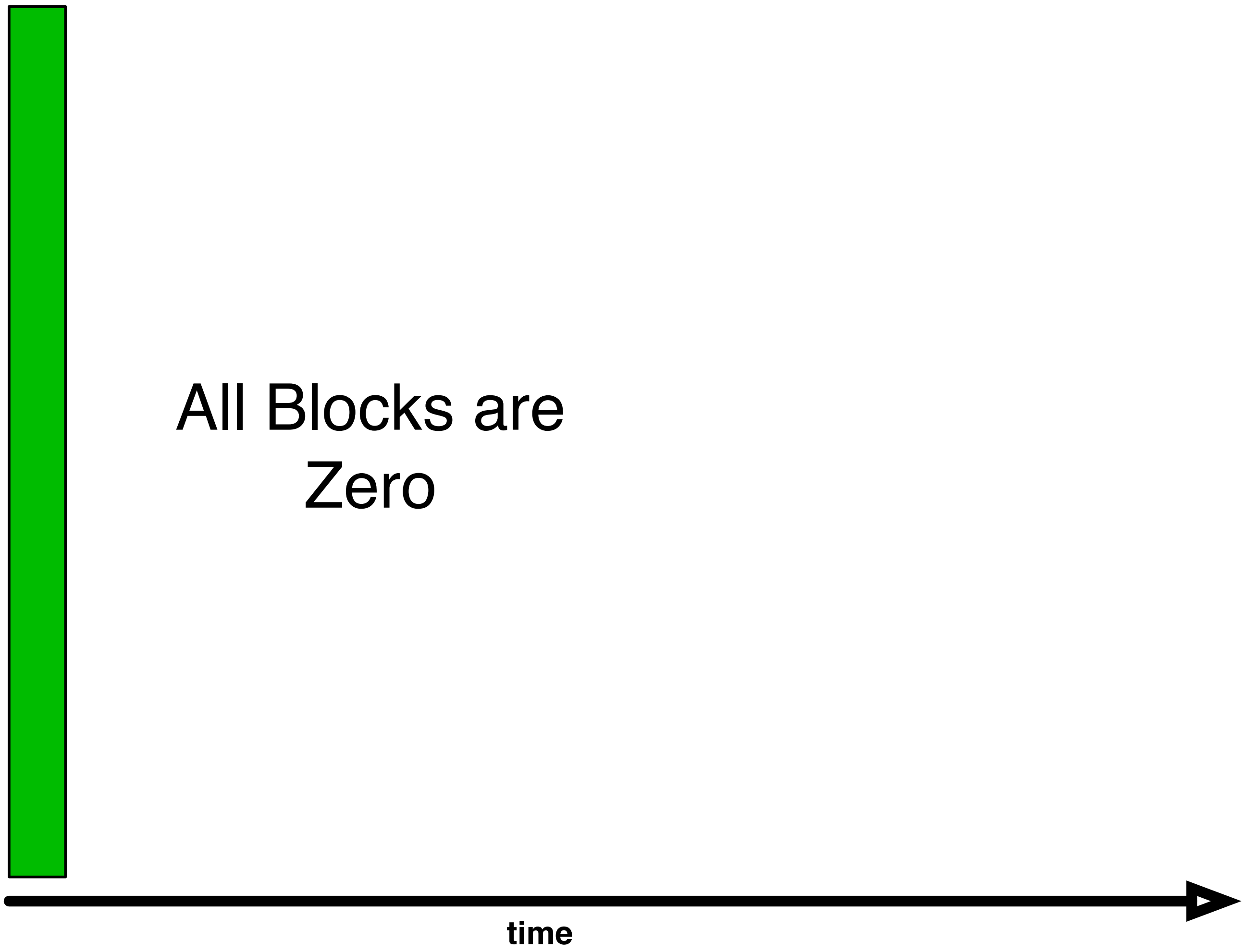



Zero Blocks

Deleted Files

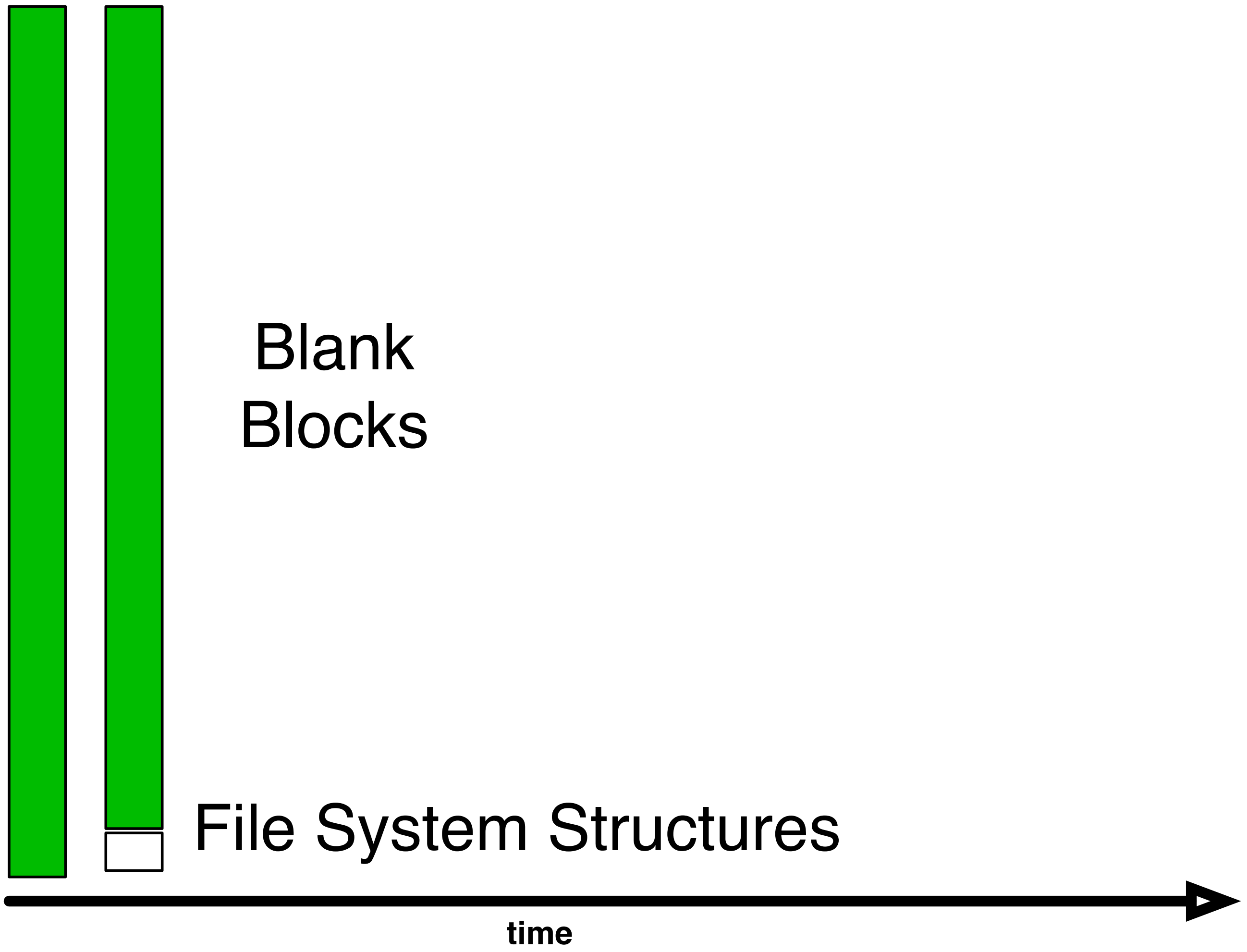
Files

NO DATA: The disk is factory fresh.

