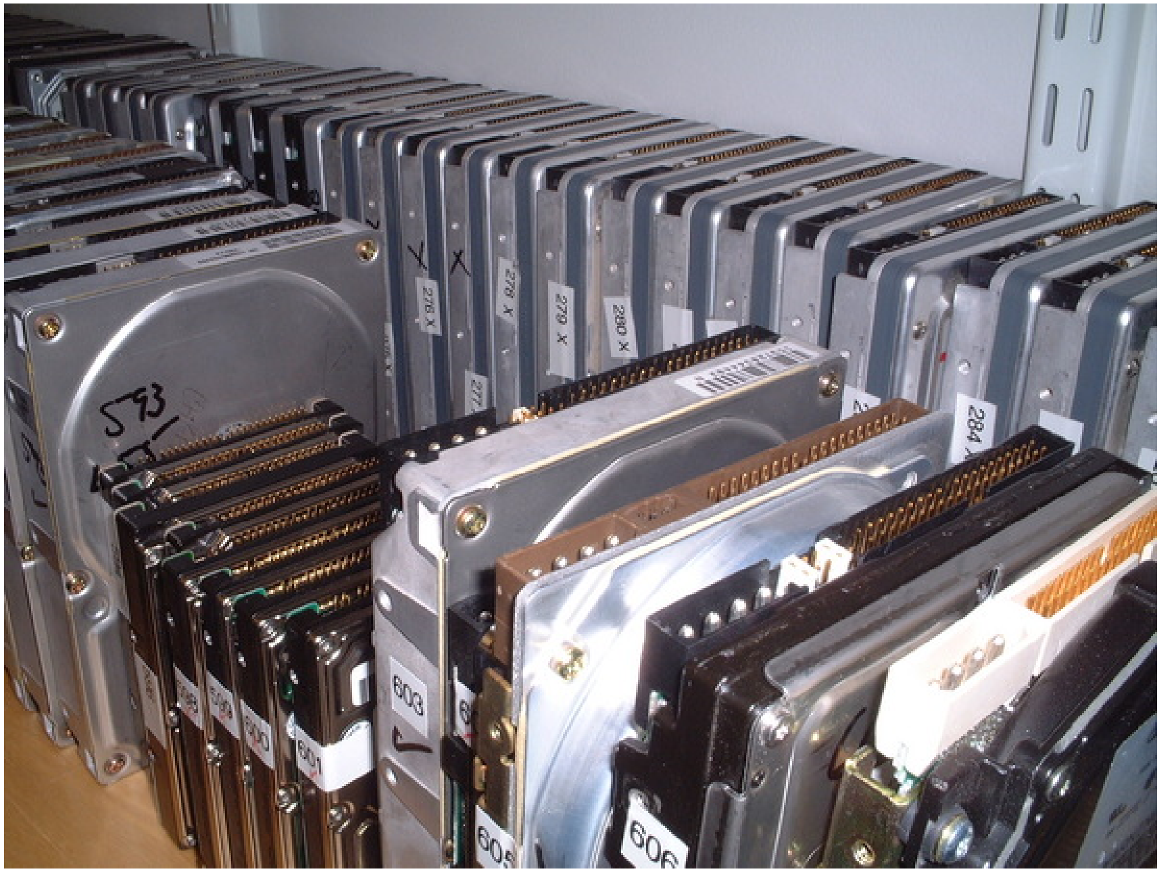
CRCS Forensics Research



Simson L. Garfinkel

February 17, 2006

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

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

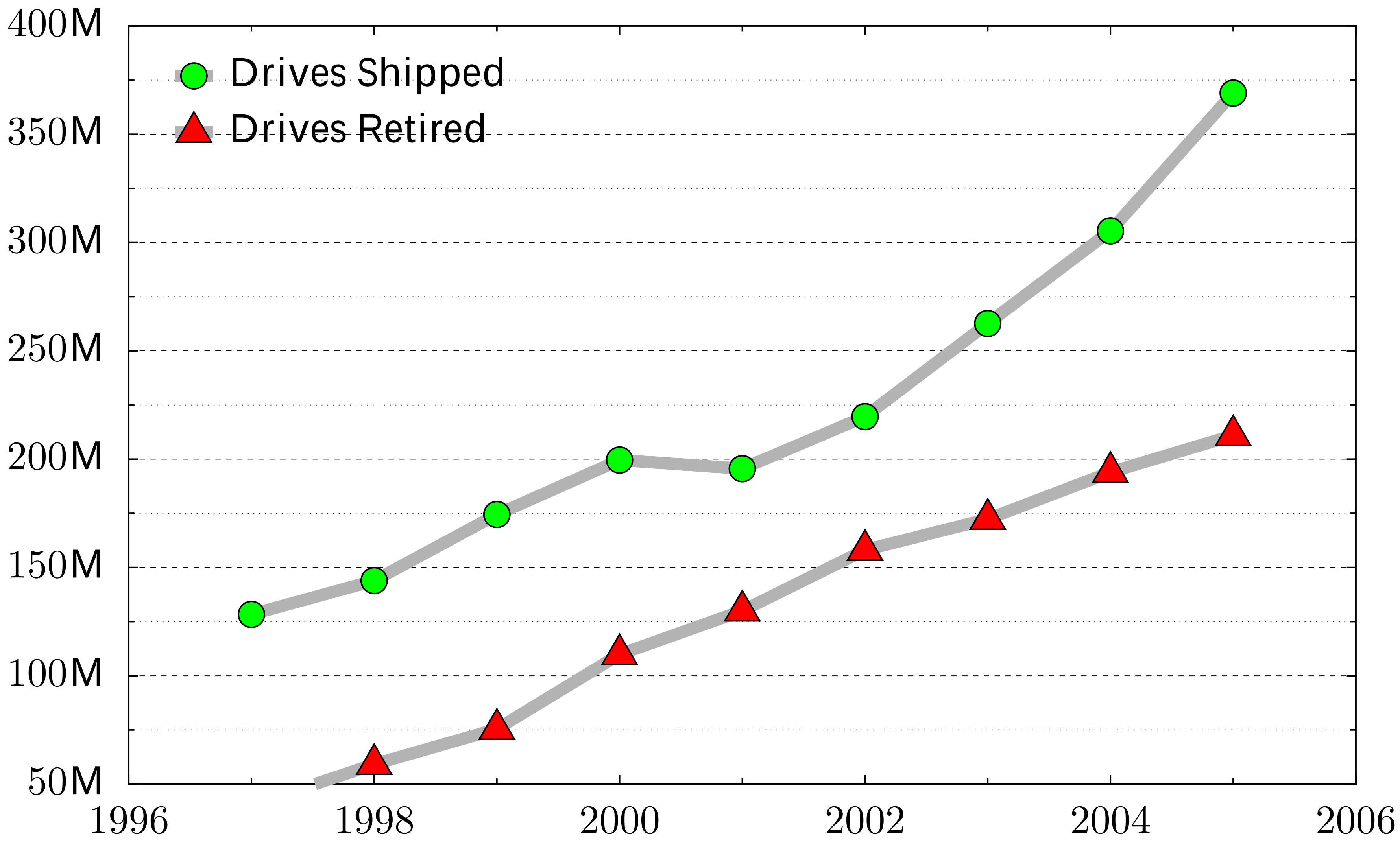
Do not forget data when power is removed.

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

Scale of the problem: huge!



210 million drives will be retired this year.

