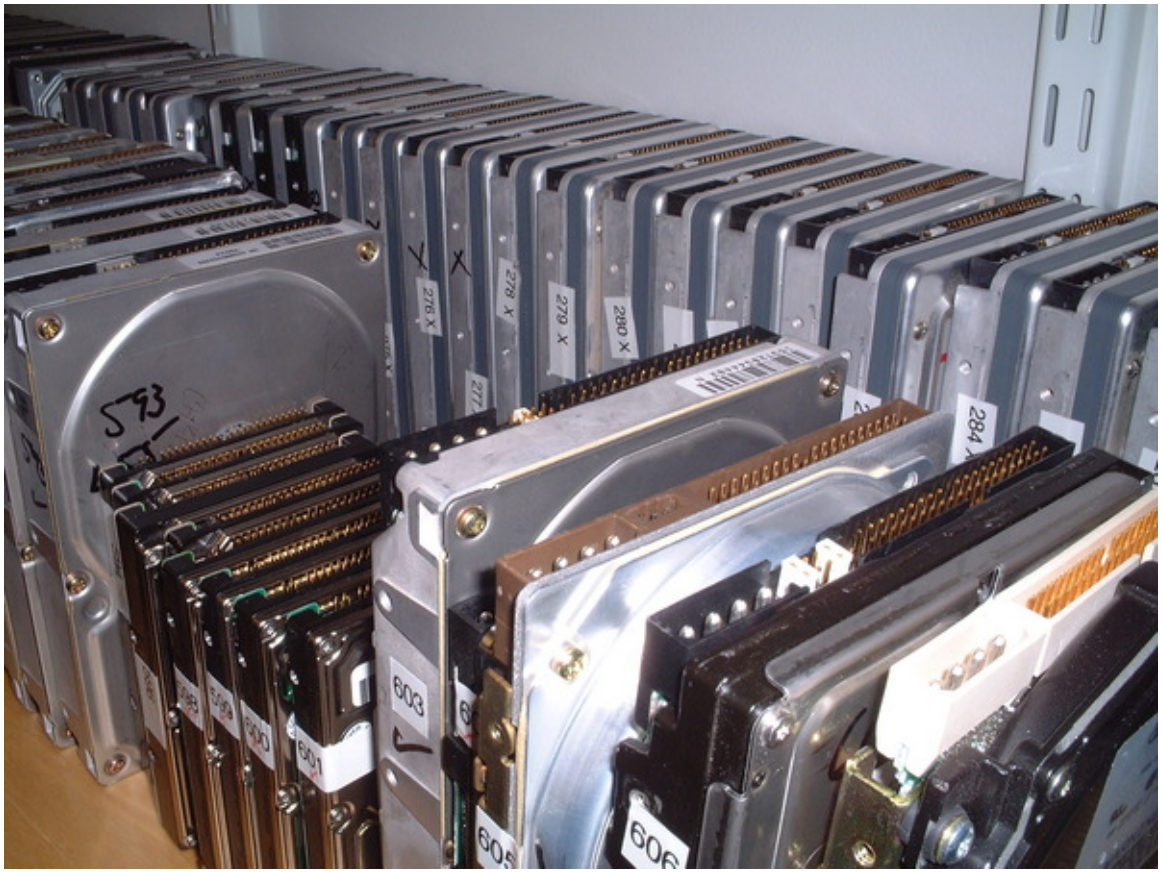
CRCS Forensics Research

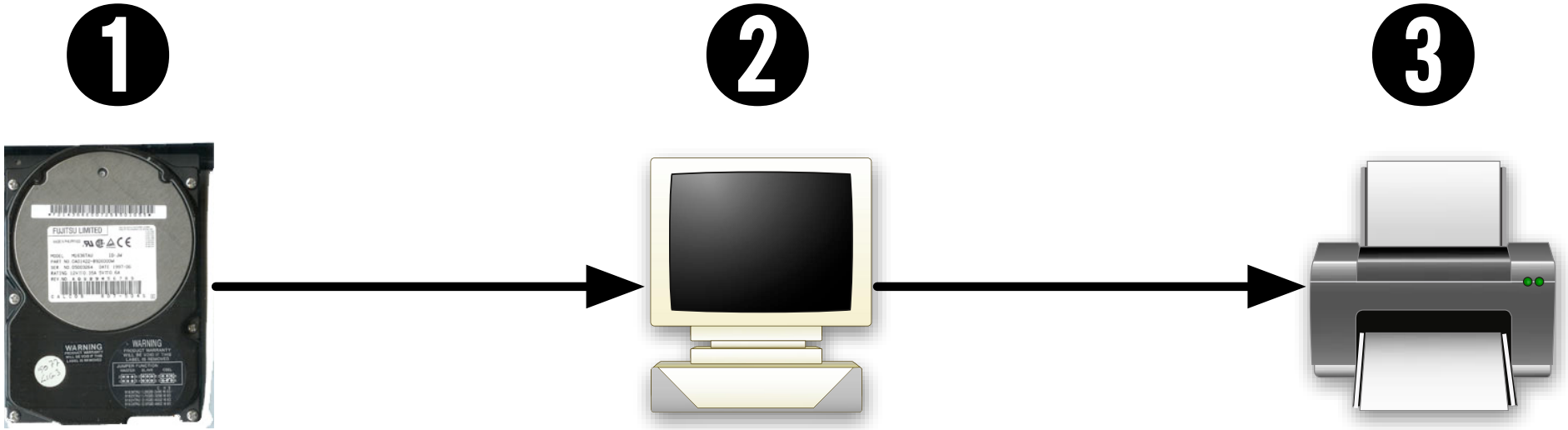


Simson L. Garfinkel

February 17, 2006

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

Today's forensic tools and file formats are designed to analyze a single drive at a time.



These tools are not adequate for today's forensic challenge.

Digital forensics have opened up a whole new world for law enforcement and intelligence.

- Recovery of “deleted” files and email