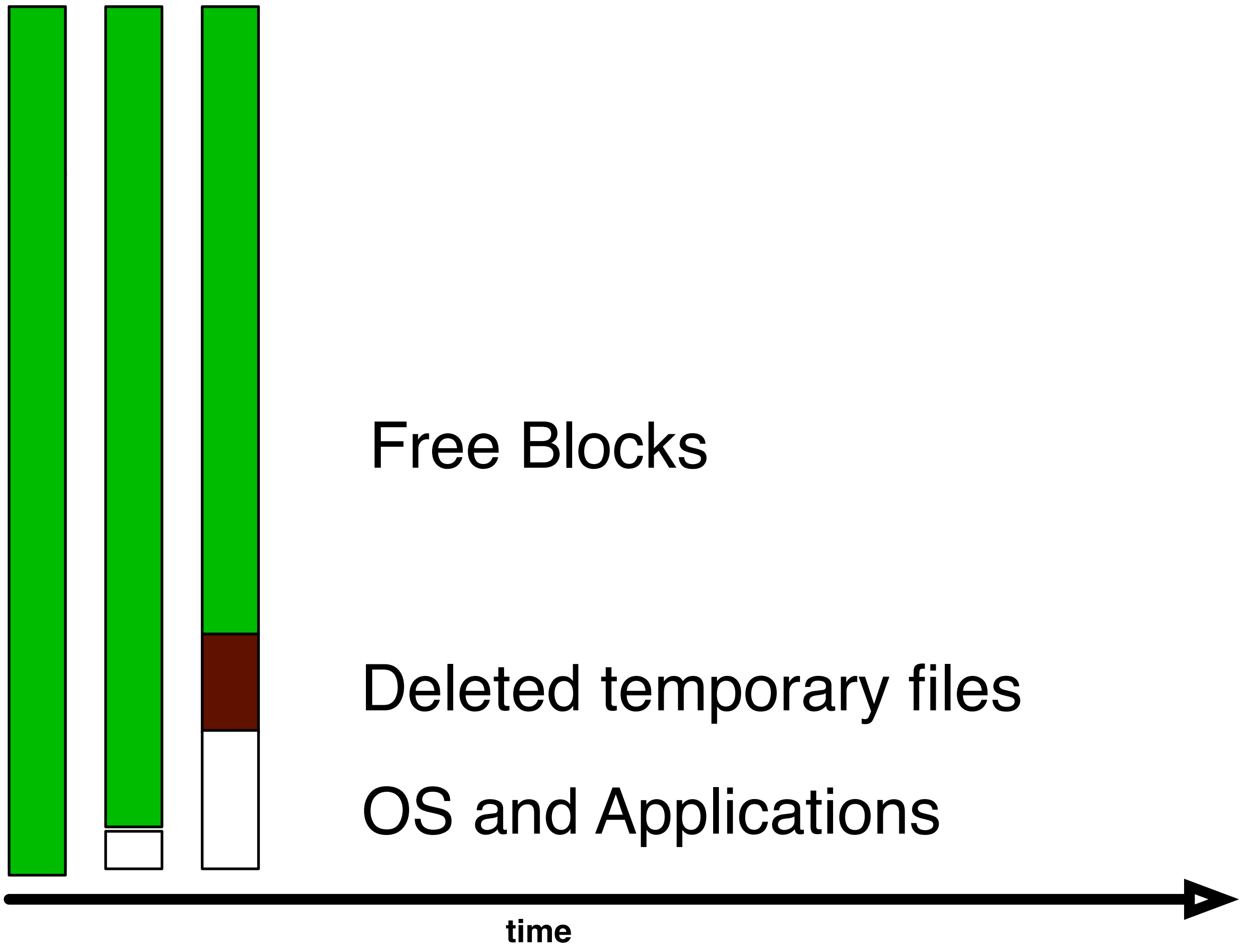


All Blocks are Zero

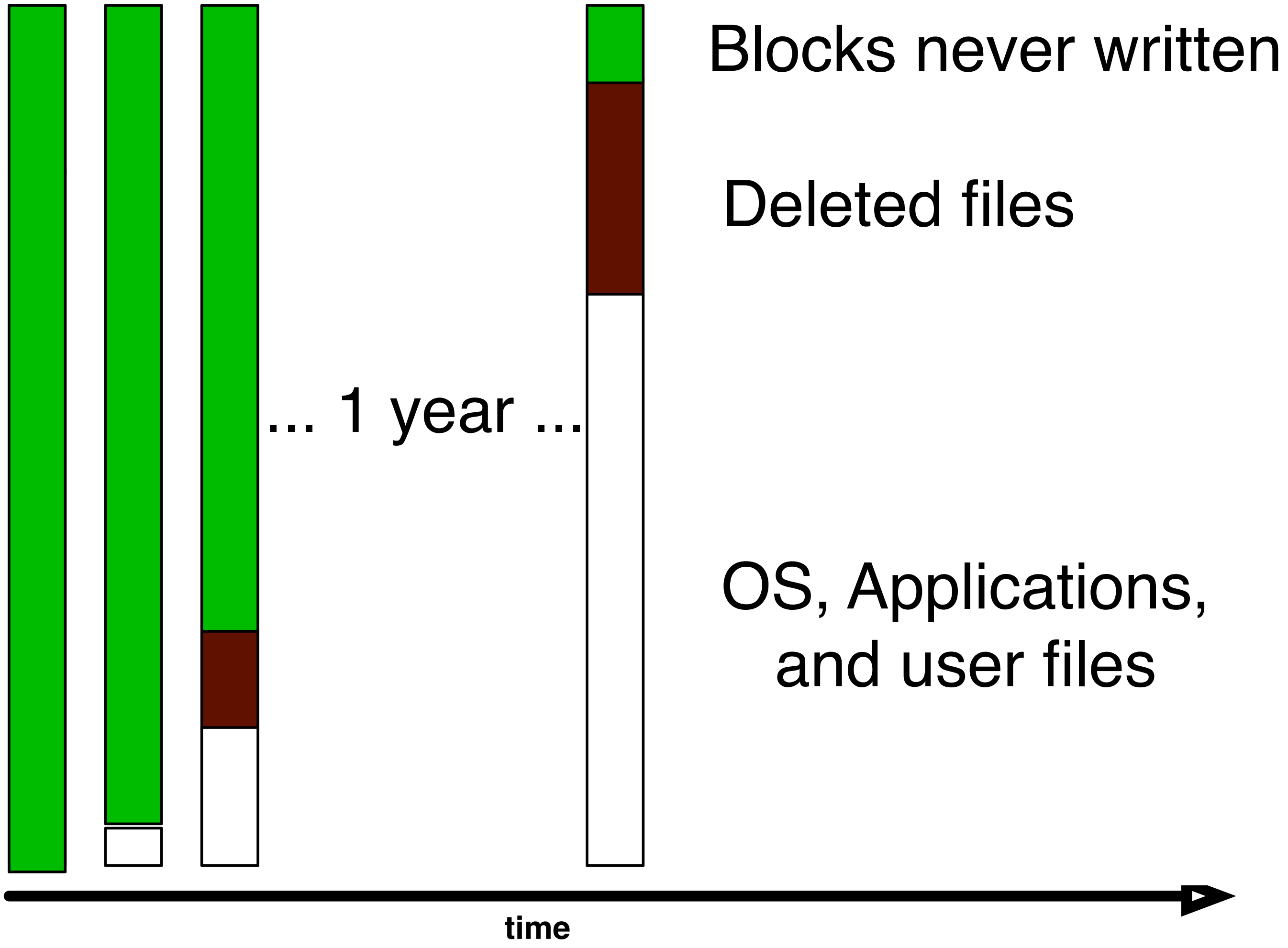
FORMATTED: The disk has an empty file system



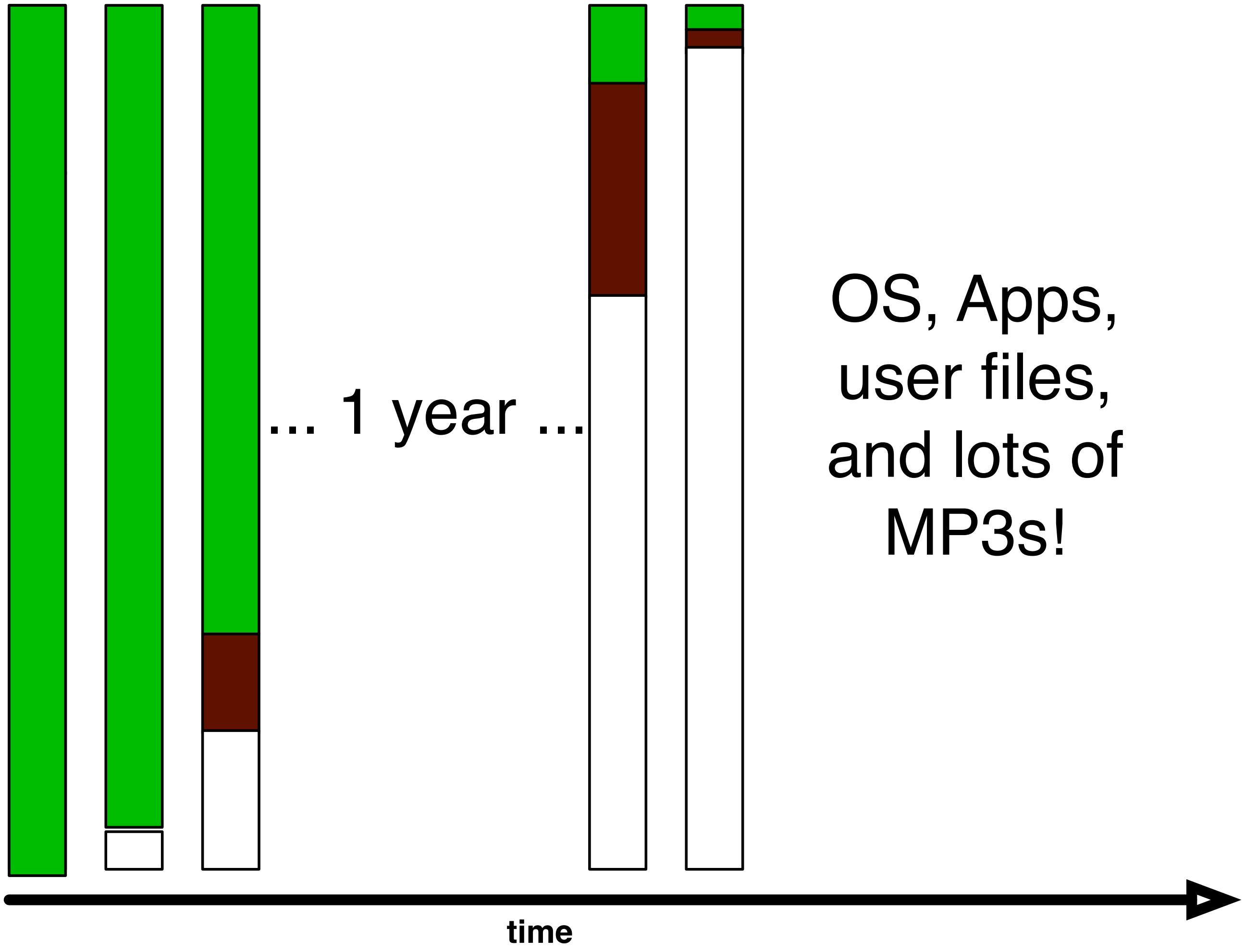
AFTER OS INSTALL: Temp. files have been deleted



AFTER A YEAR OF SERVICE

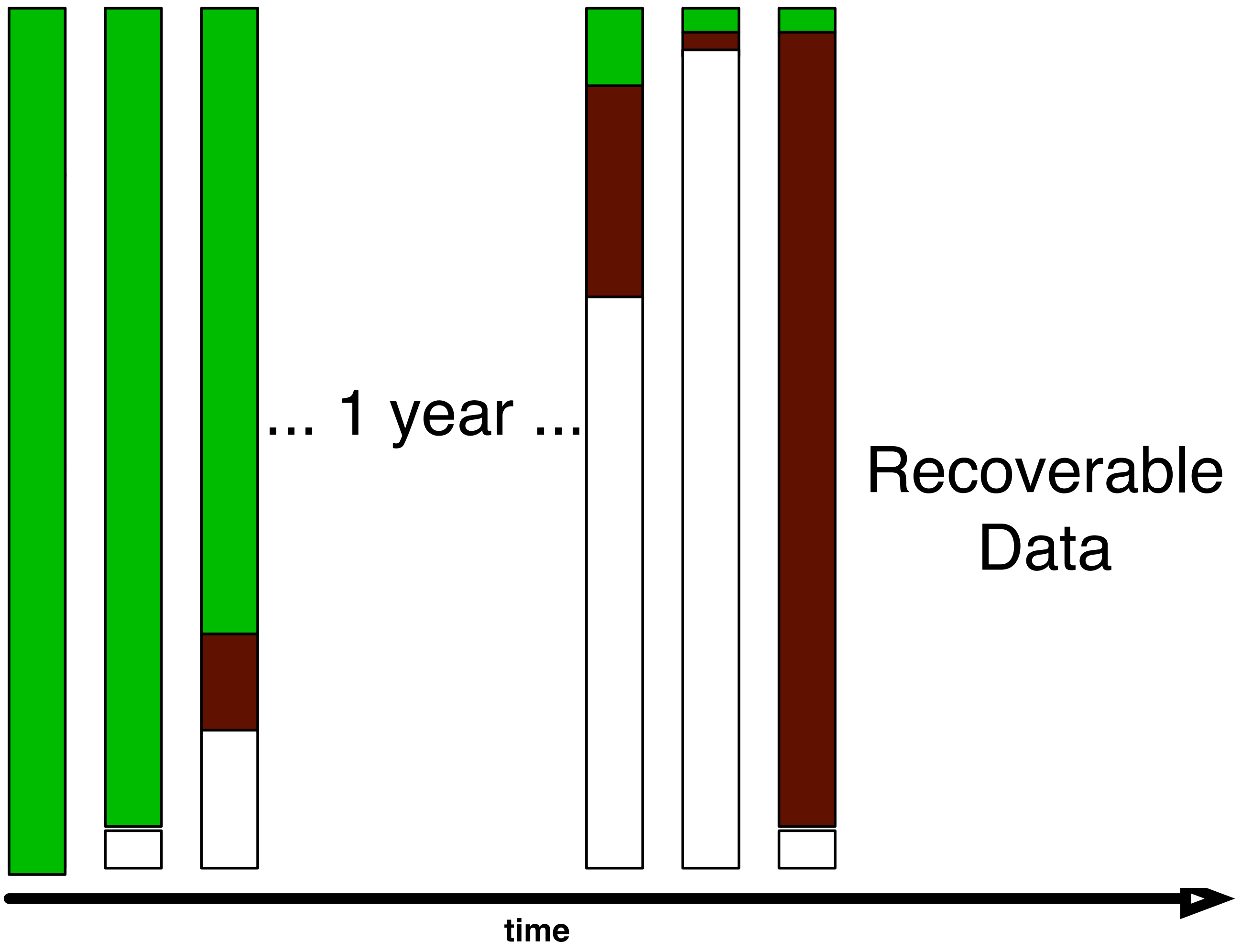


DISK NEARLY FULL!

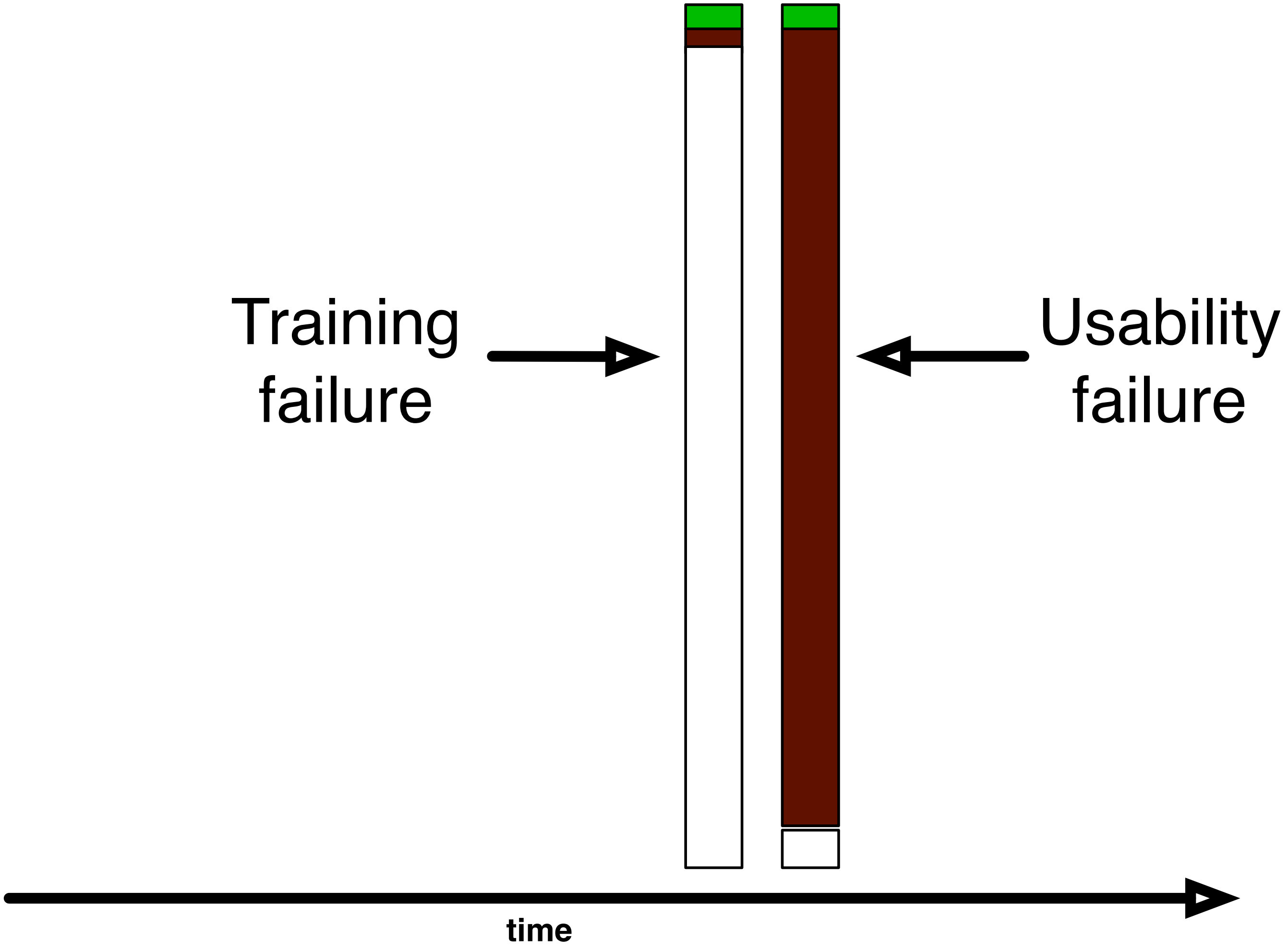


OS, Apps, user files, and lots of MP3s!