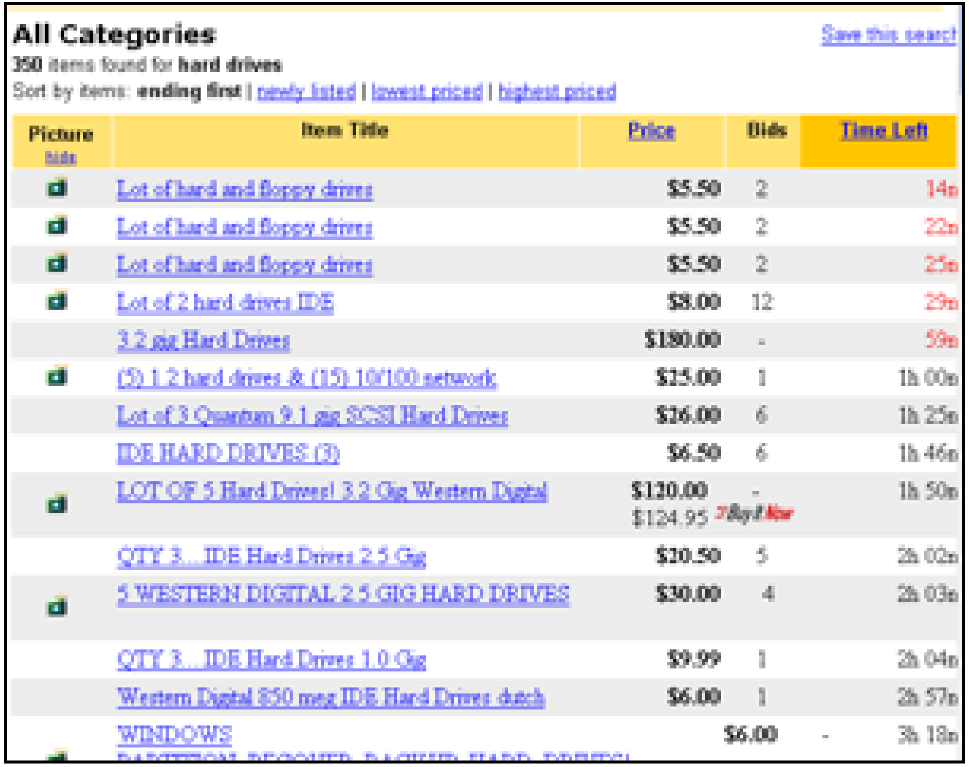
Physical destruction will remove the information...

...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

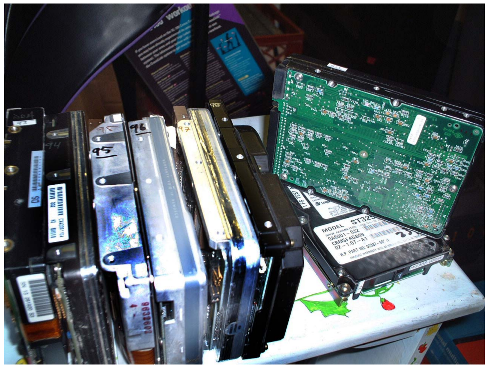
About 1000 used drives/day sold on eBay.

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

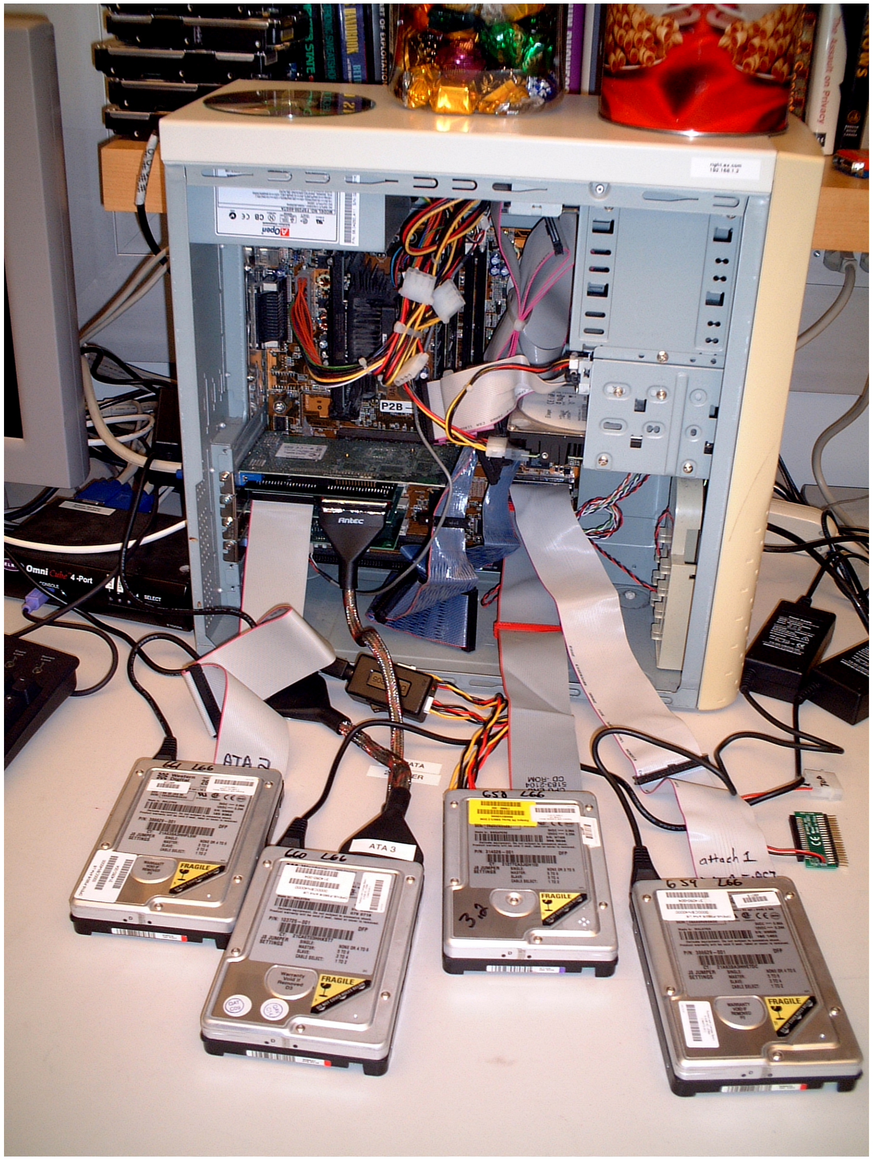
2003: 150 drives

2005: 500 drives

2006: 950 drives

Drives arrive by UPS and USPS

Drives are “imaged.”



Images stored on external firewire drives

This is 900GB of storage.