- Automatic identification of “child pornography”
- Rapid searching for target names and email addresses

Today's forensic techniques don't scale.

Process is labor intensive.

Disk drives are getting bigger.

Law enforcement seizes more drives every year.

I am developing a different approach based on a different set of requirements.

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

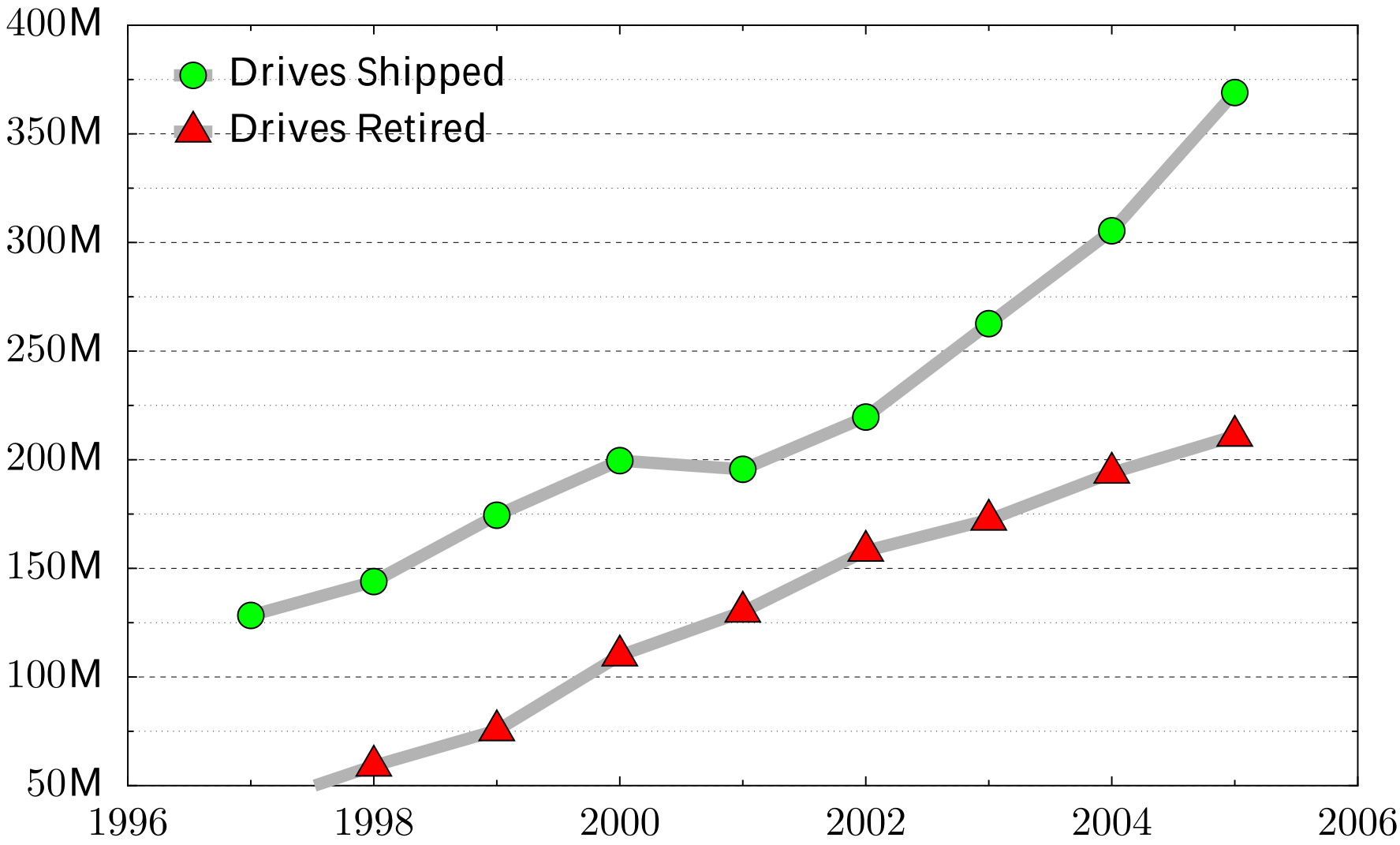
Do not forget data when power is removed.