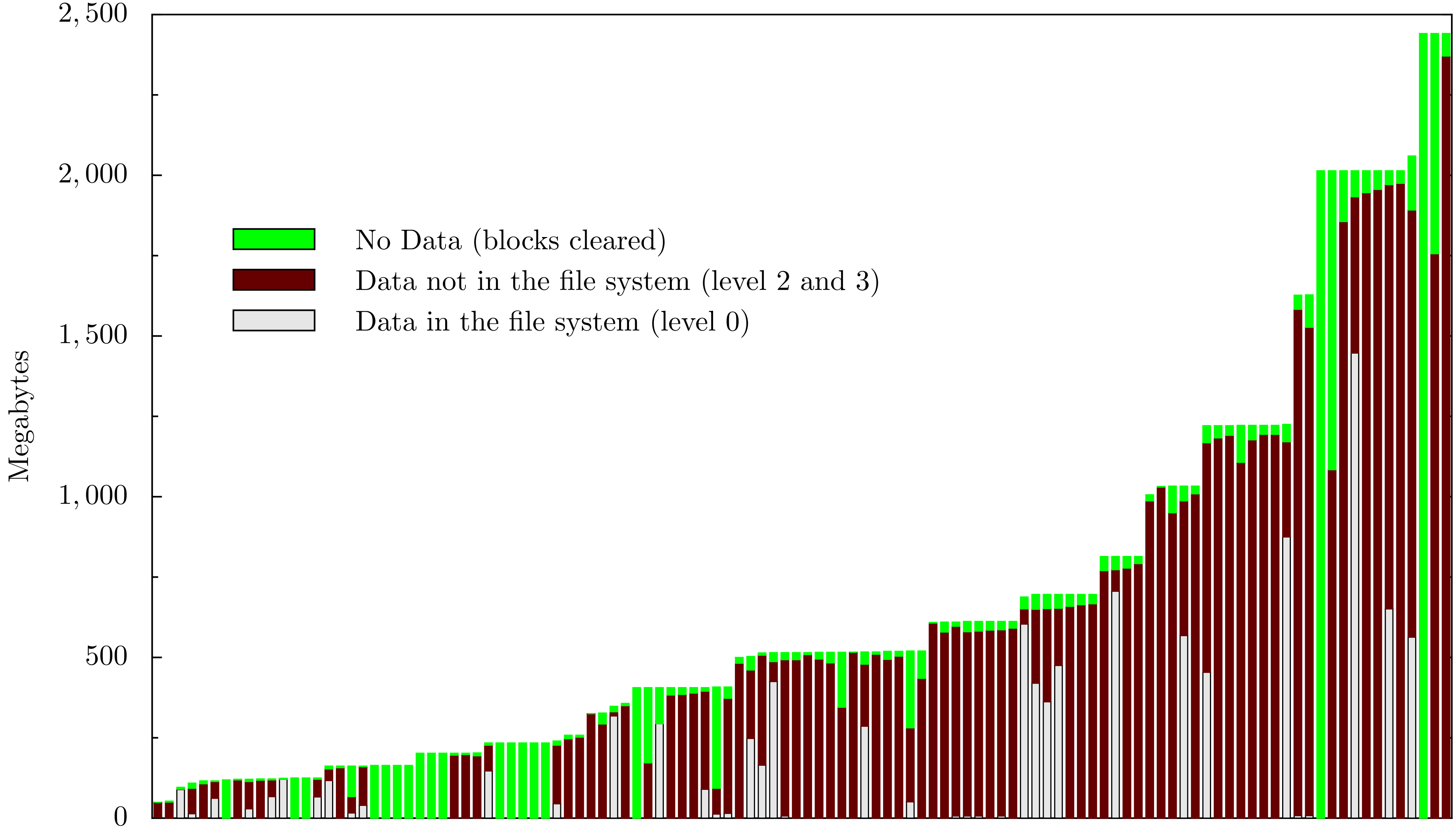
FORMAT C:\ (to sell the computer.)



We can use forensics to reconstruct motivations:



Drives 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.

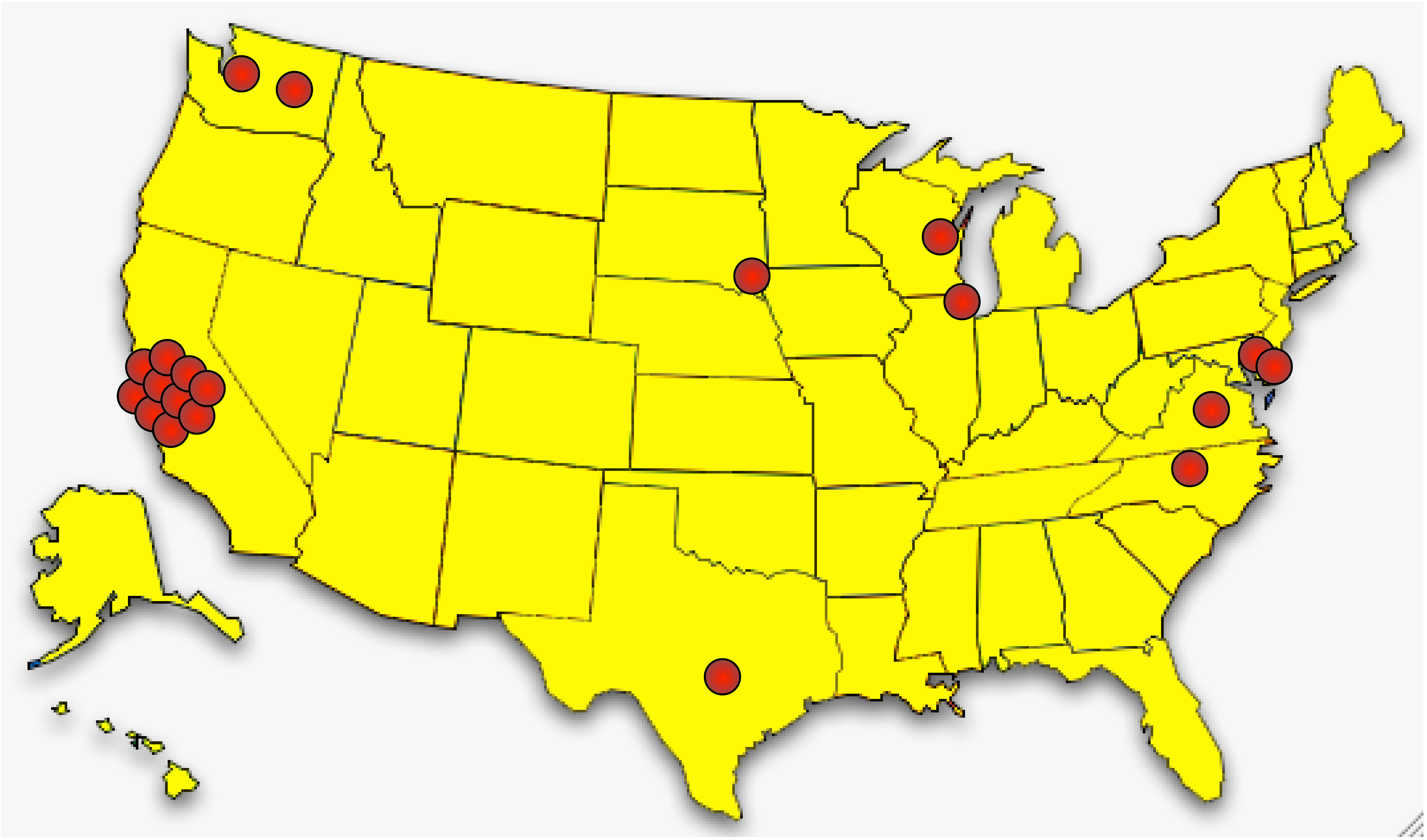
The *Remembrance of Data Passed Traceback Study.* [Garfinkel 05]

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.

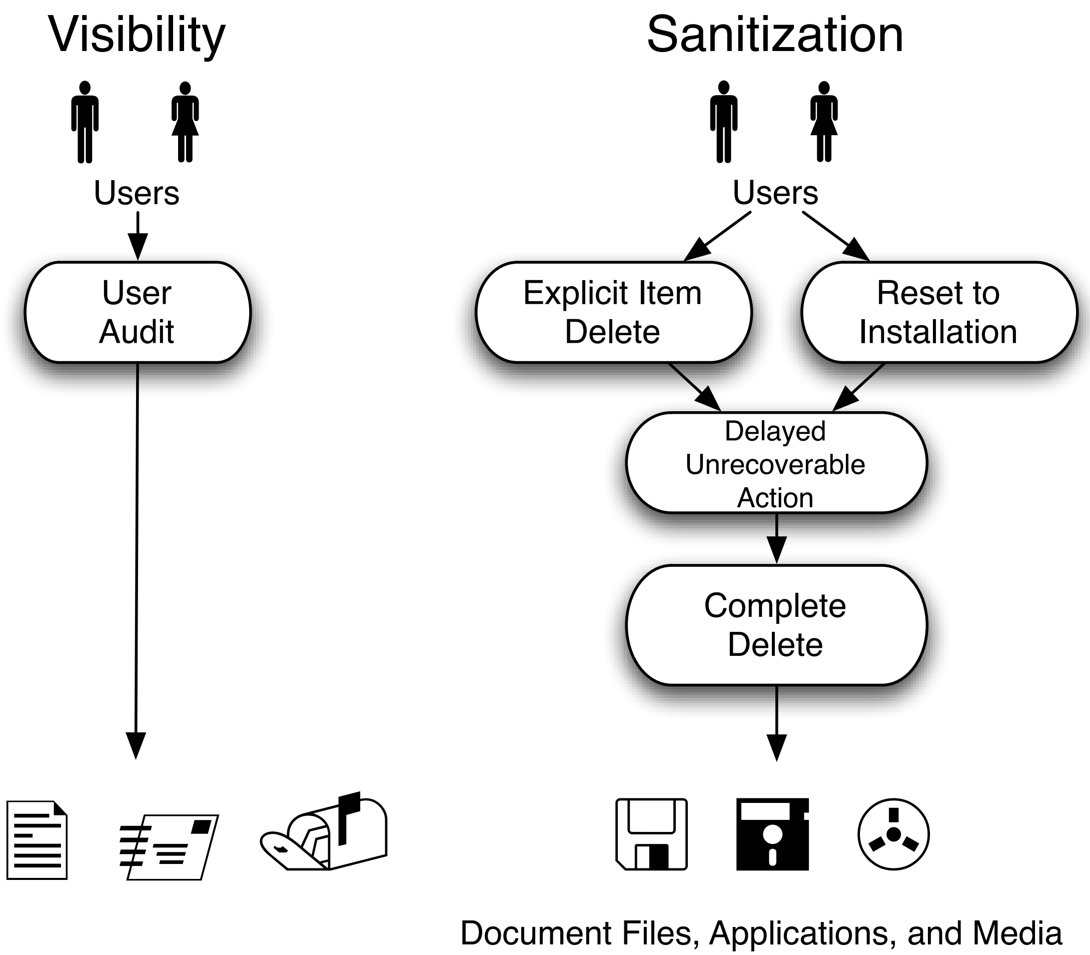
Trust Failure: 5 cases
Lack of Training: 3 cases
Lack of Concern: 2 cases

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.