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

Purchased for $5 from a Mass retail store on eBay

Copied the data off: 541MB

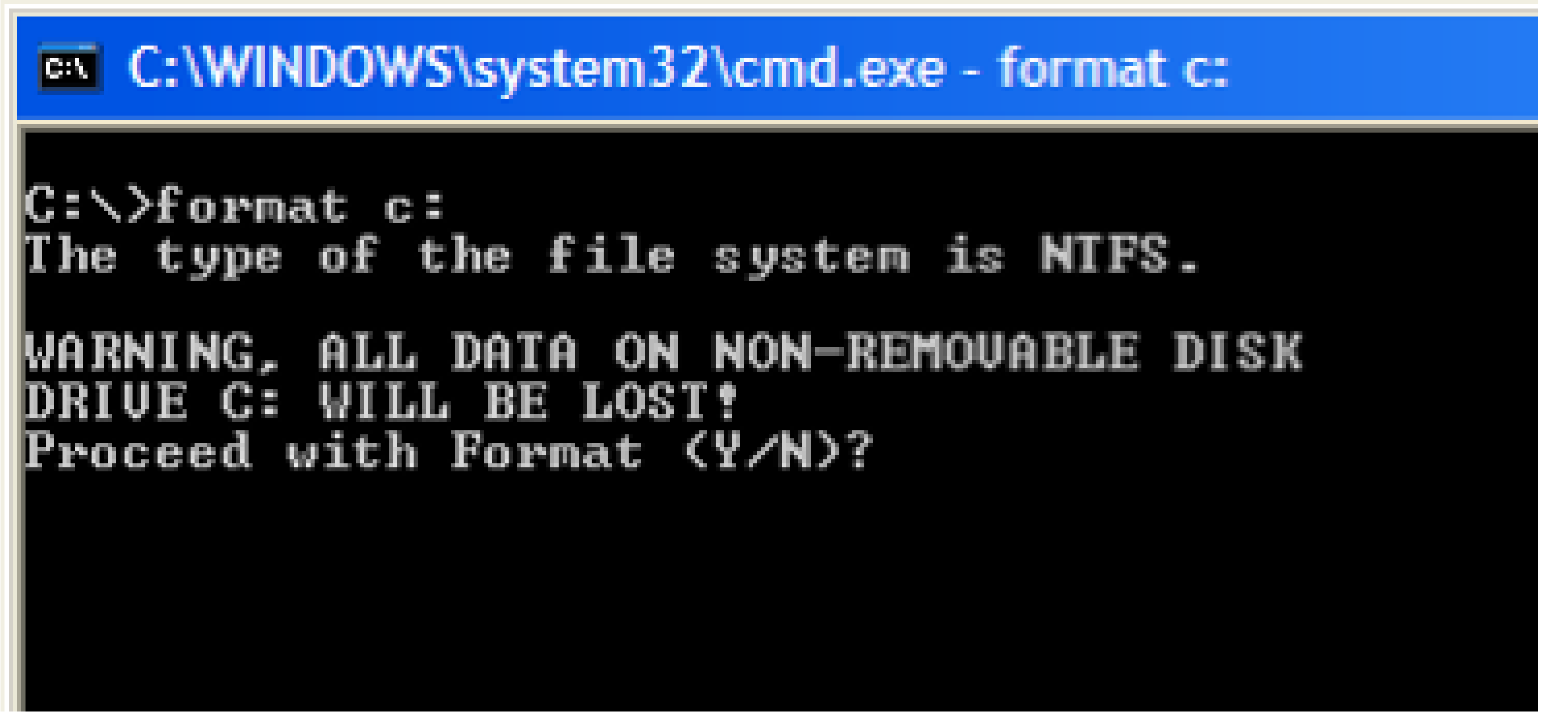
Initial analysis:

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)?
```

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:

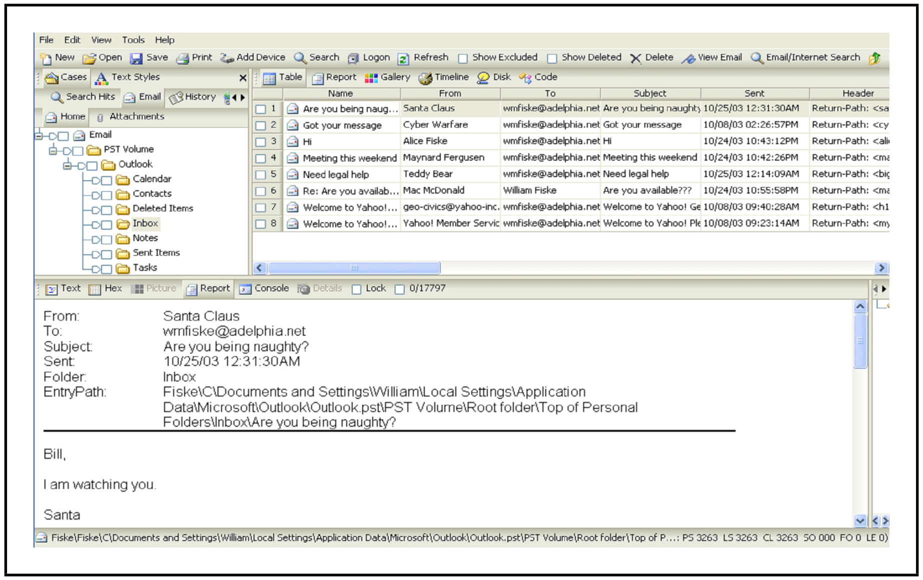
- Thousands of credit card numbers
- Financial records
- Medical information
- Trade secrets
- Highly personal information

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

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

Traditional forensics tools:

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


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,

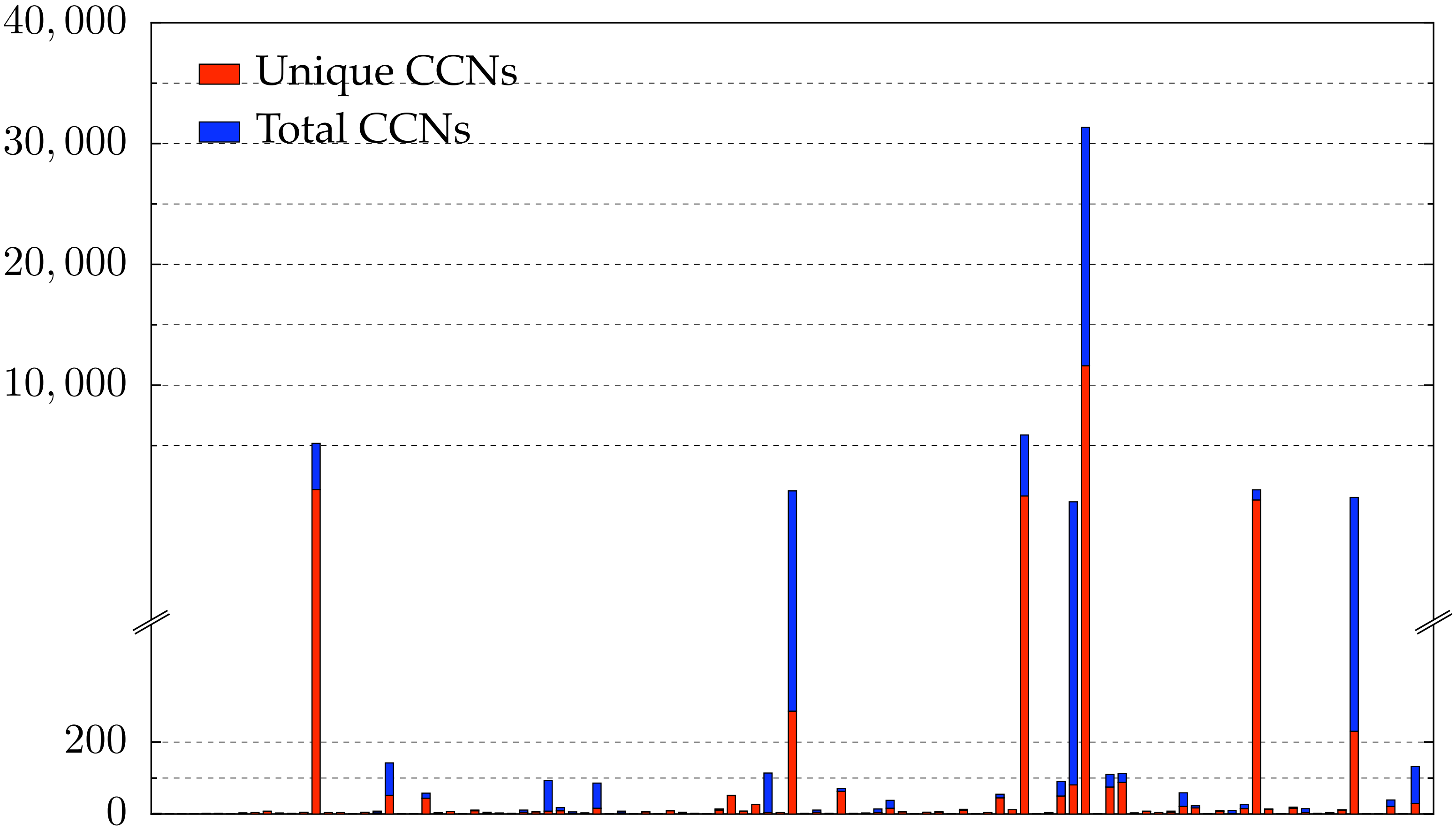