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 Bids	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	$15.00	1	1h 00m
	Lot of 3 Quantum 9.1 gig SCSI Hard Drives	$36.00	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	$10.50	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

About 1000 used drives/day sold on eBay.

I purchase hard drives on the secondary market.

2001: 100 drives

2003: 150 drives

2005: 500 drives

2006: 950 drives

Data on drives “imaged” using FreeBSD and AImage

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
```

% 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:

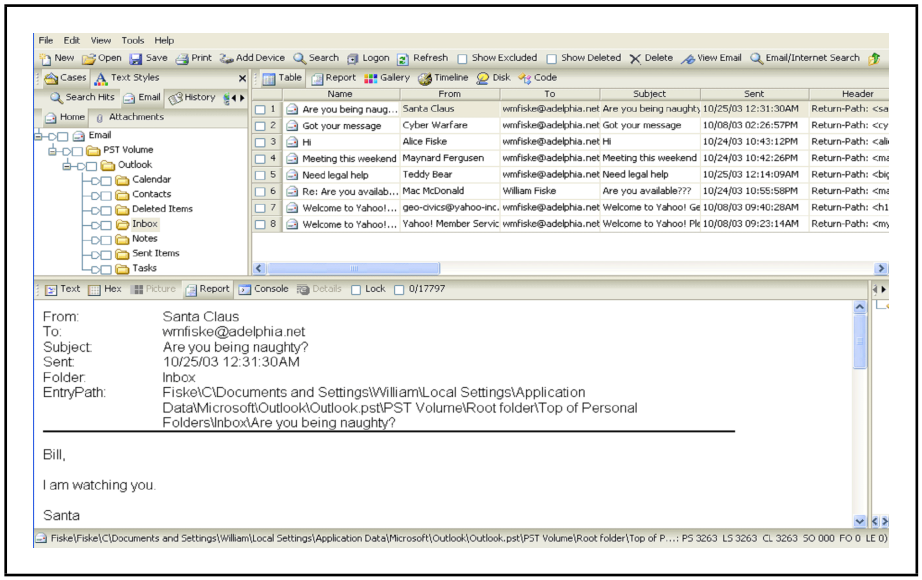
- Thousands of credit card numbers
- Financial records
- Medical information
- Trade secrets
- Highly personal information

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.