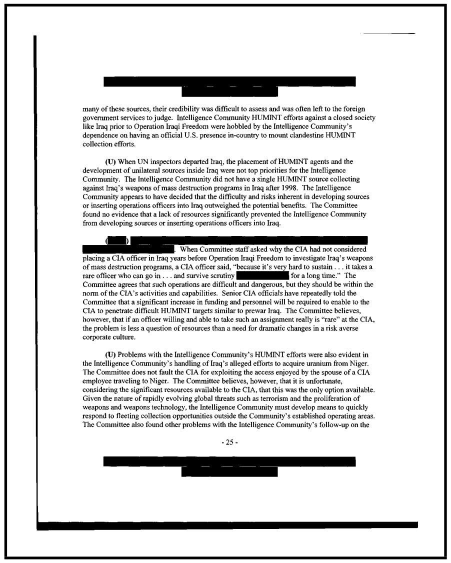
I have identified five distinct patterns for addressing the sanitization problem.



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



[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.

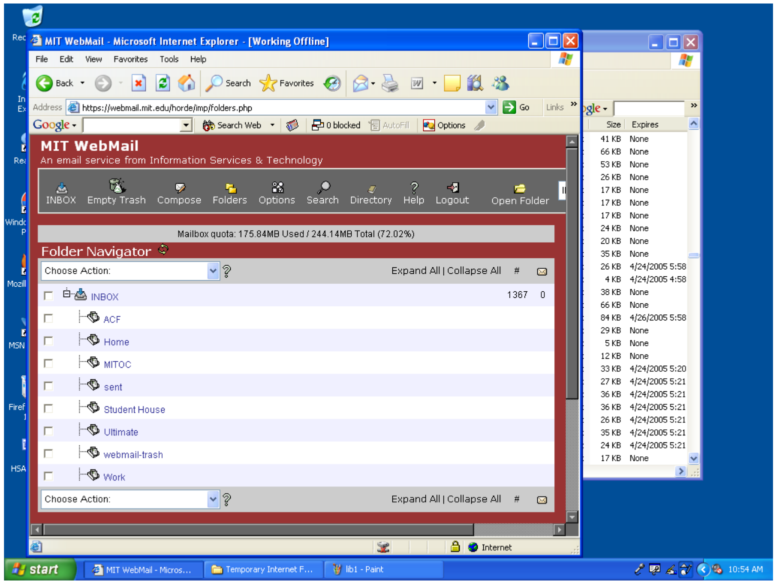
([redacted]) [redacted]. When Committee staff asked why the CIA had not considered placing a CIA officer in Iraq years before Operation Iraqi Freedom to investigate Iraq's weapons of mass destruction programs, a CIA officer said, "because it's very hard to sustain . . . it takes a rare officer who can 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



MIT WebMail - Microsoft Internet Explorer - [Working Offline]

File Edit View Favorites Tools Help

Back Search Favorites

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

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

Folder Navigator

Choose Action: Expand All | Collapse All #

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

Choose Action: Expand All | Collapse All #

Internet

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

start MIT WebMail - Micros... Temporary Internet F... lib1 - Paint 10:54 AM

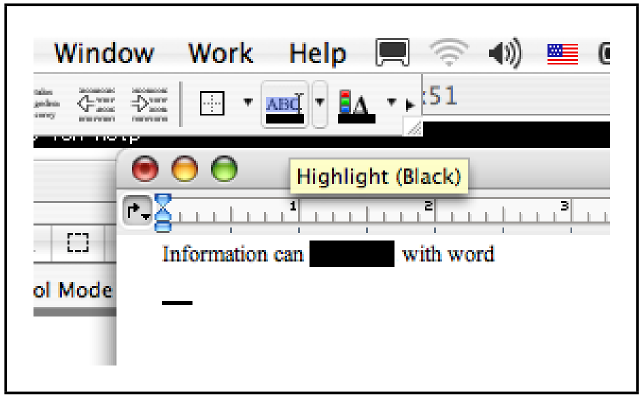
Information is left in document files.

- 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]

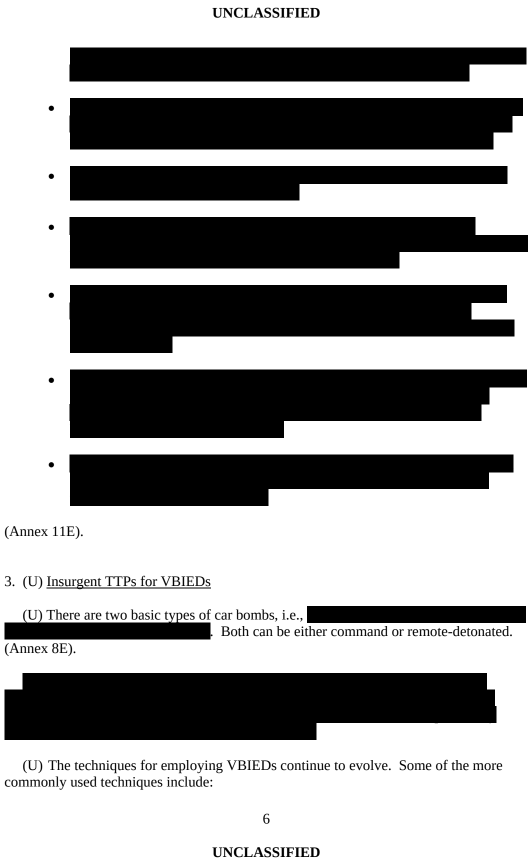
- SCO gave a **Microsoft Word file** to journalists that revealed its Linux legal strategy. [Shankland 04]

- Multinational forces in Iraq published classified information about insurgency methods.



Information can [redacted] with word

Word’s highlight feature is literally a threat to national security.



UNCLASSIFIED

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

(Annex 11E).

3. (U) Insurgent TTPs for VBIEDs

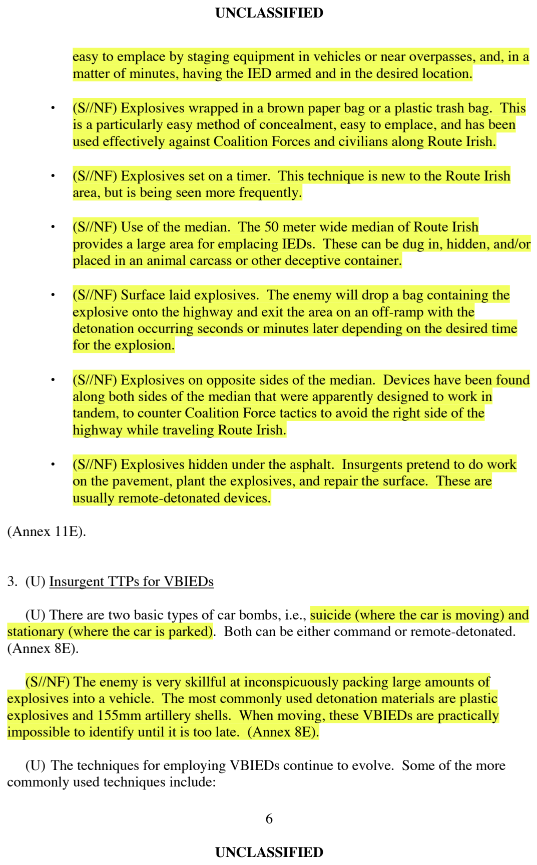
(U) There are two basic types of car bombs, i.e., [redacted]. Both can be either command or remote-detonated. (Annex 8E).

[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).

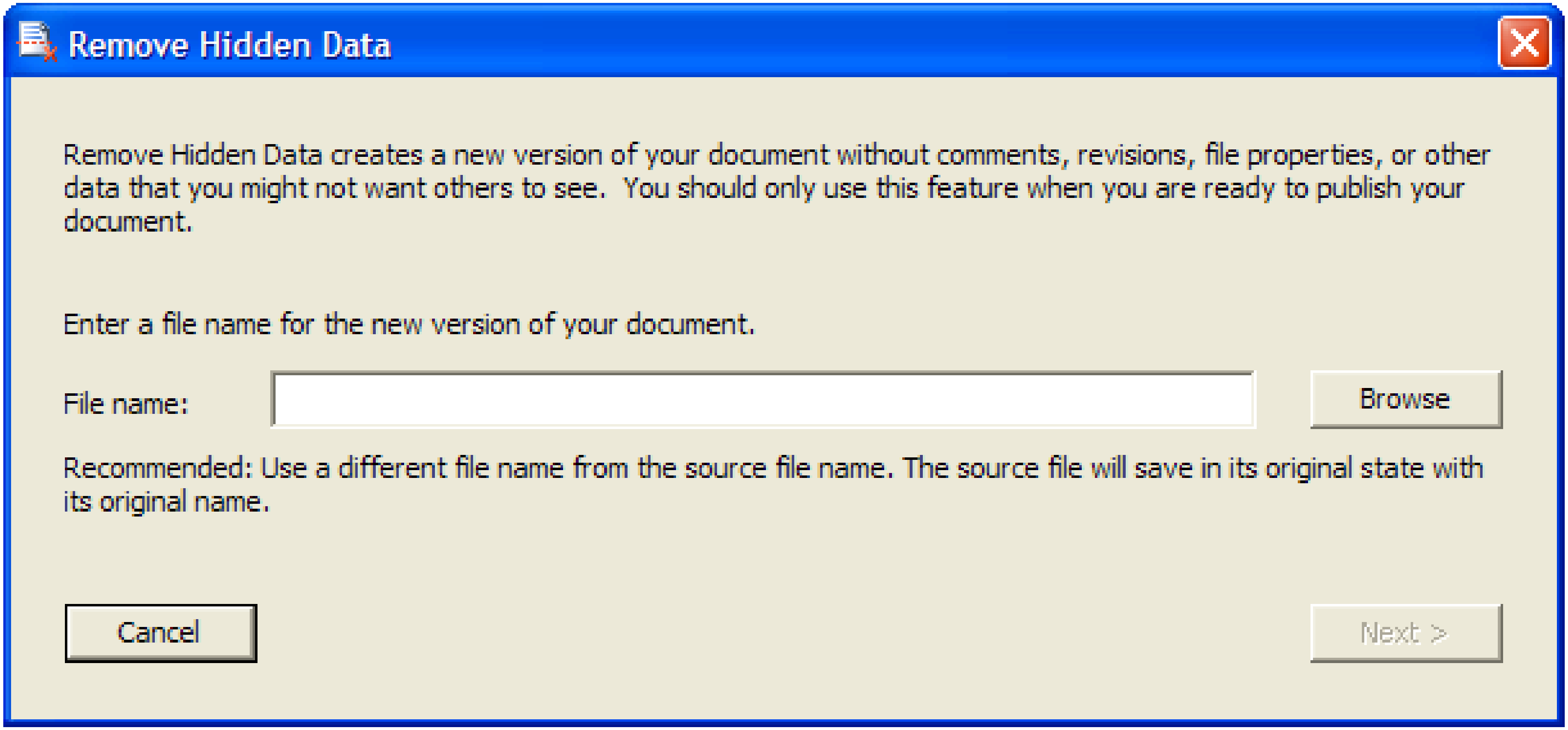
(U) The techniques for employing VBIEDs continue to evolve. Some of the more commonly used techniques include:

6

UNCLASSIFIED

NSA recently published a “how to sanitize” guide.