I am watching you.

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)

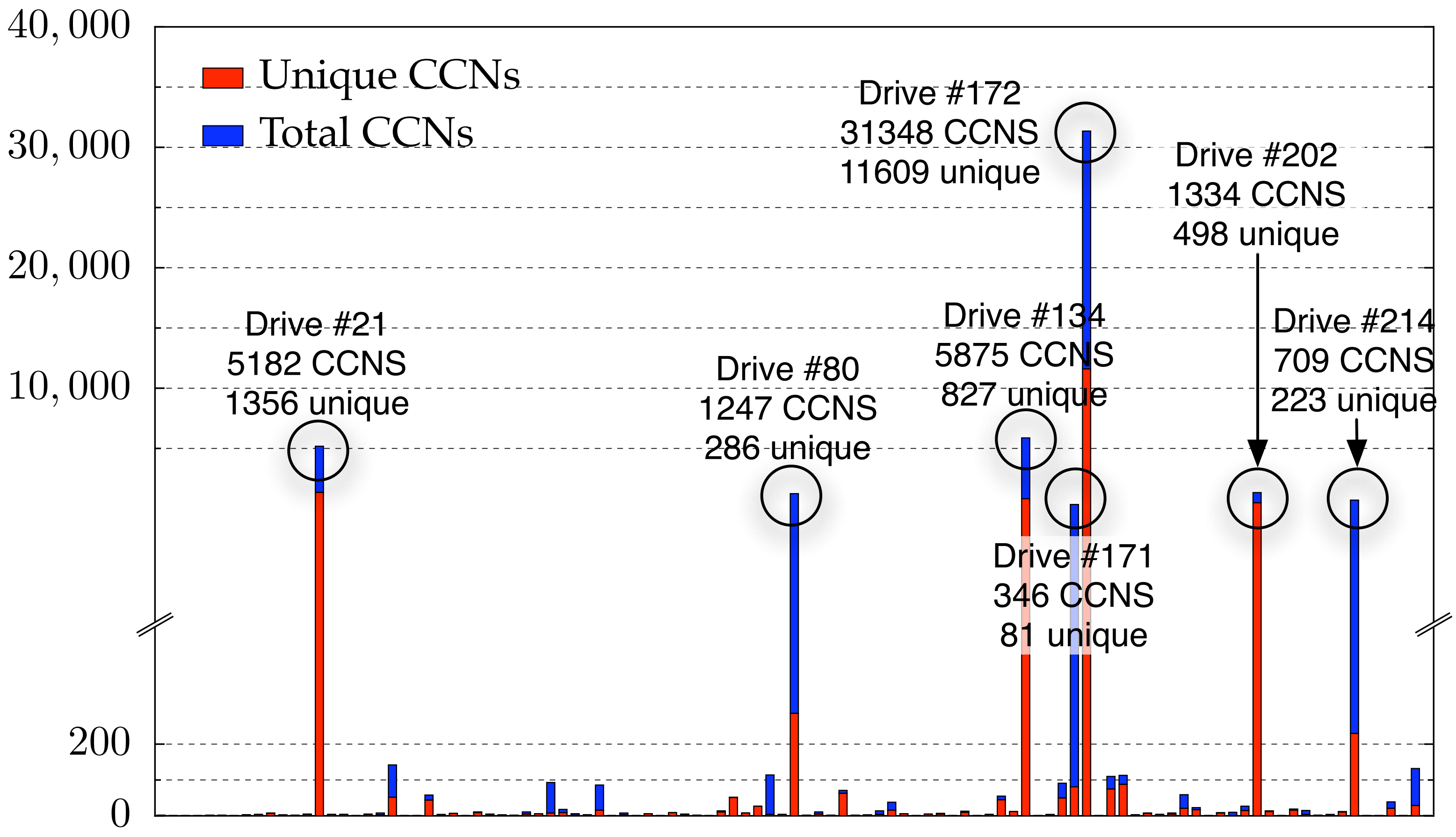
In [Garfinkel ’05], there were *hundreds* of disks to analyze.

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



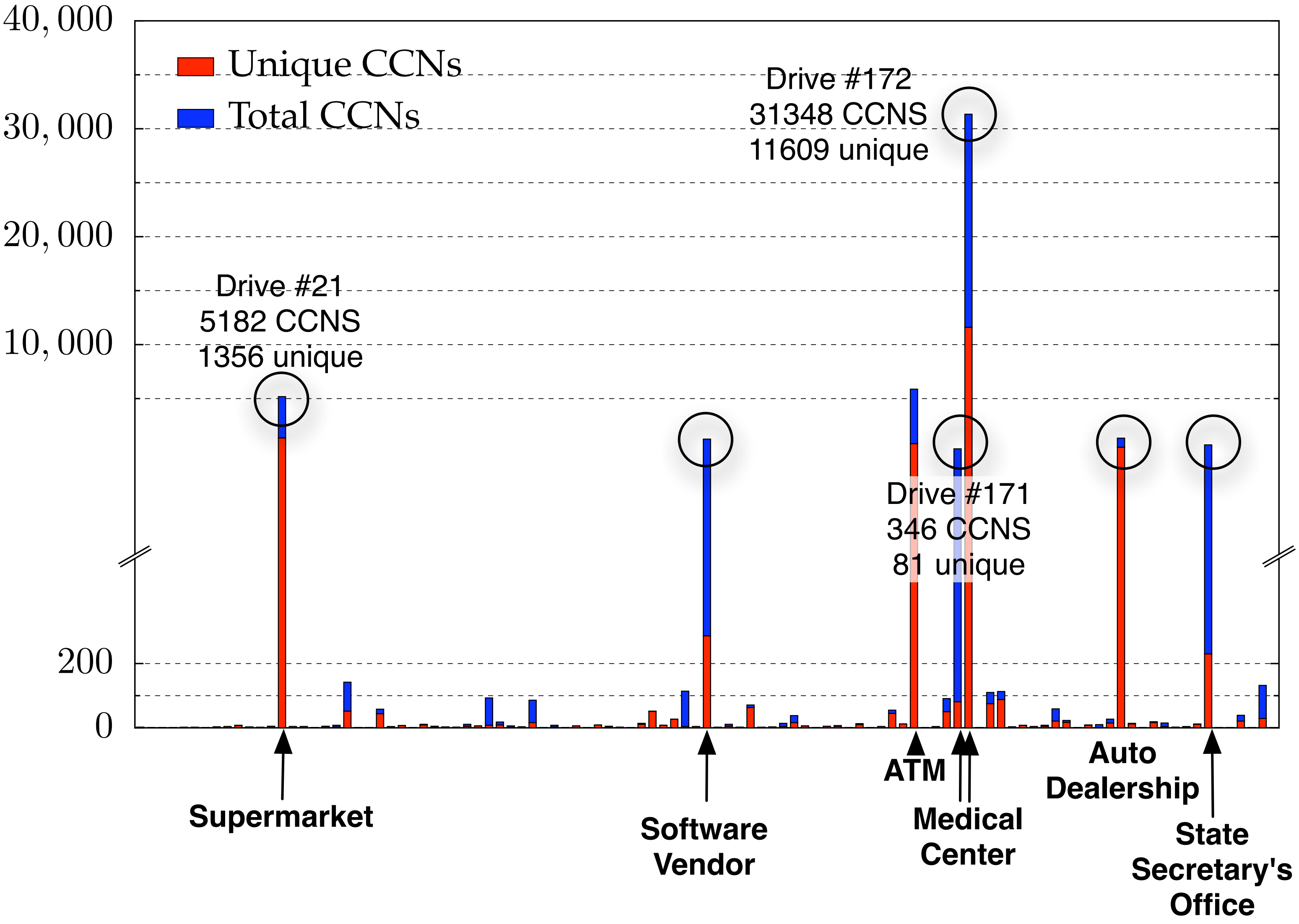
.

This filter looks for Credit Card Numbers.



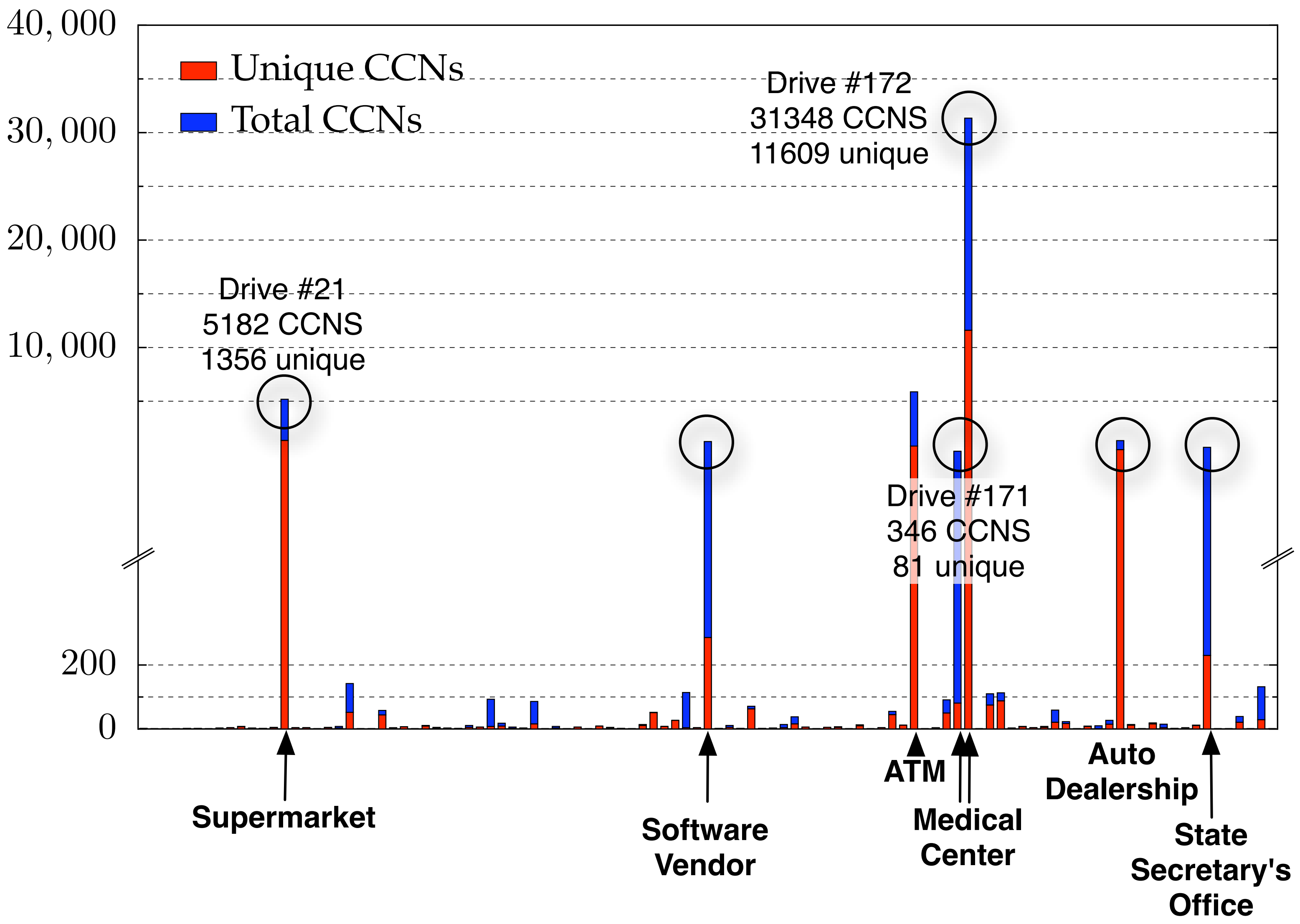
Only 7 drives had more than 300 credit card numbers.

These drives were traced back to their original owners.



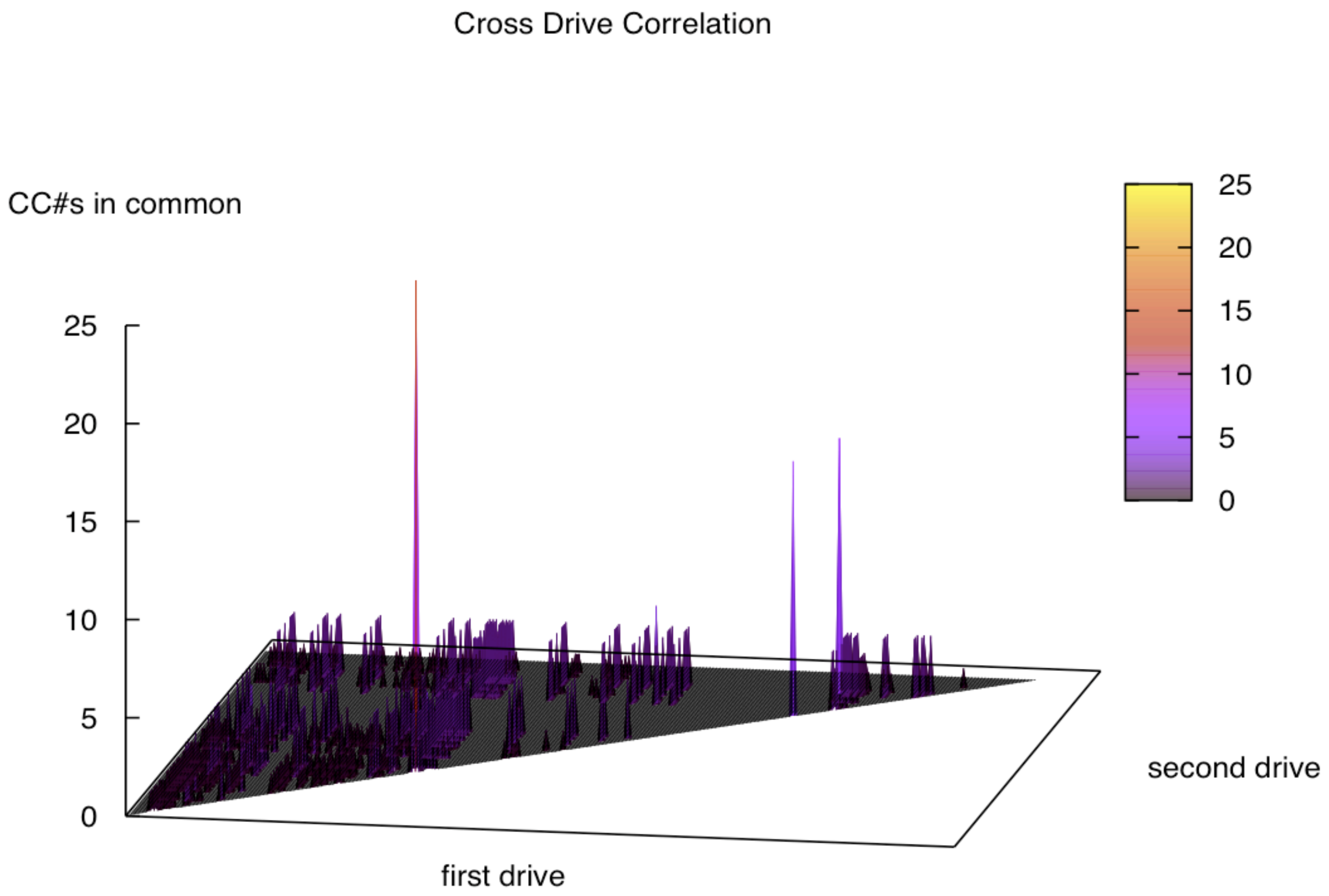
.

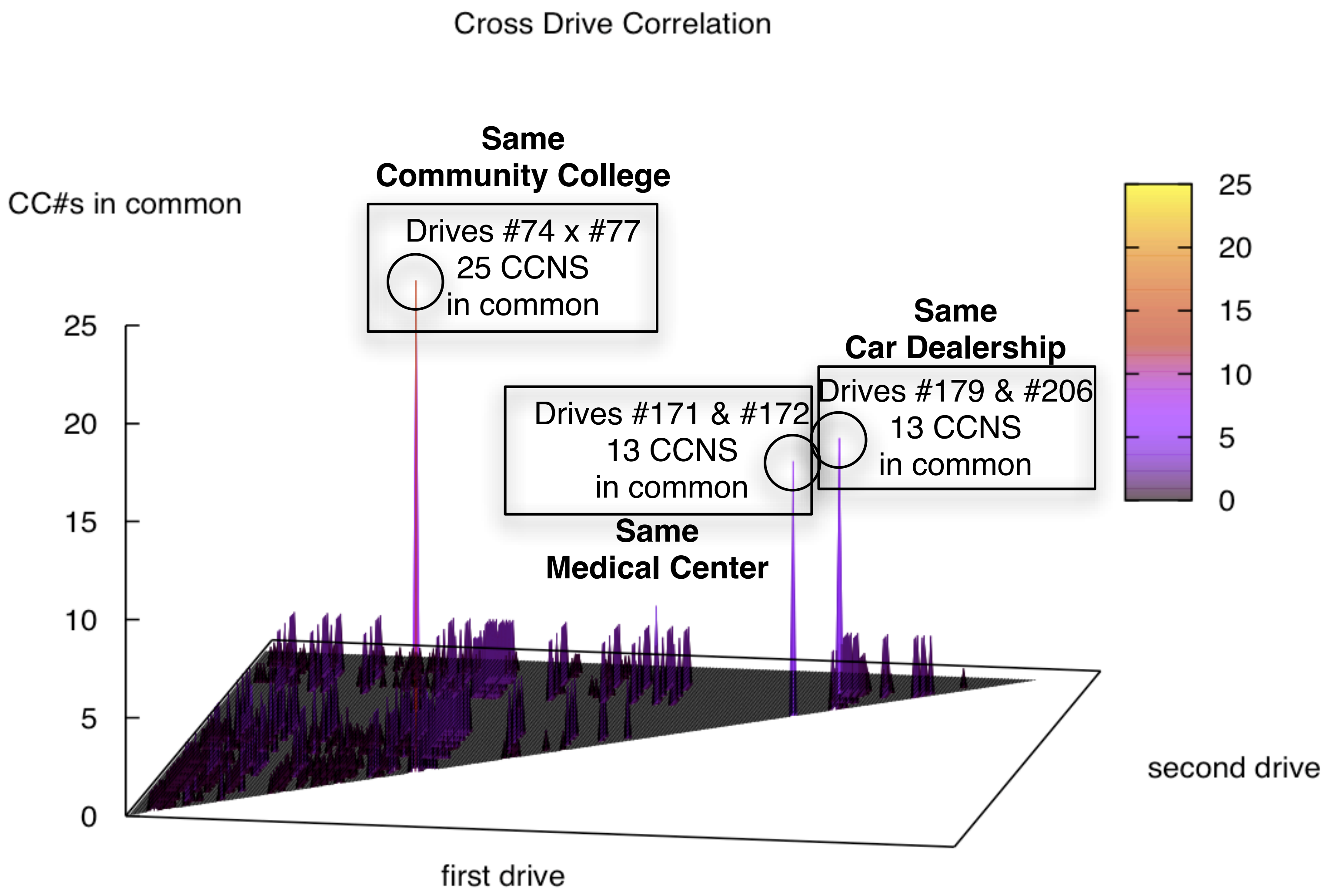