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

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

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

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

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

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

Example problem: Who owned this disk drive?

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

- Needs forensic skill.
- Requires complete documents.

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

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

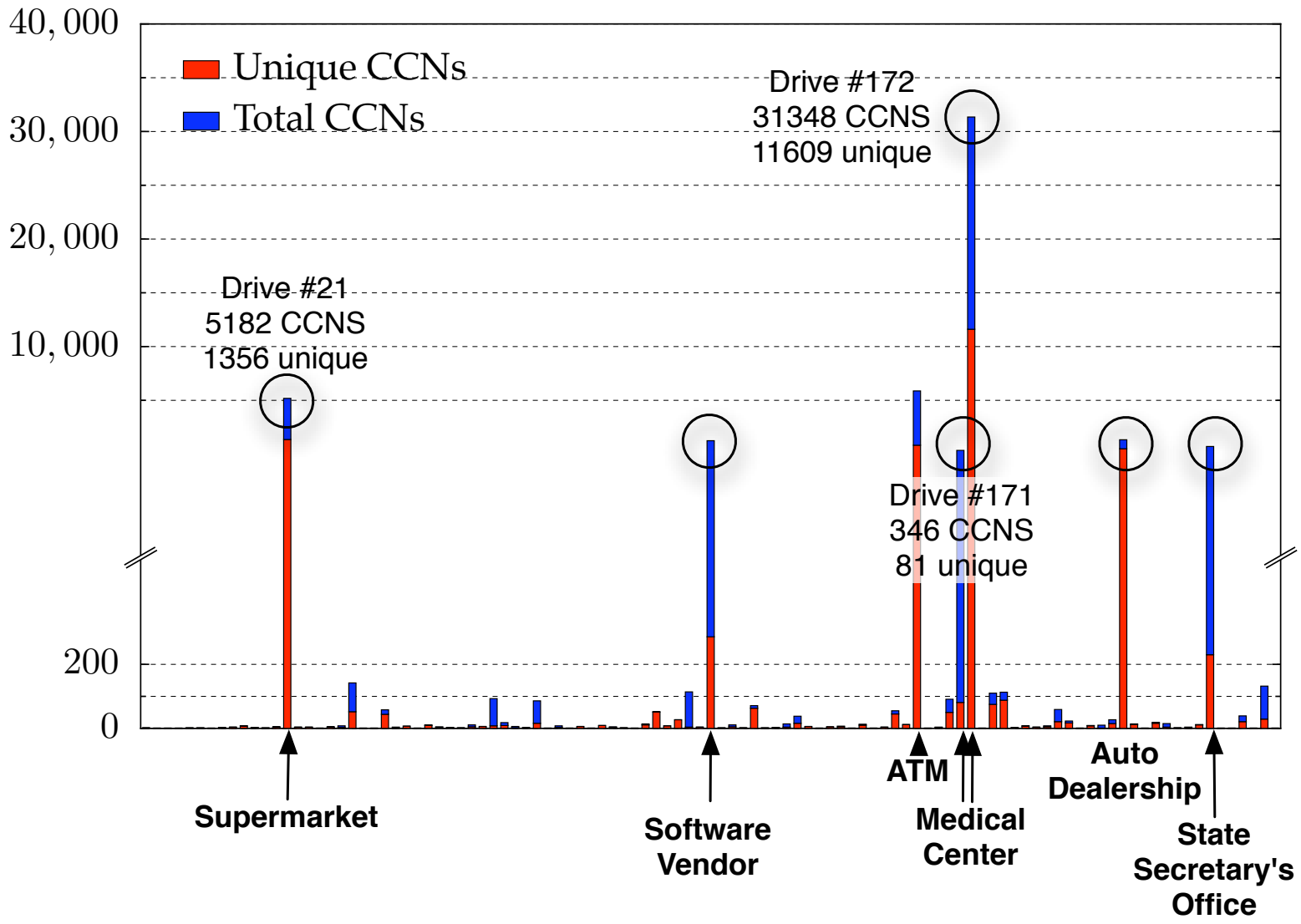
The email histogram works even if you can't find any files.

The email histogram approach works quite well.

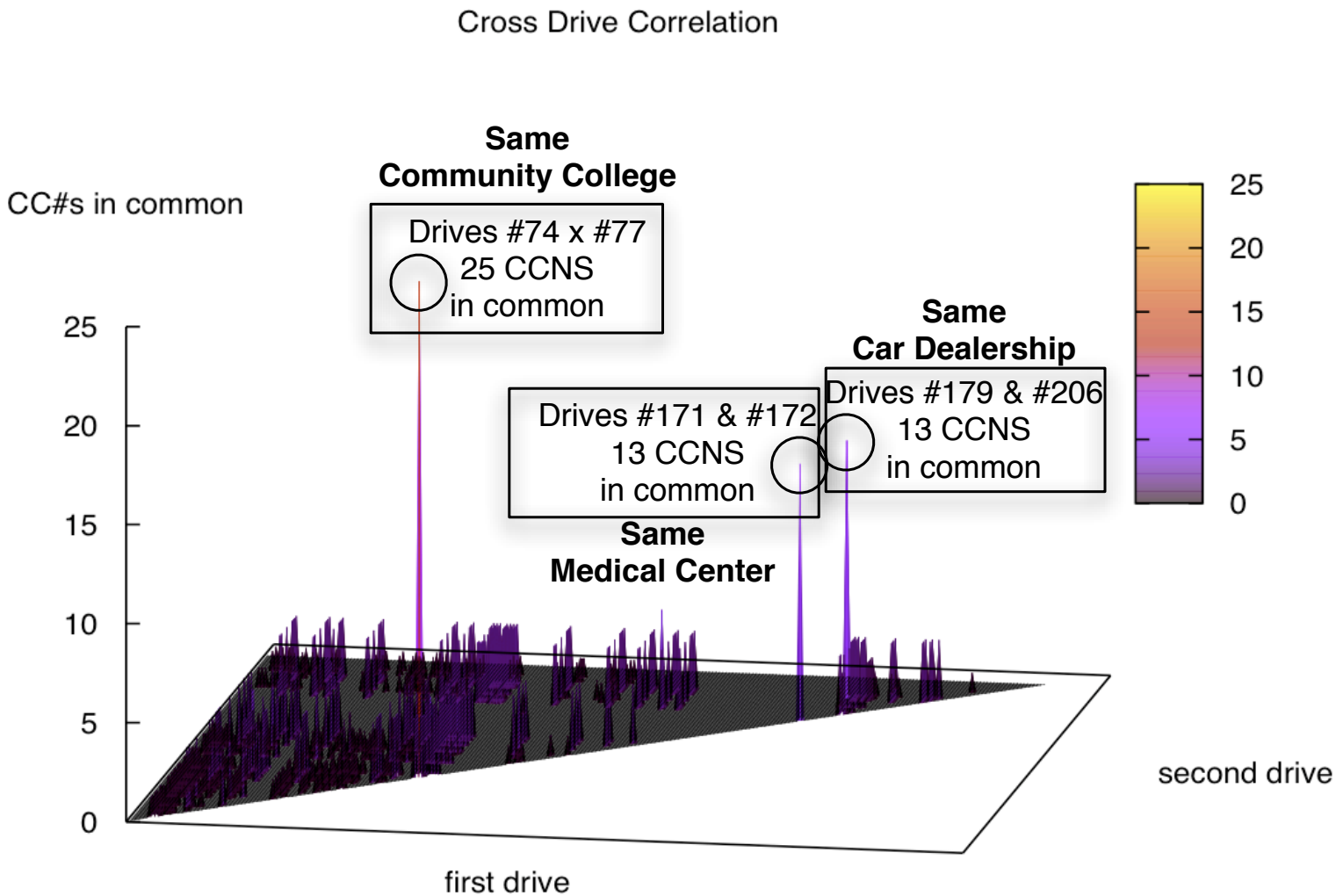