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



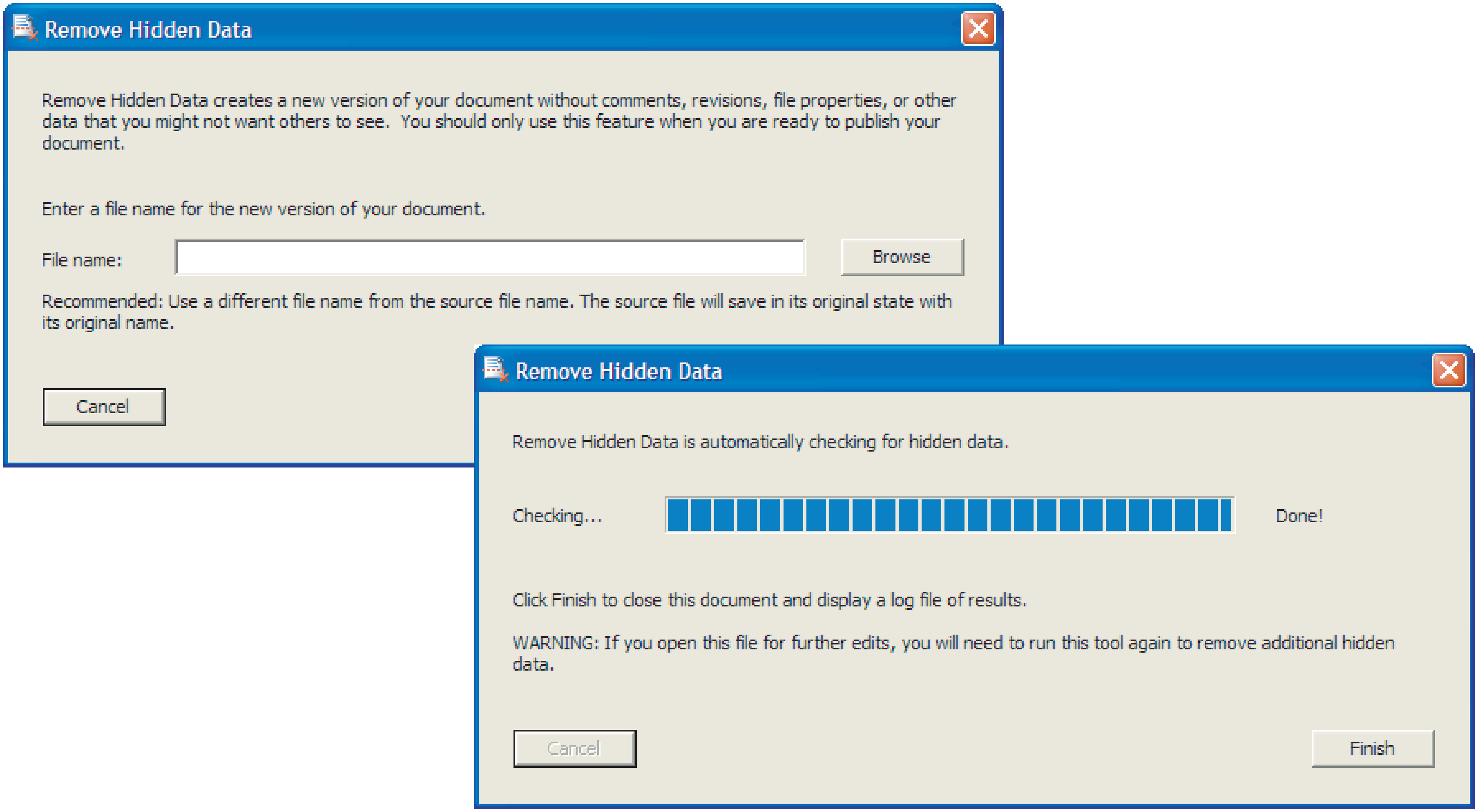
Remove Hidden Data creates a new version of your document without comments, revisions, file properties, or other data that you might not want others to see. You should only use this feature when you are ready to publish your document.

Enter a file name for the new version of your document.

File name:

Recommended: Use a different file name from the source file name. The source file will save in its original state with its original name.

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 of your document.

File name:

Recommended: Use a different file name from the source file name. The source file will save in its original state with its original name.

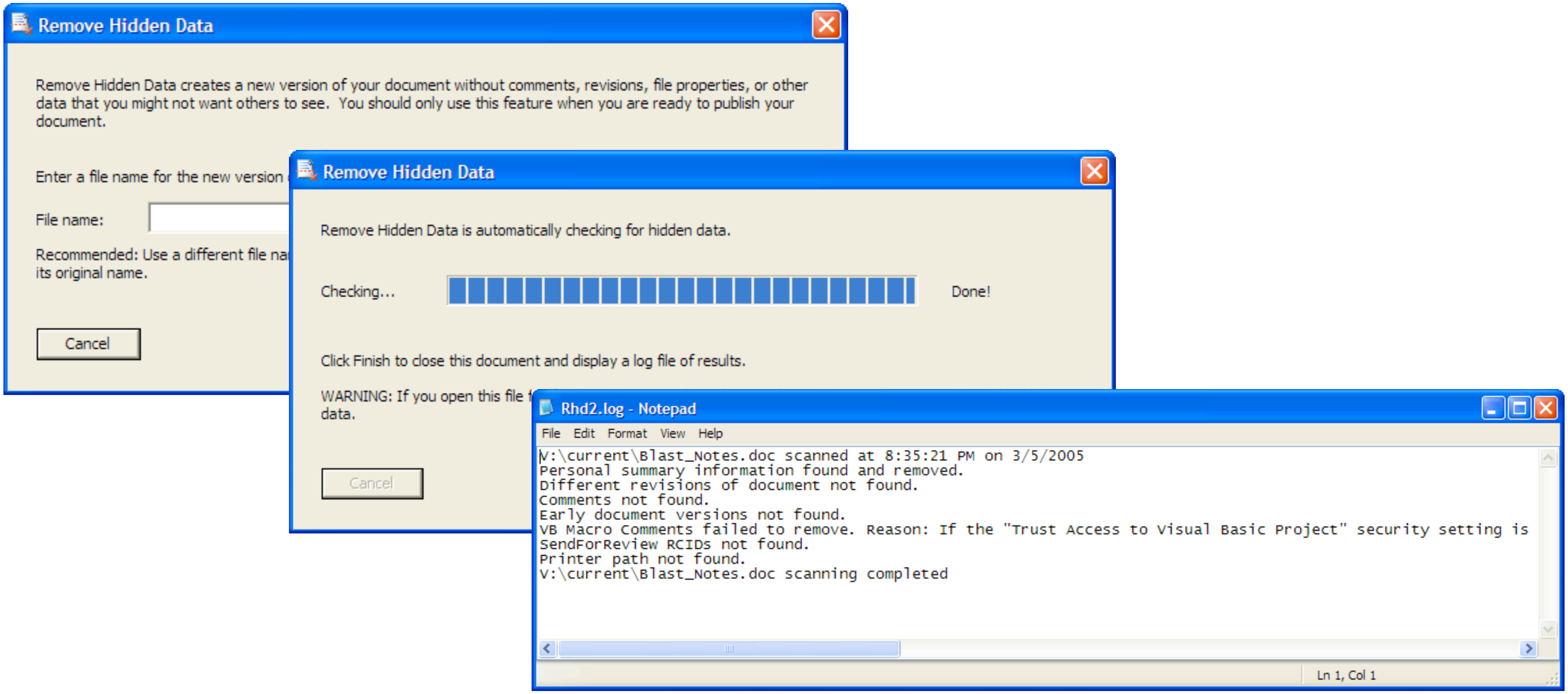
Remove Hidden Data is automatically checking for hidden data.

Checking... Done!

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

WARNING: If you open this file for further edits, you will need to run this tool again to remove additional hidden data.

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.

Checking... Done!

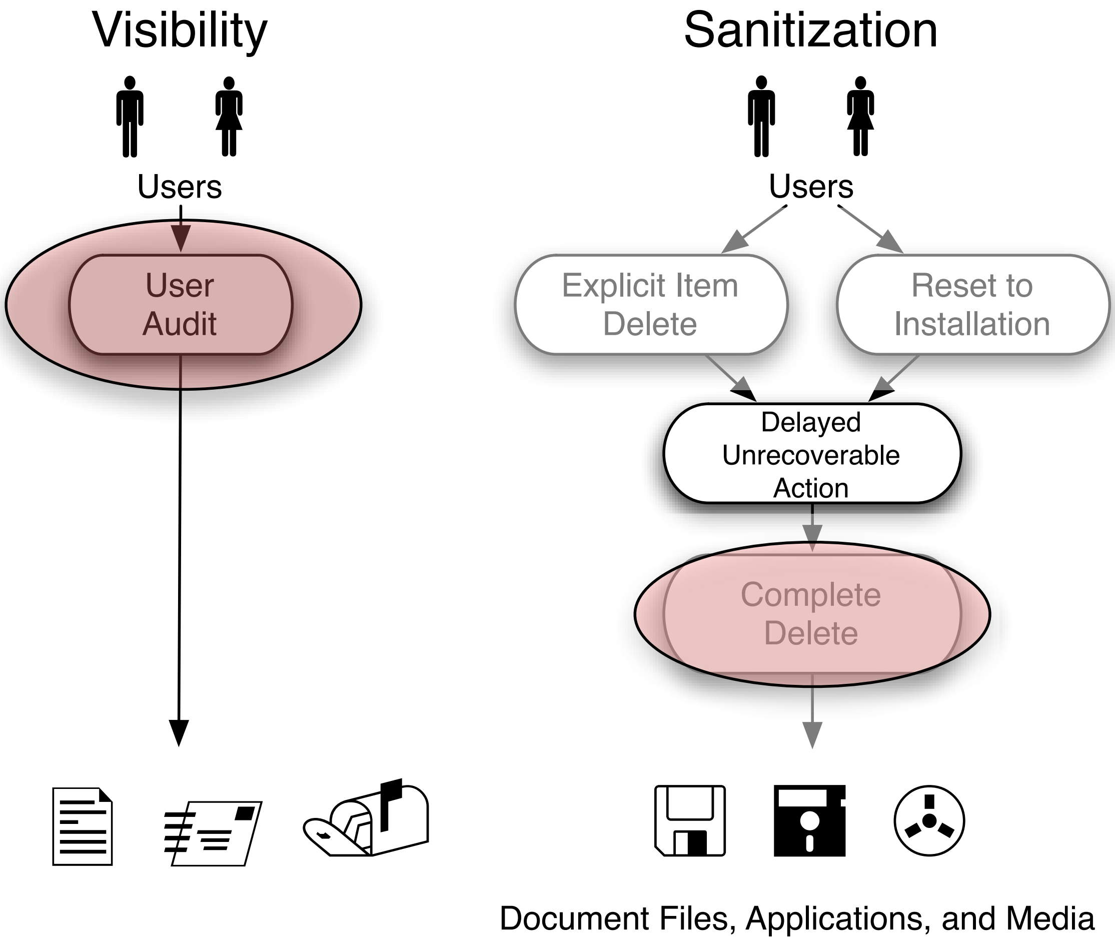
Click Finish to close this document and display a log file of results.