With a “credit card number detector,” we can rapidly identify drives with leaked consumer information.



.

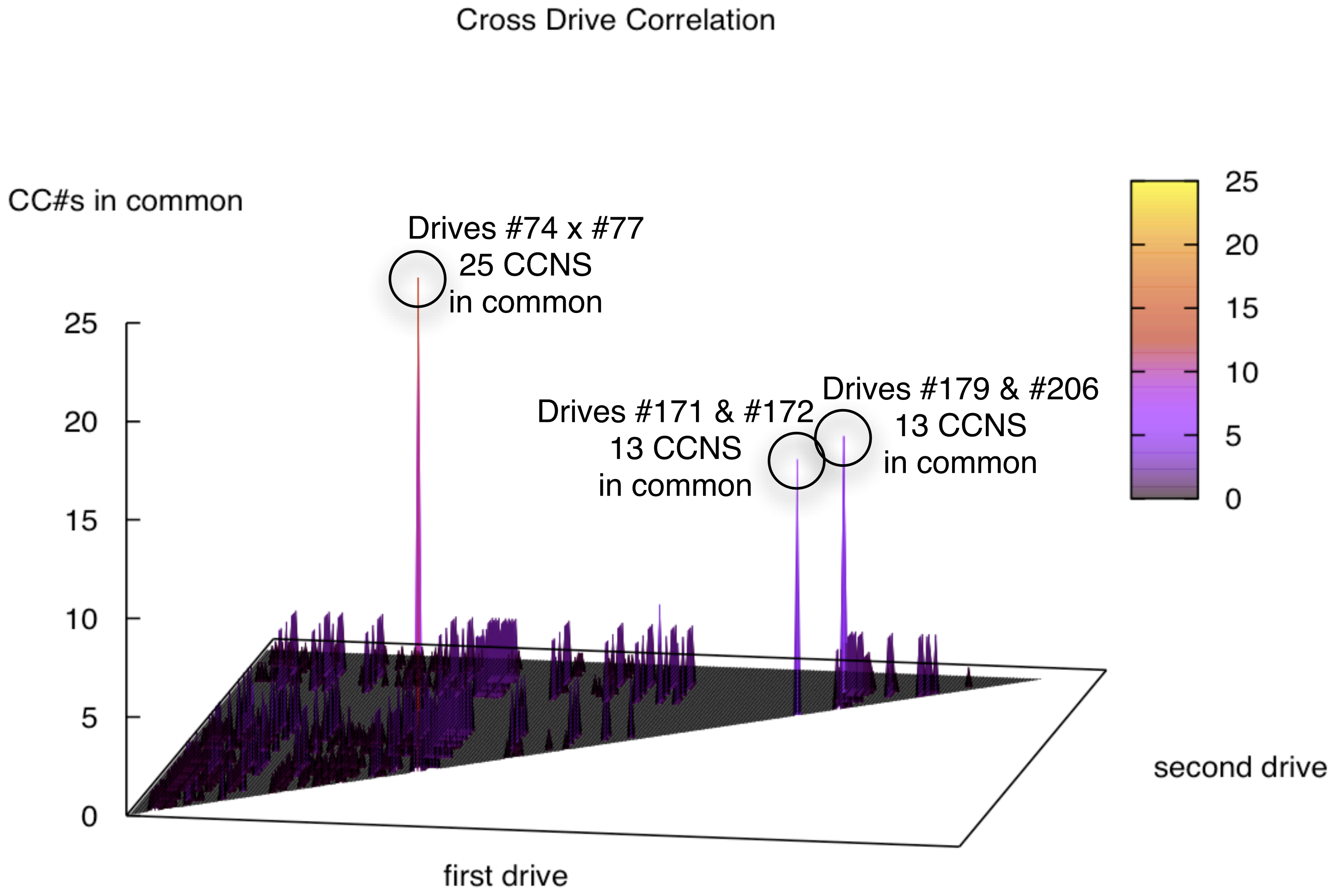
Second-order analysis uses correlation techniques to identify drives of interest.





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

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

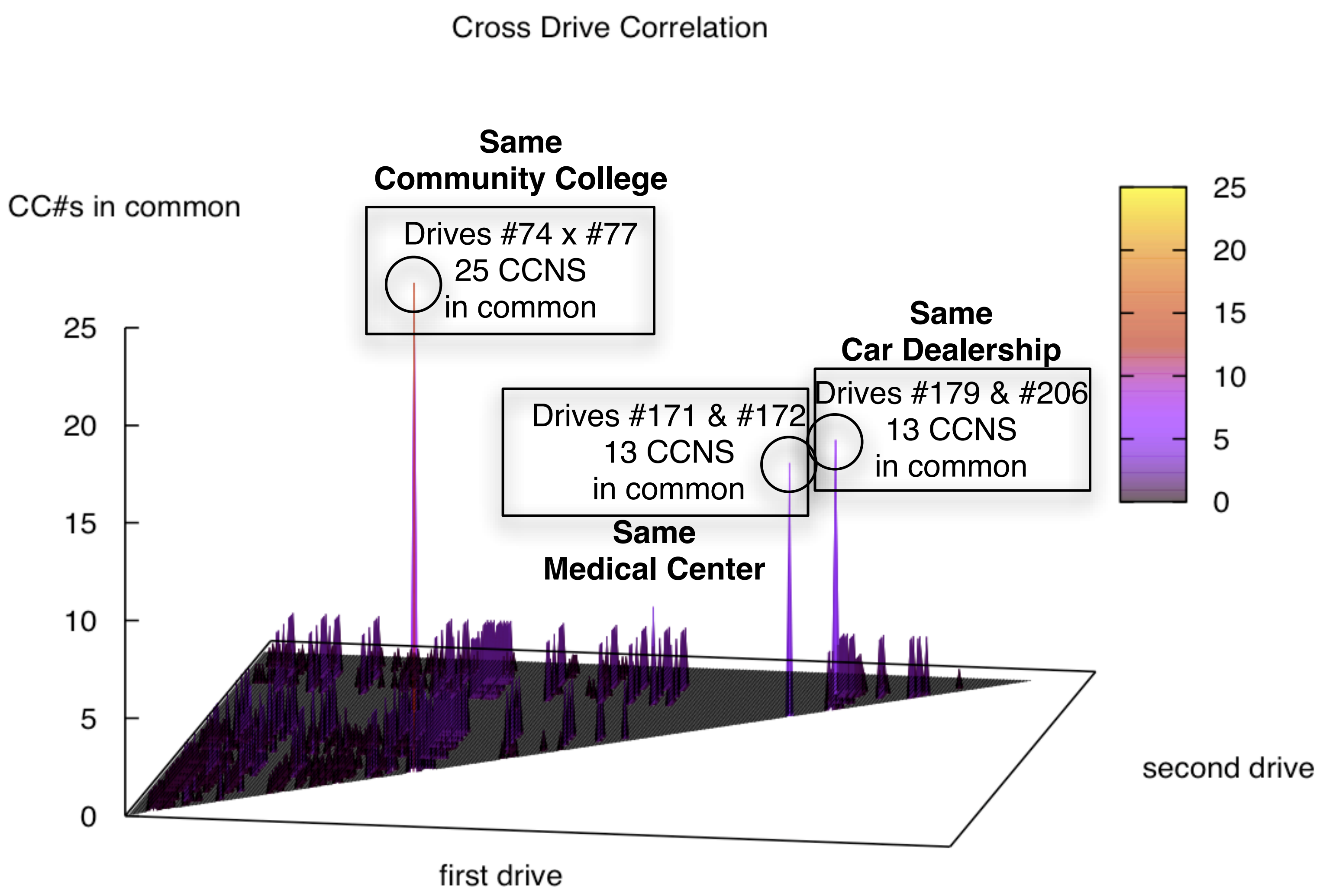


Drives #74 x #77
25 CCNS
in common

Drives #171 & #172
13 CCNS
in common

Drives #179 & #206
13 CCNS
in common

Manual analysis of on-drive data reveals that these drives are from the same organization.