Drive #51: Top email addresses (sanitized)

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

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 applications:

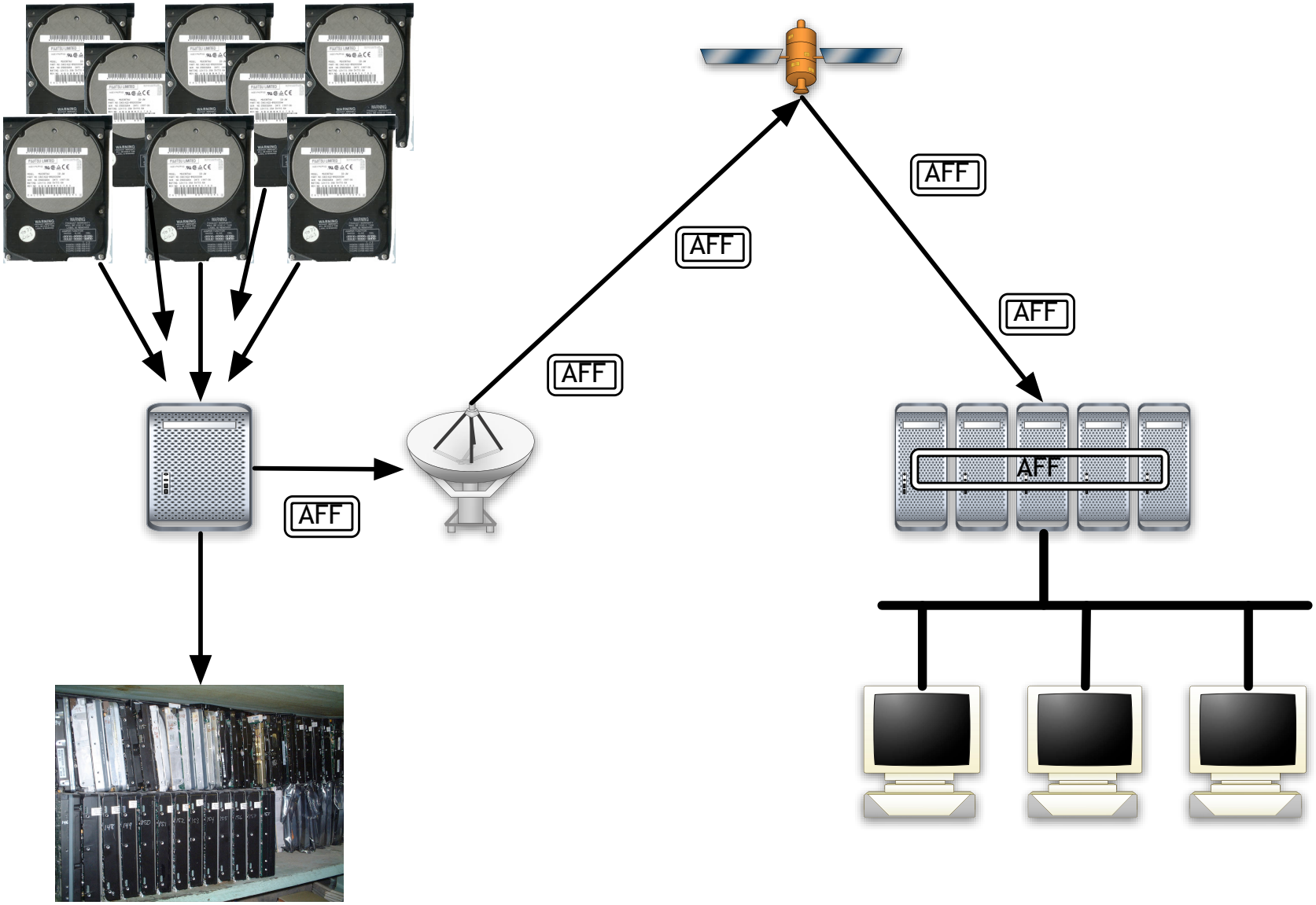
Possible Identifiers:

- CCNs
- Email addresses
- Message-IDs
- MD5 of disk sectors

Possible Uses:

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

AFF is a simple, compact, self-describing, and open way to capture and move around disk images.



AFF consists of four parts:

AFF specification	— Defines schema and data storage.
AFFLIB	— Open Source C/C++ implementation.
AFF Tools	— Utilities for working with AFF files.
AImage	— The Advanced disk imager.

All can be downloaded today from `http://www.afflib.org`**.**

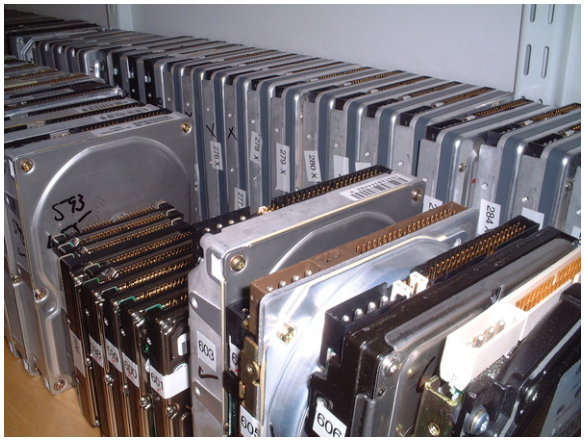
Example: AFF file of a good disk (converted from raw).