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
```

My patterns predict that Microsoft’s tool will fail.

The information leaks because two patterns were not implemented.

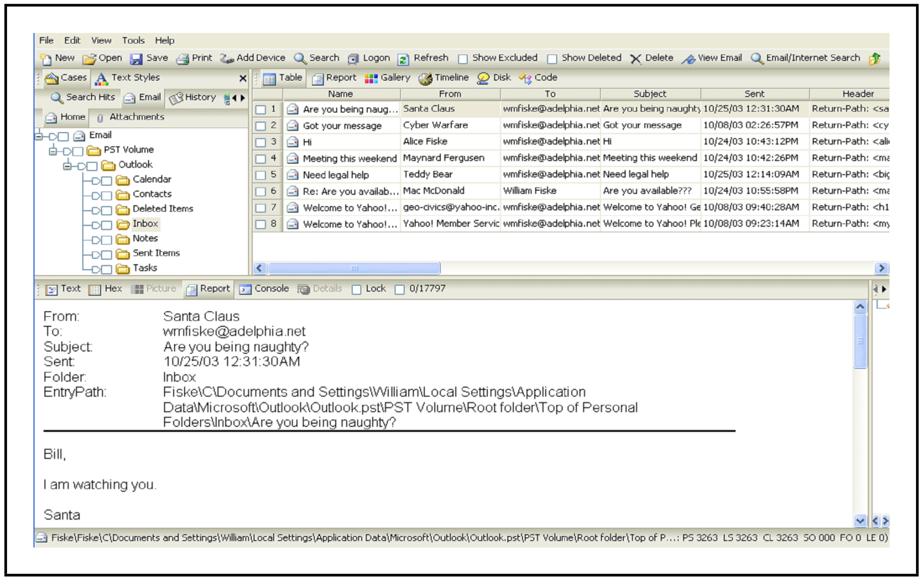


Current agenda: getting vendors to implement these patterns.

Cross Drive Analysis:
Applying to tools of [Garfinkel ’05] to computer forensics.

Today’s forensics tools:

- Interactive user interface.
- Recovery of “deleted” files.
- Child porn scanning.
- Trial preparation.
- Focus on one disk at a time.


File Edit View Tools Help
New Open Save Print Add Device Search Logon Refresh
□ Show Excluded
□ Show Deleted
Delete View Email Email/Internet Search
Cases Text Styles
Table Report Gallery Timeline Disk Code
Search Hits Email History
Home Attachments
Email
PST Volume
Outlook
Calendar
Contacts
Deleted Items
Inbox
Notes
Sent Items
Tasks
Text Hex Picture Report Console Details
□ Lock
□ 0/17797

	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 of disks.

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

But a large police department or small intelligence mission can generate thousands of disks...

New intelligence capabilities can be enabled by correlating information from multiple drives.

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

Single-Drive Statistical Techniques

Example problem: Who owned this disk drive?

Approach #1: Find Microsoft Word files; determine 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.

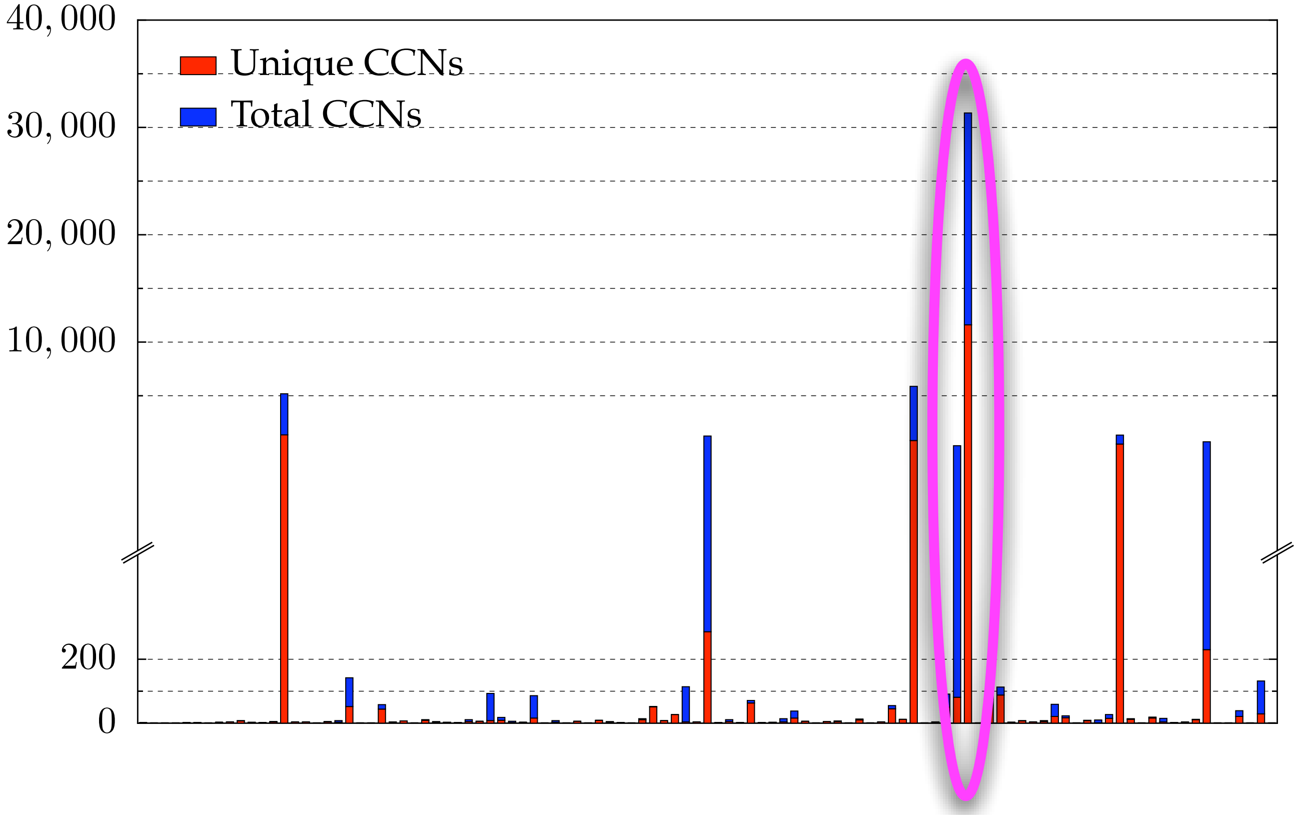
The email histogram approach works quite well.

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?)

“First Order Cross-Drive Analysis” 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 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 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
```

Even with the tests, there are occasional false positives.

CCN scan of Disk #115: (772MB)

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

```
.................@:|4444448666108|:<@<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
%.....}}}}}}.......|4444322233345|.....}}}}}}......   pos=228610295
%6"!) .&*%,,%-0)07.|373484553420378|<67<038+.5(+0+.3.  pos=638491849
%6"!) .&*%,,%-0)07.|373484553420378|<67<038+.5(+0+.3.  pos=645913801
```

Results of scanning 2003 corpus with CCN scanner:

Total number of image files:	178
Number of CCNs found:	47,771
Total number of distinct cards:	15,613
Most popular CCN	6404 6521 6029 6650 (Seen 34 times on 30 drives)

Context analysis shows this is not a valid CCN:

```
 [6]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
 [7]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
 [8]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
[10]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
[11]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
[11]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
[15]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
[18]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
[18]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
[24]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
[25]  6213 l 6758 6367 ..|6404 6521 6029 6650| v 6025 6646 l -138
```

A “stop list” can be used for these common number.

Ignore “6404 6521 6029 6650’ and we repeat the experiment:

Total number of image files:	178	
Number of CCNs found:	47,737	(was 47,771)
Total number of distinct cards	15,612	(was 15,613)
New “most popular CCN”	5501 8501 3501 3705	
	(Seen 35 times on 27 drives)	

Once again, this does not appear to be a valid CCN:

```
 [14] 3201 4901  : |5501 8501 3501 3705| 5102....yes.%d\0ff
[112] 3201 4901  : |5501 8501 3501 3705| 5102....yes.%d\0ff
[121] 3201 4901  : |5501 8501 3501 3705| 5102....yes.%d\0ff
[128] 3201 4901  : |5501 8501 3501 3705| 5102....yes.%d\0ff
[133] 3201 4901  : |5501 8501 3501 3705| 5102....yes.%d\0ff
[181] 3201 4901  : |5501 8501 3501 3705| 5102....yes.%d\0ff
[182] 3201 4901  : |5501 8501 3501 3705| 5102 13505....yes.
[184] 3201 4901  : |5501 8501 3501 3705| 5102 13505....yes.
[186] 3201 4901  : |5501 8501 3501 3705| 5102 13505....yes.
```

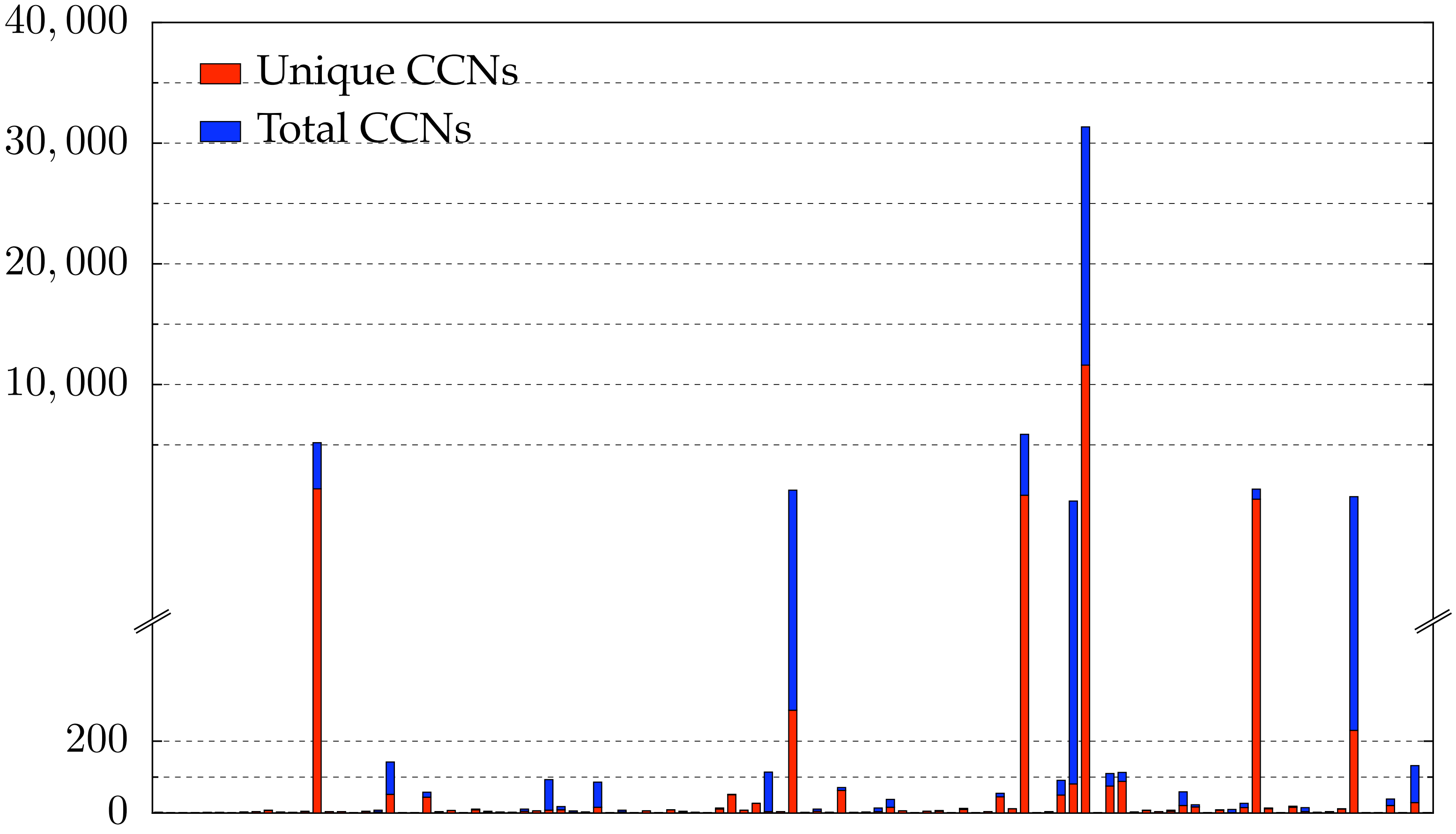
There are several problems with the “stop list” approach:

The list must be:

- Constructed.
- Maintained.
- Tuned for different applications.

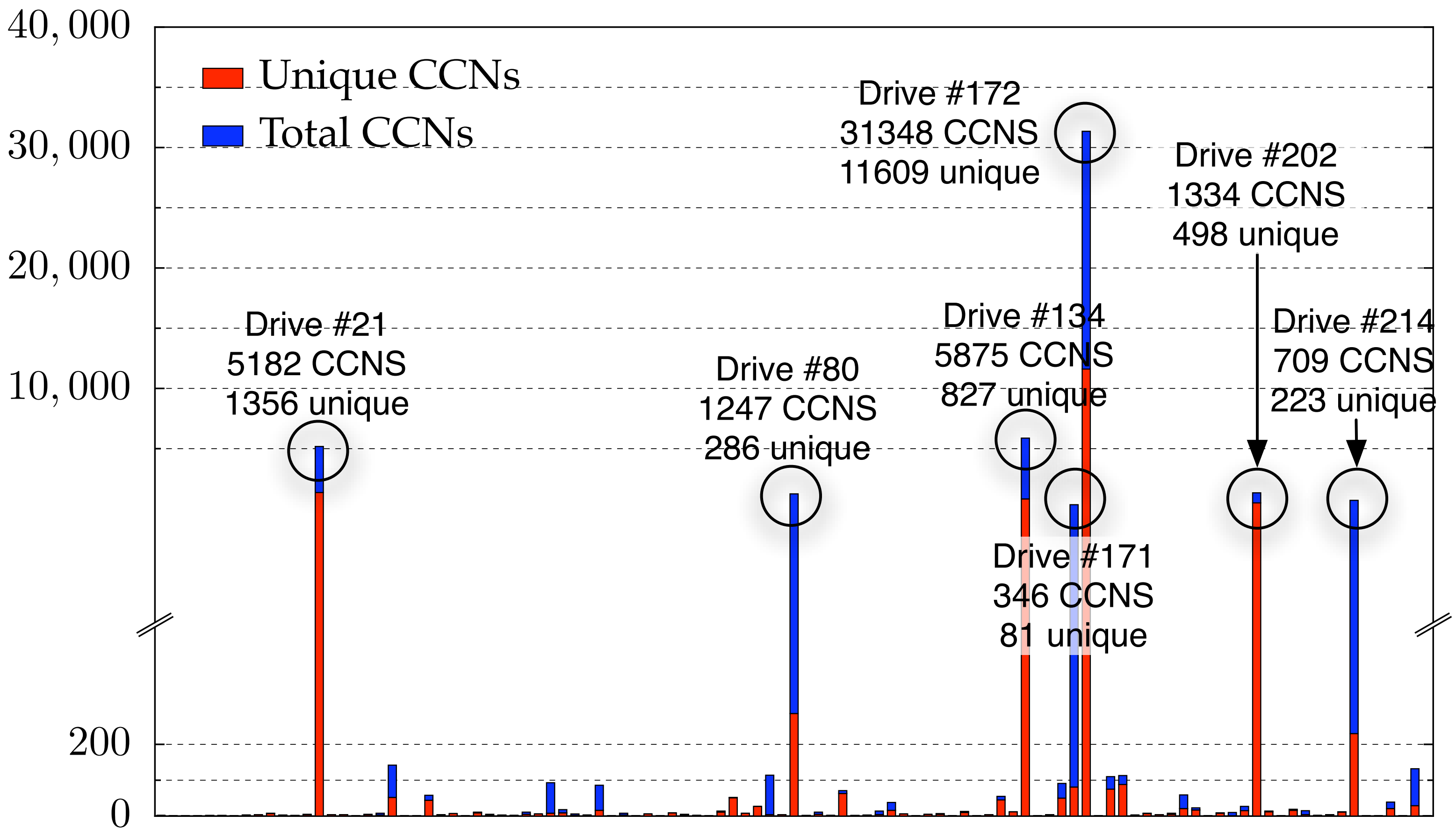
Building a “stop list” requires judgement and patience.