Cross Drive Correlation

Same Community College

Drives #74 x #77
25 CCNS
in common

Same Medical Center

Drives #171 & #172
13 CCNS
in common

Same Car Dealership

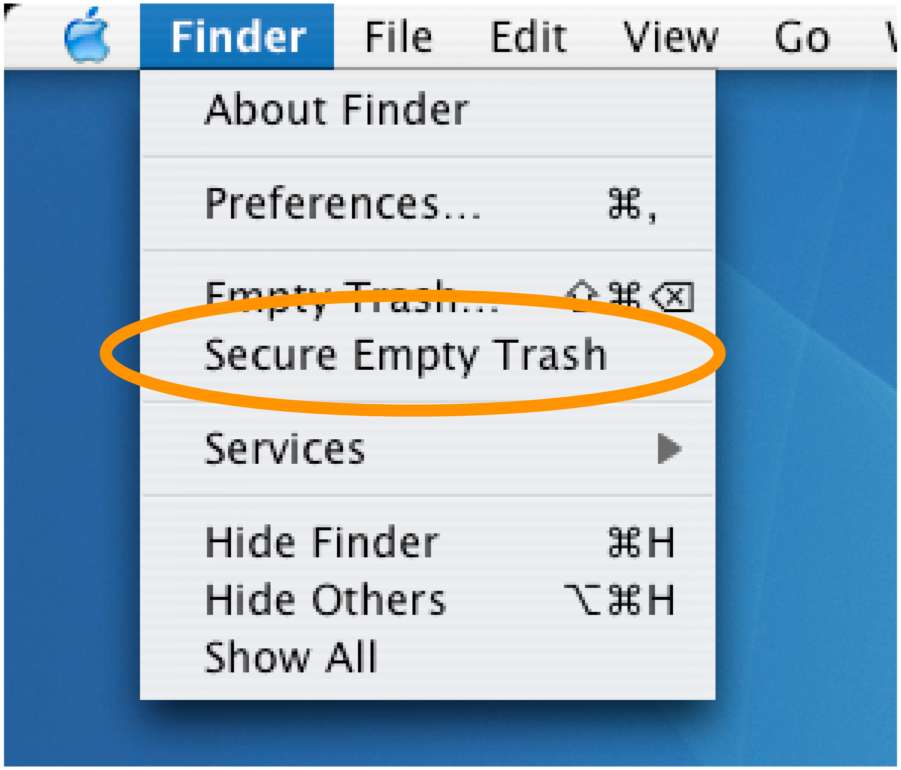
Drives #179 & #206
13 CCNS
in common

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.”

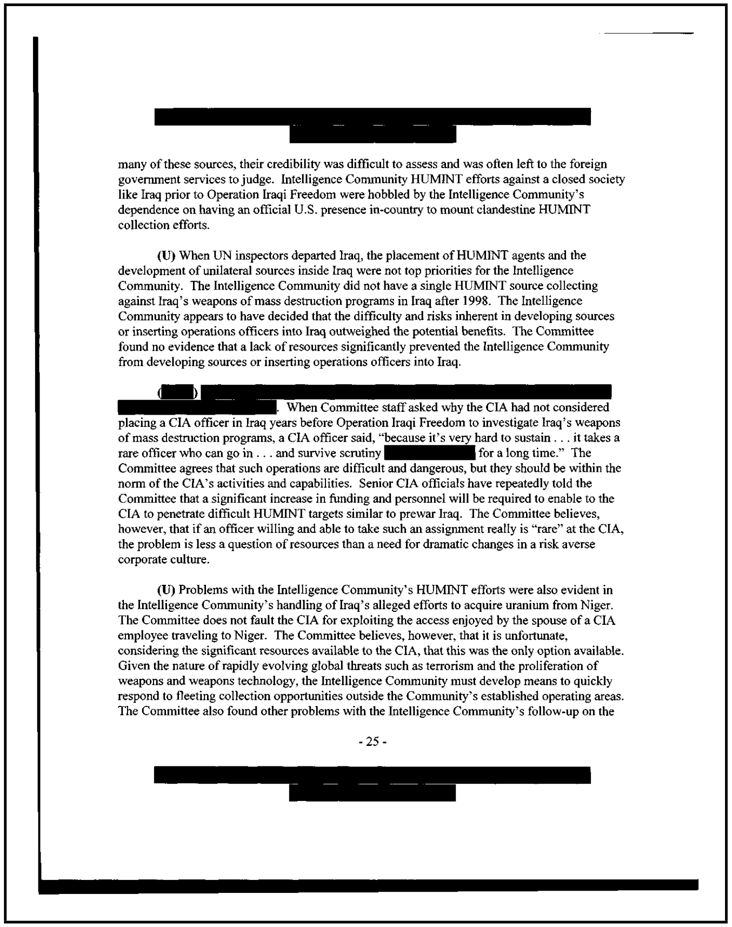
Testimony: `http://tinyurl.com/cd2my`

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



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 -

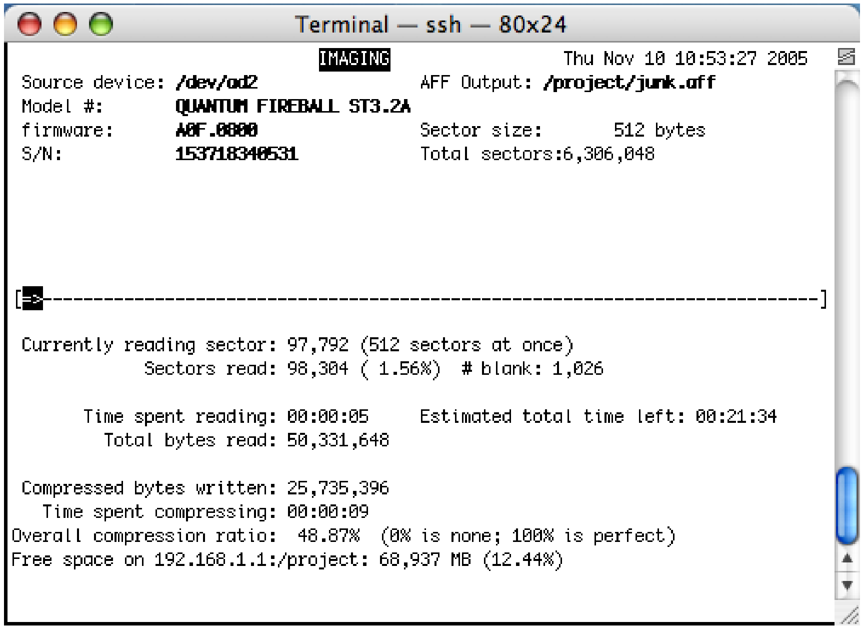
[redacted]