```
% afinfo -a /project1/affs/47.aff
                        file                 data
Segment               offset       arg     length    data preview
=======               =======  =======     ======    ============
pagesize                  585 16777216          0
imagesize                 616        0          8    = 1629831168 (64-bit value)
md5                       657        0         16    .U[...'L....ng..
sha1                      700        0         20    .8..l..).(......kMcI
page0                     748        1   10488912    x....|\u.7..4...^...E.....
page1                10489688        1   16398437    x...@..=|....6Mh.jz...A...
page2                26888153        1   16305513    x..}..%E.v.T....`W.t.D...(
page3                43193694        1   16665964    x...@....E.A..8...05.x..`.
page4                59859686        1   16742440    x.DyS.O.....m..m{..m..m..m
page5                76602154        1   16726198    x...@..4.....,}....x.O.M..
page6                93328380        1   16768092    x...@..9..Vd.|3...NF..u..
...
```

Almage is the Advanced Disk Imager.

```
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.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%)
```

Almage has been used to image more than 800 hard drives.



Automatic capture of metadata is exceedingly important!

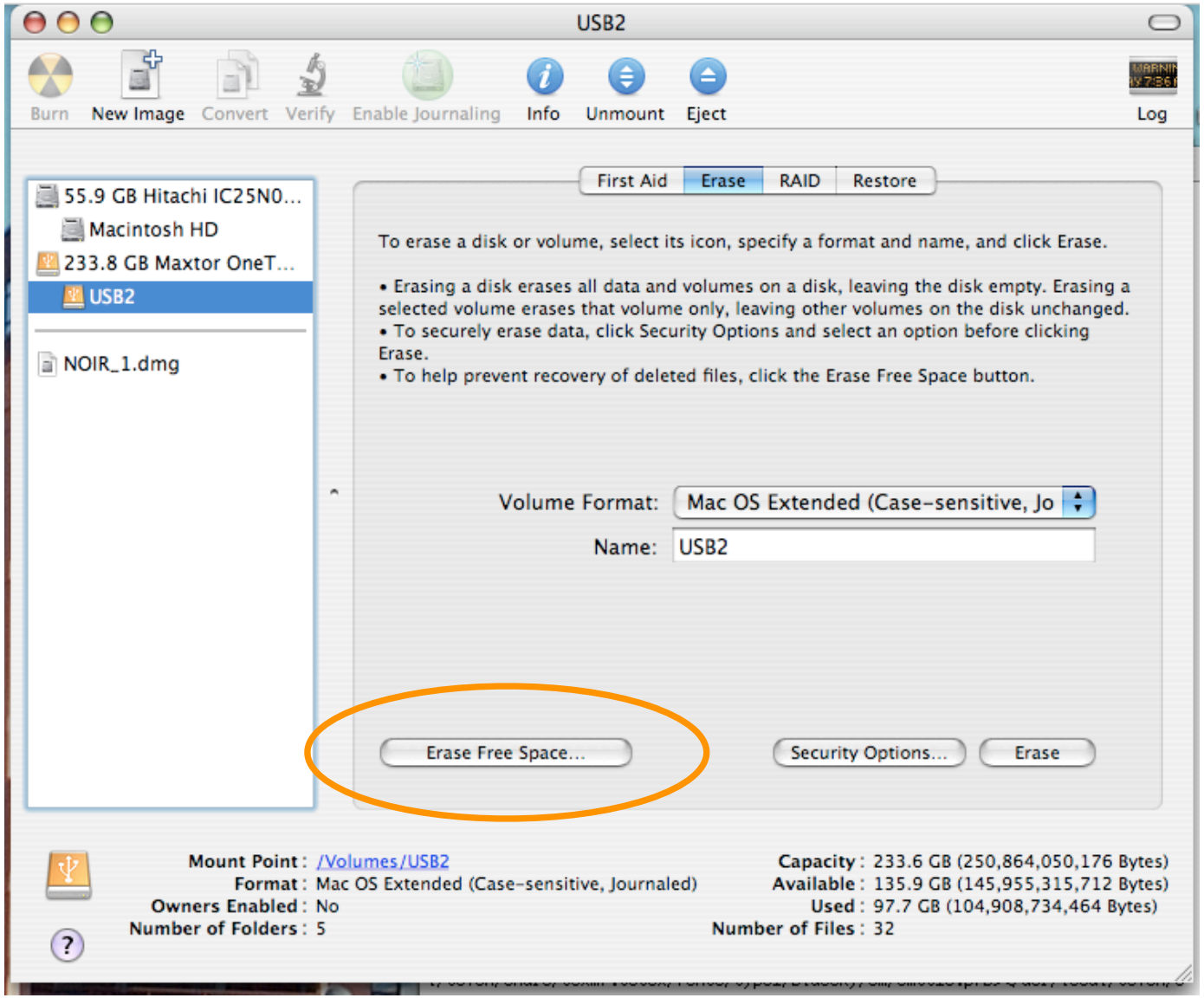
Legislative reactions to this research: "Fair and Accurate Credit Transactions Act of 2003" (US)