An alternative is to assume that “false positives” are rare and focus on those drives with high response.



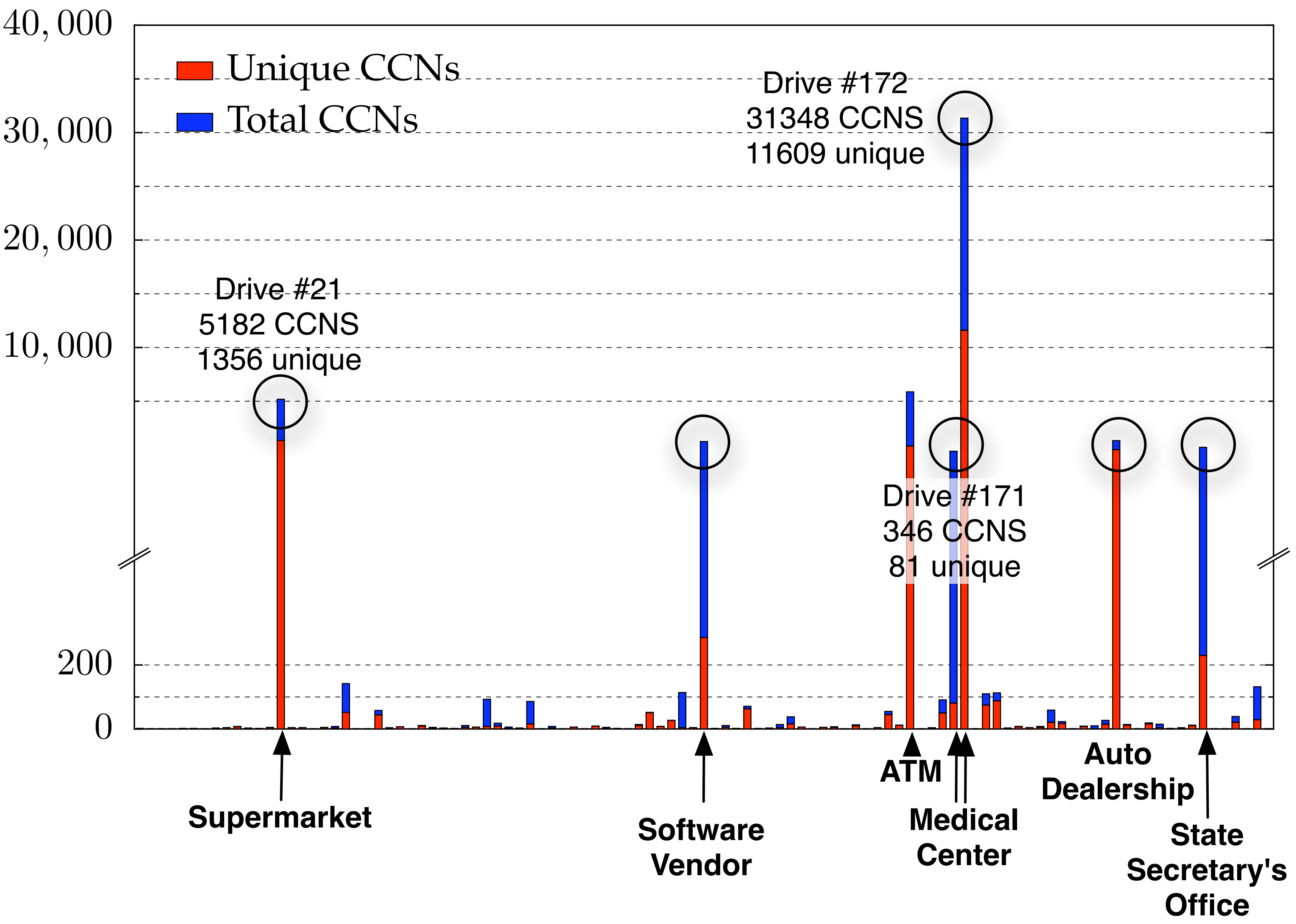
By definition, no drive should contain a large number of CCNs, so these drives are all interesting.

An alternative is to assume that “false positives” are rare and focus on those drives with high response.

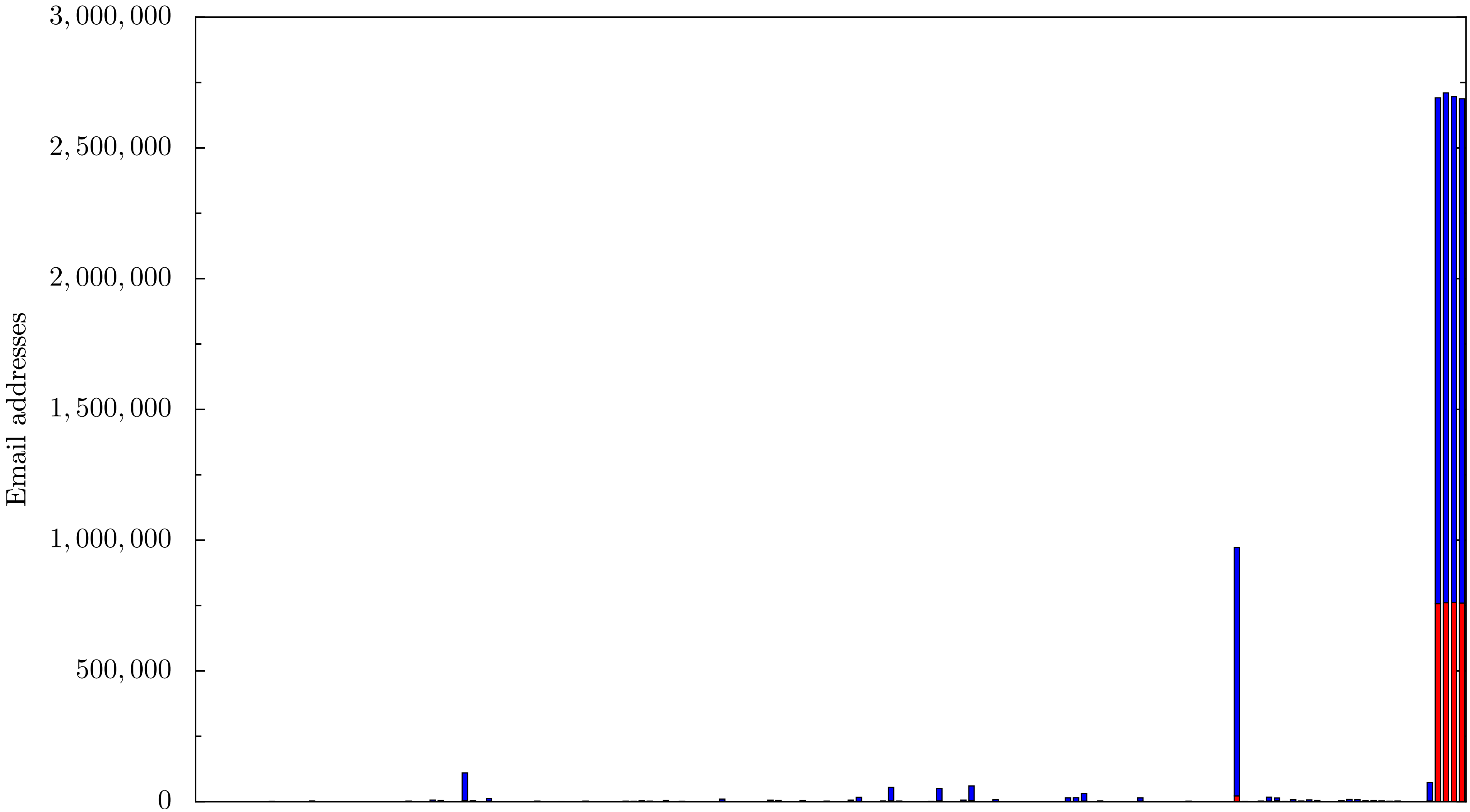


Only 7 drives had more than 300 credit card numbers.

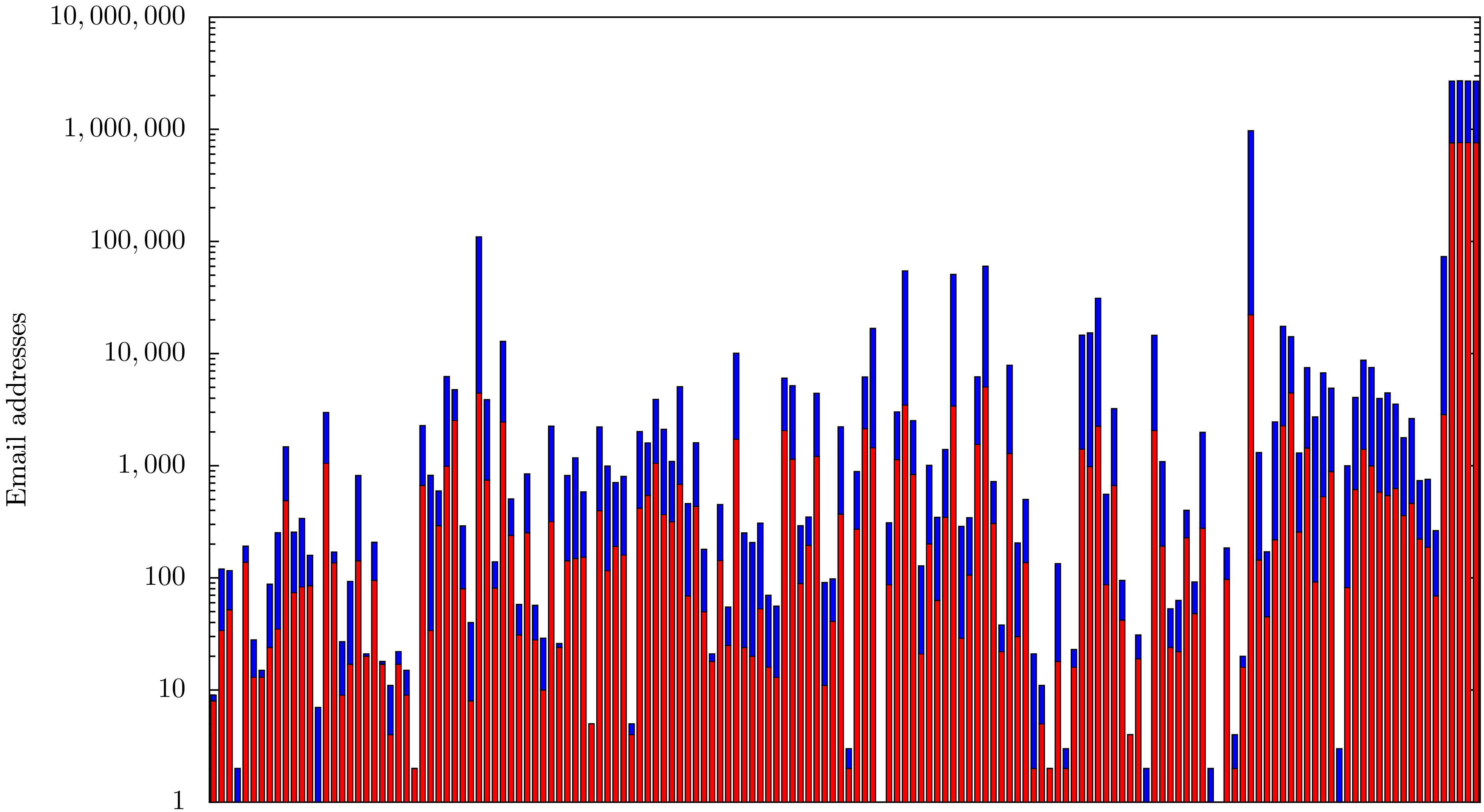
With a "credit card number detector," we can rapidly identify drives with leaked consumer information.