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

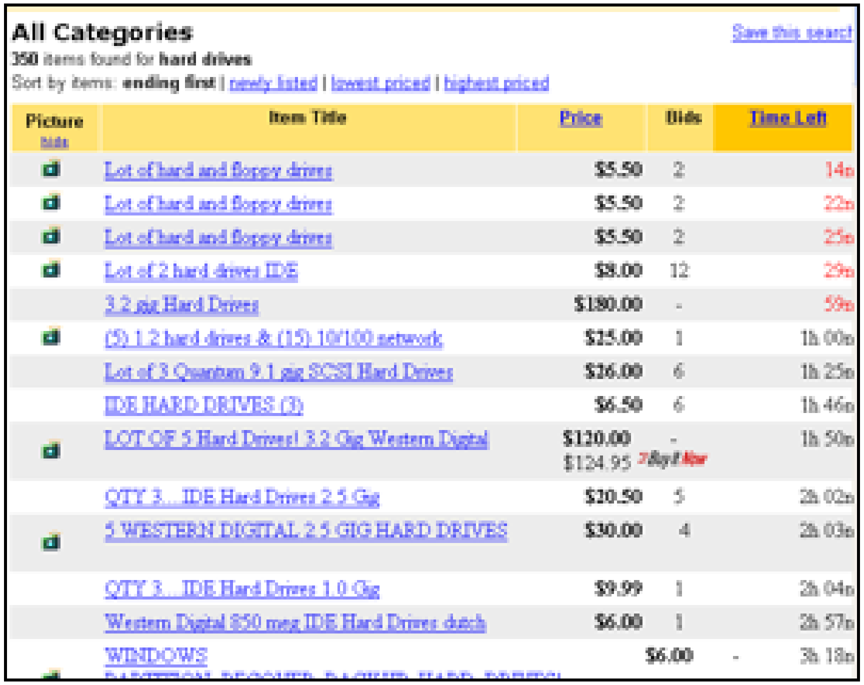
- Improved imaging, storage and backup.
- Web-based database of hash codes.



```
                             IMAGING                 Thu Nov 10 10:53:27 2005
 Source device: /dev/ad2               AFF Output: /project/junk.aff
 Model #:       QUANTUM FIREBALL ST3.2A
 firmware:      A0F.0800               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	$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.

There is a lot more work to do.

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.