- 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."

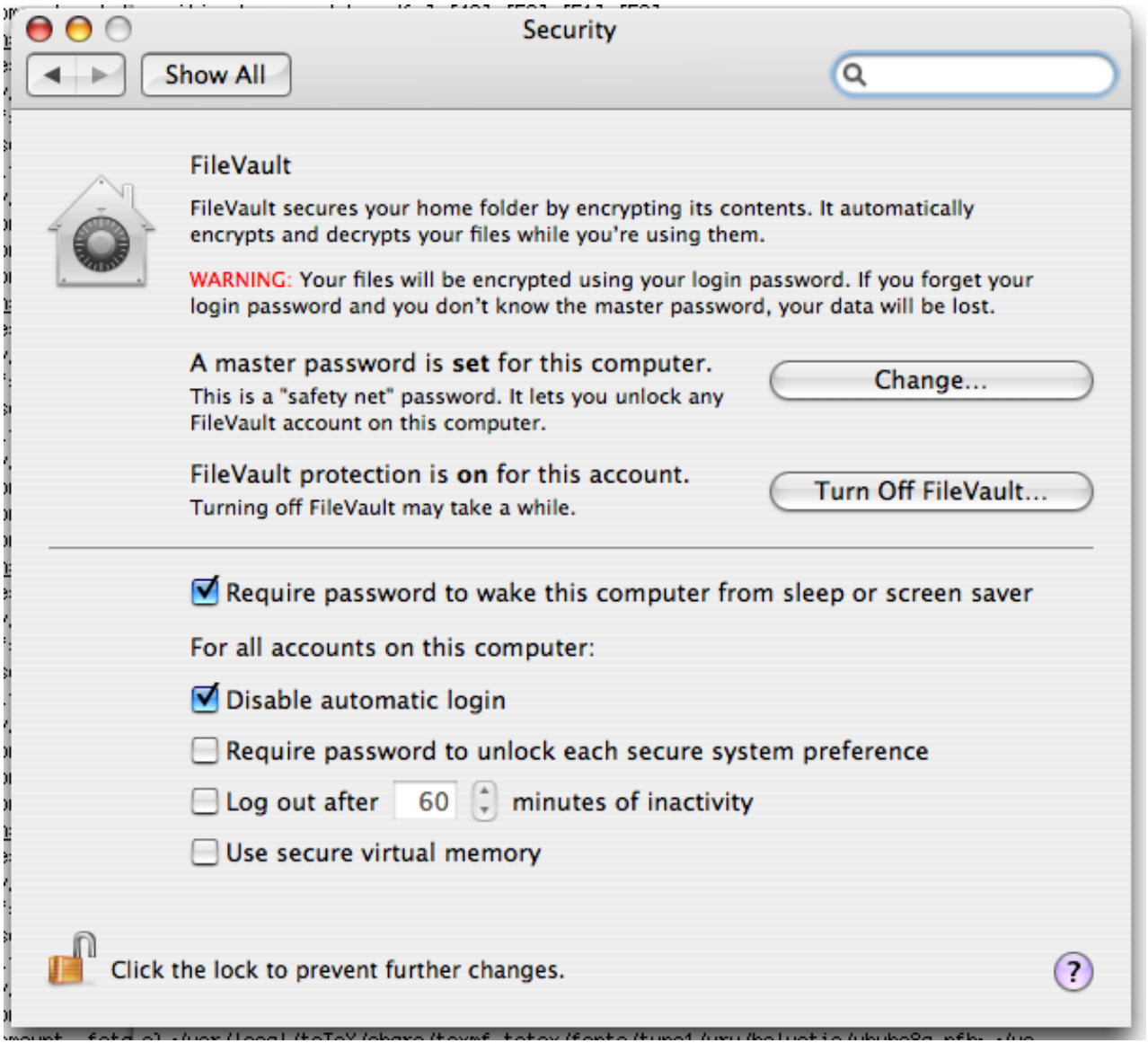
Testimony: `http://tinyurl.com/cd2my`

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

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

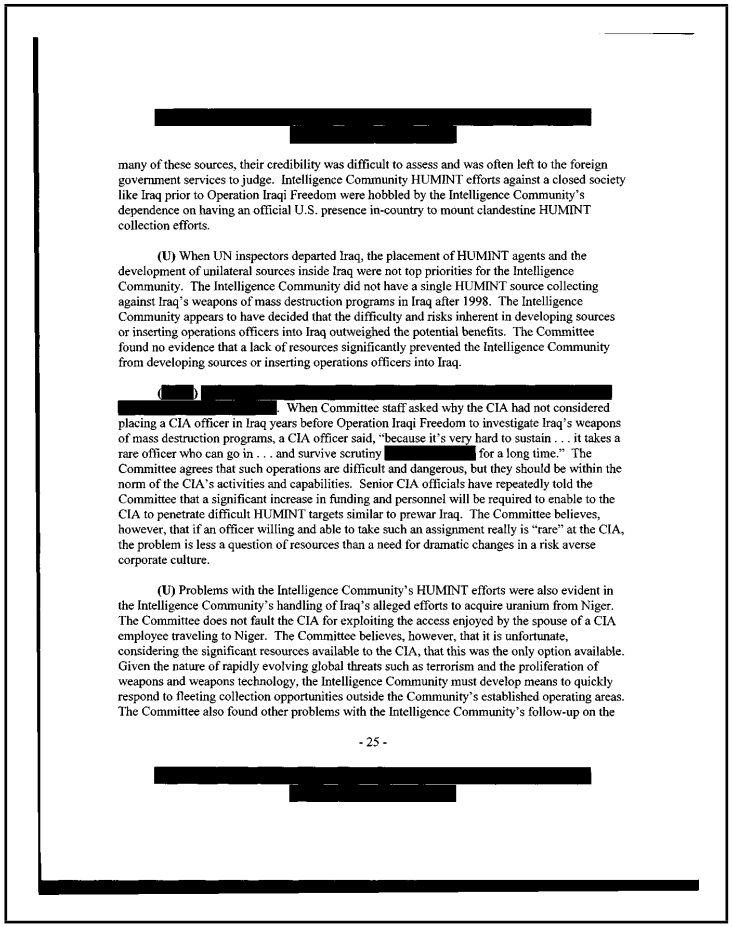


MacOS “File Vault” gives users an encrypted file system.



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.



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.

(U) . 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 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

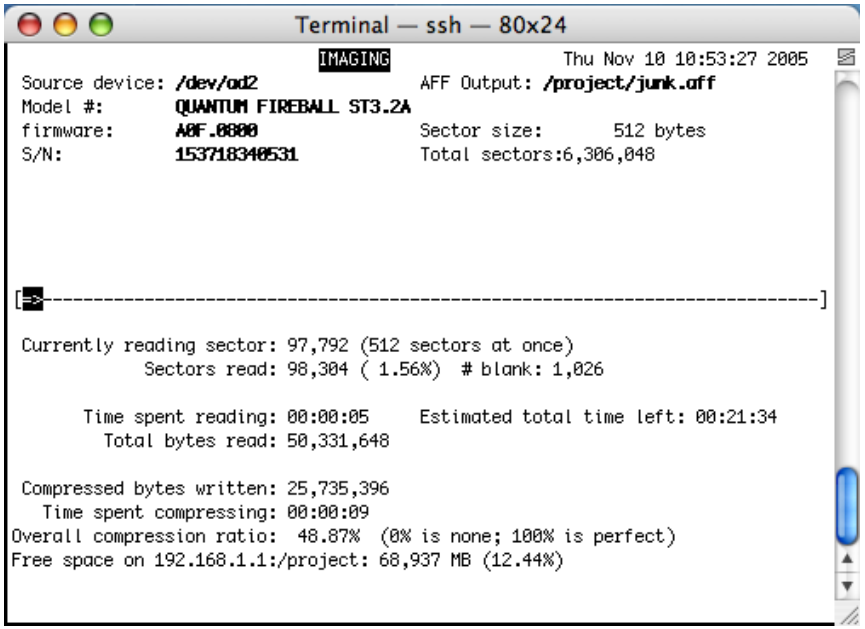
- 25 -

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.



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.