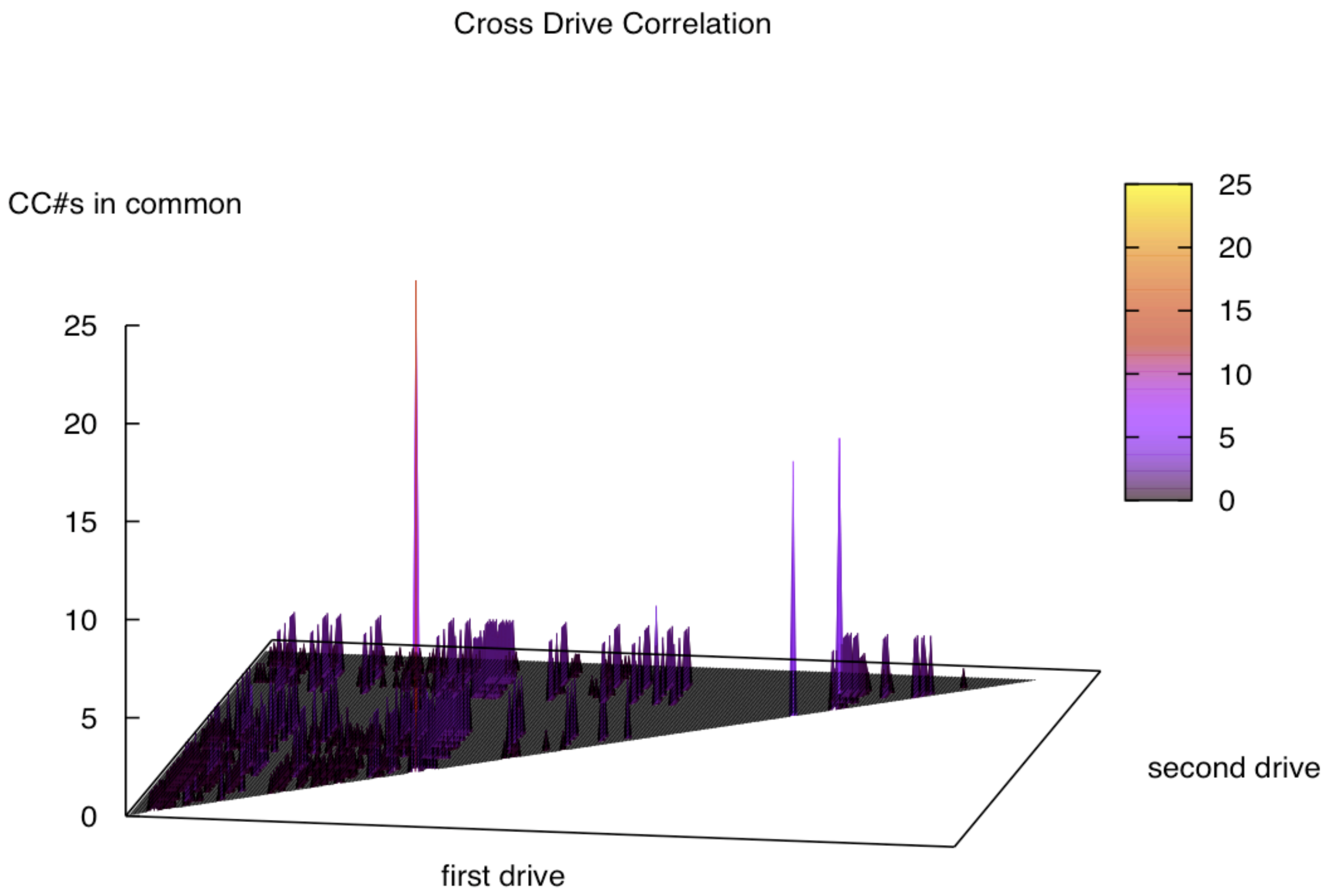
Email Addresses



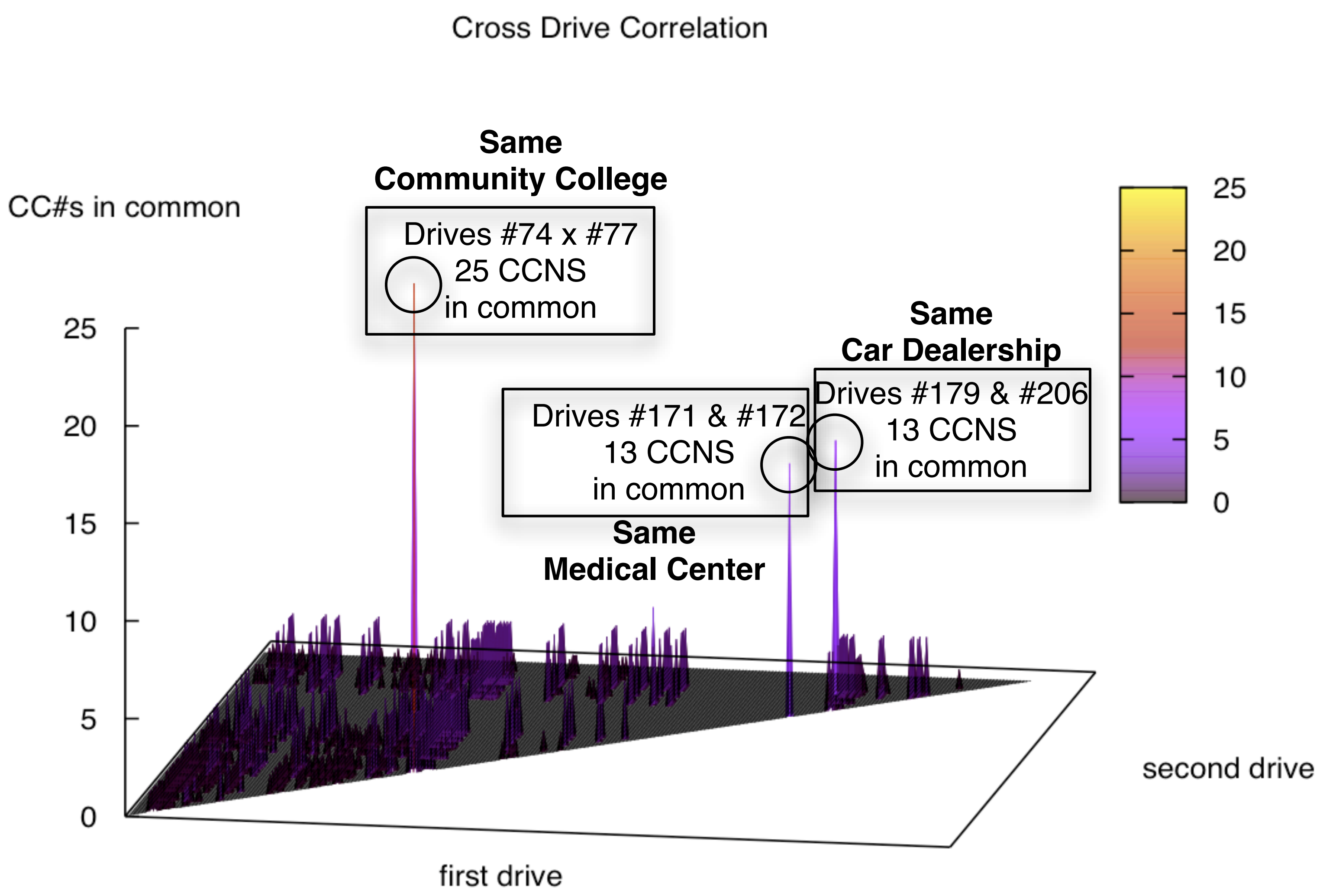
Email Addresses



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

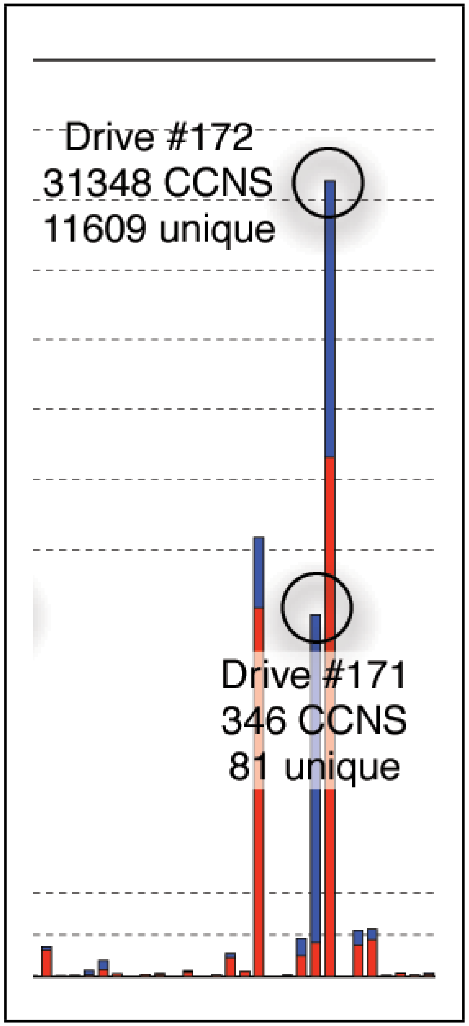


Second-order analysis uses correlation techniques to identify drives of interest.



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

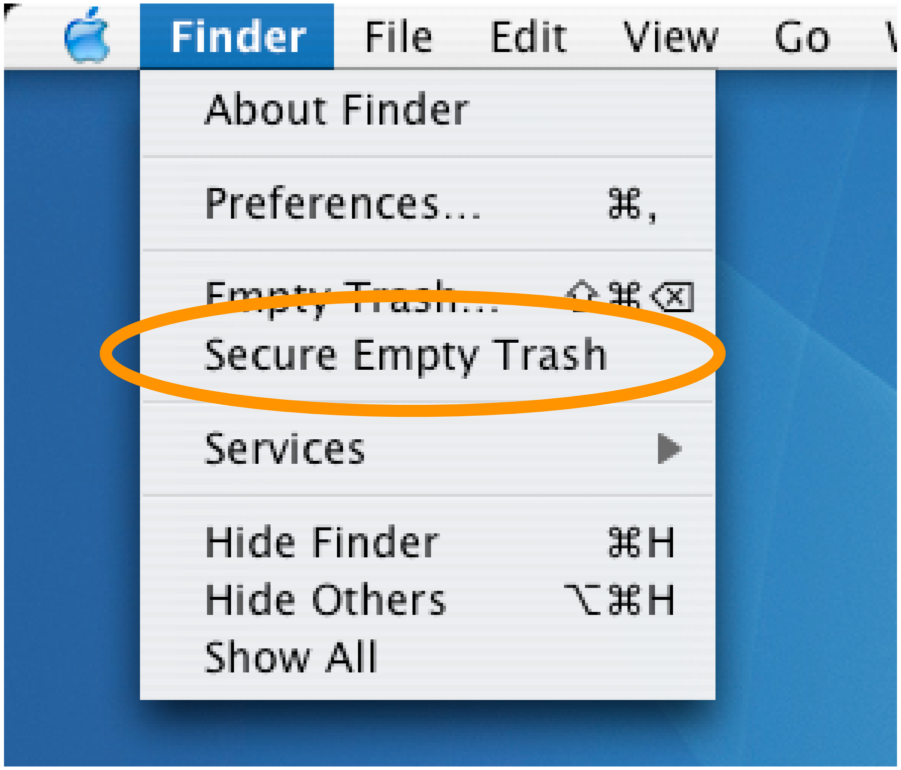
Possible Uses:

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

Reactions to this research:

Legislative: “Fair and Accurate Credit Transactions Act of 2003”

Technical: Modifications to MacOS & Windows



Current Research Projects

- ✔ Evaluating "big file" sanitization technique.
- Scaling up cross-drive analysis
- Continued development of AFF and AFFLIB

- S/MIME
- "Computation and Human Thought" book.

Long-term Agenda

- Fix security & privacy in current systems.
- Next-generation forensic tools.

- New tools for secure personal data management.
- Resolving privacy and ubiquitous data collection.

Questions?